AD-A050 752 DEFENSE SYSTEMS MANAGEMENT COLL FORT BELVOIR VA F/G 5/1
DEFENSE SYSTEMS MANAGEMENT REVIEW. VOLUME I, NUMBER 5.(U)
1977 W E LOTZ, I E DONOVAN, P W POWERS

UNCLASSIFIED NL

1 OF 2

AD
A050 752

12

DEFENSE SYSTEMS MANAGEMENT

REVIEW

DEPARTMENT OF DEFENSE
SYSTEMS MANAGEMENT COLLEGE
DUCTUS DOCTRINA DOMINATO

VOL I, NO 5.

WINTER 1977

DEFENSE SYSTEMS MANAGEMENT REVIEW

PURPOSE

The purpose of the Defense Systems Management Review is to disseminate information concerning new developments and effective actions taken relative to the management of defense systems programs and defense systems acquisition.

The Review is designed as a vehicle to transmit, between persons in positions of leadership and responsibility in the program management and systems acquisition communities, information on policies, trends, events and current thinking affecting the practice of program management and defense systems acquisition. The publication serves as a means for providing an historical record of significant information associated with defense systems acquisition/management concepts and practices.

The Review supports the assigned mission of the Defense Systems Management College, and serves as a medium for continuing the education and professional development of persons in the field.

Review: US ISSN 0363-7727

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
VOLUME I, NO. 5

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
6 DEFENSE SYSTEMS MANAGEMENT REVIEW. Volume I, Number 5.

5. TYPE OF REPORT & PERIOD COVERED
PERIODICAL 9 QUARTERLY rept.
1 Sep 77 - 1 Jan 78,

6. PERFORMING ORG. REPORT NUMBER
1 SEP 77 - 1 JAN 78

7. AUTHOR(s)
10 [redacted] Walter E. Lotz, Jr., [redacted] Ian E. M. Donovan,
[redacted] Patrick W. Powers, [redacted] Thomas A. Fitzgibbons
[redacted] Howard W. Leaf, CAPT T. H. Sherman

8. CONTRACT OR GRANT NUMBER(s)
Approved by: OSD(PA)
18 May 76

9. PERFORMING ORGANIZATION NAME AND ADDRESS
DEPARTMENT OF RESEARCH AND PUBLICATIONS
DEFENSE SYSTEMS MANAGEMENT COLLEGE
FORT BELVOIR, VA 22060

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
DEFENSE SYSTEMS MANAGEMENT COLLEGE
FORT BELVOIR, VA 22060

12. REPORT DATE
[redacted]R 1977 11 1977

13. NUMBER OF PAGES
176 12 184p.

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)
UNDER SECRETARY OF DEFENSE RESEARCH AND ENGINEERING
OFFICE OF THE SECRETARY OF DEFENSE
ROOM 3E1006, PENTAGON
WASHINGTON, DC 20301

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
N/A

16. DISTRIBUTION STATEMENT (of this Report)
Distribution of this document is unlimited. (As defined in AR 70-31 and prescribed in DOD Directive No. 5200.20.)

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Test and Evaluation
TRI-TAC
Air Force Test and Evaluation
Test and Evaluation in the Navy
DOD Test and Evaluation Directives
Specifications and Standards for Test and Evaluation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
This issue is devoted to the practice of test and evaluation of military equipment and materiel used in the US Armed Services.

The purpose of the Defense Systems Management Review is to disseminate information regarding new developments and effective actions taken relative to the management of defense systems programs and defense systems acquisition.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

410 036 JOB

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ITEM 7 Continued

George G. Gleghorn
John W. McCord
LtCol David L. Washington
Howard W. Kreiner
Charles W. Karns
LtCol Michael J. Butchko
LtCol Martell D. Fritz
David T. Love
COL Ben H. Swett
John W. Rustenburg
Mark A. Owen
William L. Geese
LtCol Robert B. Machen

ITEM 20 Continued

The Review is designed as a vehicle to transmit, between persons in positions of leadership and responsibility in the program management and systems acquisition communities, information in policies, trends, events, and current thinking affecting the practice of program management and defense systems acquisition. The publication serves as a means for providing an historical record of significant information associated with defense systems acquisition/management concepts and practices.

The Review supports the assigned mission of the Defense Systems Management College and serves as a medium for continuing the education and professional development of persons in the field of defense systems acquisition/management.

DEFENSE SYSTEMS MANAGEMENT COLLEGE

Dear Reader:

This issue of the Review is devoted to the Acquisition Management area of Test and Evaluation.

The problems associated with Test and Evaluation are many and are very critical at every milestone of the acquisition process.

As the costs of testing increase, the Project Manager has a very difficult task in resource management against the requirements for technical decisions, the rigidity of schedule and funding constraints. The adequacy of the data base, the extent of credible instrumentation, and the use of simulations vs actual tests, all play a major role.

The need for Test and Evaluation planning early in all phases of the acquisition process cannot be overstressed. It is of particular importance to determine the criteria for success or failure well before the tests are planned and conducted so that evaluation of the results will have integrity. One of ~~our~~ the greatest difficulties is insuring the testing environment is as close to the operational environment as we can make it. Failure to do so may adversely affect the ultimate customer—the soldier, sailor, marine or airman.

possible.

In this issue, Test and Evaluation is addressed from different perspectives. Viewpoints vary as do the reported approaches to management. This issue was designed to stimulate dialogue and improve our Test and Evaluation plans and processes while, most importantly, achieving credible results.

R. G. FREEMAN III
Rear Admiral, USN
Commandant

In consonance with the intentions of the Secretary of Defense to reorganize his department, the functions and scope of the former Deputy Director for Test and Evaluation (DD(T&E)) are undergoing revision. The principal change involves the delegation of responsibility for operational testing to the Undersecretary for Program Analysis and Evaluation (PA&E). The Test and Evaluation organization in the Office of the Undersecretary of Defense (Research & Engineering) retains primary responsibility for developmental testing as well as test resources and remains the focal point within the Office of the Secretary of Defense (OSD) for all Test and Evaluation matters. This organizational realignment, coupled with a recent OSD-wide staff reduction, has the effect of reducing the former DD(T&E) staff by two-thirds. These changes in organization and functions will be reflected in a forthcoming revision to Department of Defense Directive, Number 5000.3; however, they do not affect the principles of test and evaluation stated in the articles in this issue.

DEFENSE SYSTEMS MANAGEMENT

REVIEW

VOL I, NO 5. WINTER 1977

CONTENTS

CONTENTS, Continued

DEPARTMENT OF DEFENSE SYSTEMS MANAGEMENT COLLEGE
DUCTUS DOCTRINA DOMINATIO

TEST AND EVALUATION IN THE DEPARTMENT OF DEFENSE

by

Lieutenant General Walter E. Lotz, USA (Ret)
Former Deputy Director of Defense Research and Engineering
(Test & Evaluation)

Test and Evaluation has played an important part in the significant improvements made in Defense systems acquisition management during the 1970's. Prior to this decade, emphasis was placed on the "total package procurement" concept in which a contract would be let for a complete system development and procurement program after an initial paper definition phase. The theory was that if a program was sufficiently defined in the beginning, a contractor could be expected to deliver the required product at a predetermined cost. Unfortunately, the total package procurement concept did not work well in practice largely because of overoptimistic cost and performance estimates and inaccurate and unverified inputs to initial definition phases. The result was that many programs managed under this concept experienced large cost overruns and significant performance deficiencies.

POLICY EVOLVEMENT

About 1970, many groups, including the Blue Ribbon Defense Panel, the Commission on Government Procurement and the Defense Science Board, began to recognize the deficiencies of then current defense acquisition management practices. Partly as a result of their recommendations, new policies evolved that emphasized demonstrated performance as the pacing function for defense programs. The key feature of the new policies is the periodic review of acquisition programs at critical milestones. During these periodic reviews, measured program progress is compared with program goals and objectives and a decision made to continue, reorient or cancel the program. A significant improvement offered by the new policies relative to those that had prevailed in the past is that program deviations can be detected quickly and corrective actions can be taken early.

To be effective, these new acquisition policies are heavily dependent upon reliable and accurate measurement of program progress as compared to program objectives. Test and evaluation, the primary means for making such measurements, became the cornerstone of the new acquisition policies and received emphasis in the implementation of these policies. One significant result of this emphasis was the establishment of the Office of the Deputy Director (Test and Evaluation) in the Pentagon's Directorate of Defense Research and Engineering. In addition, each Service was required to establish independent test agencies and strong headquarters staff focal points to assist in conducting required test and evaluation. As a further step, appropriate DOD and Service directives were promulgated to spell out the increased emphasis on test and evaluation in the acquisition process.

TEST, EVALUATION AND SYSTEMS ACQUISITION

There are two principal kinds of test and evaluation conducted in the systems acquisition process—Development Test and Evaluation (DT&E) and Operational Test and Evaluation (OT&E). Development Test and Evaluation is that testing conducted by, or under the supervision of, the development agency to evaluate technical performance of prototype equipment. This testing is generally conducted by skilled technicians and engineers under carefully controlled conditions. Operational Test and Evaluation is that testing conducted by military personnel to determine the degree to which new equipment fulfills military operational requirements. The Operational Test and Evaluation is conducted under conditions that duplicate as closely as possible the environment expected in field operations. Further, Operational Test and Evaluation is conducted on early production models as well as on research and development prototypes of new equipment.

FUNCTIONS

Development Test and Evaluation serves two important functions in the systems acquisition process. First Development Test and Evaluation assists in the actual design and development of systems. In this role Development Test and Evaluation is an integral part of the normal development process in which initial designs are converted to hardware. The hardware is tested, deficiencies are noted, and the design is modified as necessary. This process is iterated until the system hardware reaches a final design configuration. The development testing conducted in this iterative process can be thought of as providing the essential "feed back" information required to proceed from one step to the next in converting an initial paper design into fully developed hardware.

A second important function of Development Test and Evaluation is to provide information on the progress of new system development. The progress is ascertained by comparing measured system performance with a time-phased set of goals and objectives which have been established for the program. These time-phased goals and objectives are based on the premise that a system evolves or "matures" during development. In this maturing process certain characteristics, e.g., reliability, will typically improve, while other characteristics, e.g., system weight, may typically deteriorate or remain static. If we are sufficiently familiar with this maturing process, we can determine what performance levels should be attained by the system at various stages of development to reach the final, required levels of performance. Development tests then can be performed at frequent points in the program and provide a decision maker with information on the rate of maturation of the system. Additionally, a prediction of final system performance can be provided by comparing performance level progress with similar systems.

The functions served by Operational Test and Evaluation in the acquisition process are somewhat similar to those of Development Test and Evaluation. To the extent that Operational Test and Evaluation is performed on research and development prototypes, it aids system design and development by the early detection and correction of operational deficiencies. The detection and correction of deficiencies during development provides significant time and cost advantages by procluding the necessity for retrofitting operational systems.

Like Development Test and Evaluation, Operational Test and Evaluation also provides essential information for decisionmaking by comparing system operational performance with program goals and objectives. Since Operational Test and Evaluation conducted before system production involves testing of prototypes, to predict final system operational performance test results must be extrapolated as in Development Test and Evaluation. The degree of extrapolation required will depend upon the stage of system development and the realism of the simulated test environment.

Although performing the same generic functions in the acquisition process, Operational Test and Evaluation and Development Test and Evaluation have distinctive differences. Operational Test and Evaluation is concerned more with evaluating purely operational as opposed to technical factors and gives more emphasis to predicting future performance rather than evaluating current performance. With regard to the latter distinction, recognize that there is normally only one Operational Test and Evaluation conducted on a research and development prototype, and single point projections are error-prone. A combination of Operational Test and Evaluation with Development Test and Evaluation provides the most effective means to predict mature performance.

"TRY BEFORE BUY"

While much has been made of the "try before buy" aspect of test and evaluation, this aspect is a considerable oversimplification of the current role of test and evaluation in the systems acquisition process. Although there are other significant considerations, most important decisions in the acquisition process will normally require information relating to two basic questions:

- What is current program status? and,
- What is the likely outcome of the program?

As the preceding discussion indicates, this is exactly the kind of information test and evaluation provides. It is apparent that test and evaluation is quite important not only for major production decisions but for most other decisions made during the life of an acquisition program.

For major defense programs, test and evaluation impacts on the acquisition process primarily at the Office, Secretary of Defense level through the operation of the Defense Systems Acquisition Review Council (DSARC). The DSARC is an advisory group to the Secretary of Defense and is chaired by the DOD Acquisition Executive. Membership is made up from those who head major DOD staff offices. When a defense program reaches a major milestone, the DSARC meets to consider whether the program should be advanced to its next phase. The DSARC then provides its recommendations directly to the Secretary of Defense who makes the final decision. Although not a member of the DSARC, the Deputy Undersecretary of Defense, Research and Engineering (Test and Evaluation) DUSD R&E(T&E) plays an important role in the DSARC process. Prior to a DSARC he submits to each DSARC member his detailed assessment of test and evaluation conducted to date or planned for the future on the program under consideration. The Deputy Undersecretary of Defense, Research and Engineering (Test and Evaluation) also participates in the actual DSARC discussions leading to a DSARC recommendation to the Secretary of Defense. After the DSARC meeting, the Deputy Director submits a test and evaluation assessment directly to the Secretary of Defense for use when considering the recommendations of the DSARC. The effect of these procedures is to insure that systems acquisition decisions are based on a complete and independent assessment of a program's test and evaluation status.

Recent changes to Defense Acquisition policies delegate DSARC-type functions for some systems to the Service Acquisition Review Councils. In these cases test and evaluation plays the same important role.

Although sometimes difficult to achieve, independence and objectivity are quite important to the test and evaluation assessments provided for decisionmaking. Since most Development Test and Evaluation is an integral part of the systems design and development process, it is rightly conducted or controlled by the developing agency. Development agencies are, traditionally, "success oriented" and as such do not always give adequate weight to the problems and deficiencies that sometimes arise in a program. Similarly, operational testing cannot be conducted without the extensive participation of user personnel since user personnel are needed to achieve adequate realism. Using commands, however, may tend to minimize system deficiencies in seeking early deployment of a system, or alternatively may seek refinements in a system which, although "nice to have," are not essential to mission performance. To the extent possible, the effects of such potential biases must be reduced by providing independent assessment and reporting of testing. Such independence is achieved at the Office of the Secretary of Defense level by the DD(T&E) reporting test and evaluation assessments directly to the Secretary of Defense. Similarly, within the Services, the independent test agencies report directly to the Chief of Service rather than through either a developer or user chain of command. These independent reporting arrangements are specifically required by current DOD directives.

Since 1971 considerable emphasis has been placed on Operational Test and Evaluation, particularly that portion of operational testing that immediately precedes a decision to produce and deploy a weapon system. Operational testing provides an estimate of how the system will perform in the operational environment. Short of actual warfare, operational testing is the ultimate measure of an acquisition program's output and as such will normally receive considerable emphasis when decisions are to be reached about resource commitments to specific weapon systems.

Effective operational testing requires as much test realism as possible consistent with resource constraints and hardware status. This means the use of representative models of production systems, the participation of typical operational personnel in testing and, accurate simulation of both the threat and the physical environment. Accurate simulation of the operational environment may sometimes require two-sided testing involving simulated friendly and aggressor forces.

There are several other factors that must be considered in providng for effective operational testing. Operational test designs must allow for the exercise of systems over a reasonable range of operational conditions. We cannot fall into the trap of reaching a conclusion on the overall operational effectiveness of a system based on testing under a single set of conditions that may represent only a small percentage of likely tactical situations. Similarly, the scope of the testing must allow for all significant interaction between the system being tested and other systems with which it must function in the operational environment. Finally, it is important that objective measurements of a system's performance in simulated operational employment be made without compromise to realism. In the last few years we have made great strides in developing instrumentation that will allow us to make such objective measurements without interfering in actual test conduct.

Certainly not the least important aspect of effective utilization of test and evaluation in the acquisition process is the establishment of meaningful program goals and objectives whose attainment is to be demonstrated by test. The task is difficult and requires the interaction of many segments of the Defense community. For example, the operational forces have to provide information on critical capabilities needed to meet particular operational requirements. The system development commands must provide inputs on resource constraints and critical technical parameters. The test community must provide inputs on test feasibility and test facility requirements. Lastly, the analytical community must contribute by helping define measures of effectiveness, confidence levels, and data requirements. If the efforts of these various groups can be integrated effectively, we can be successful in defining proper goals and objectives against which program progress can be measured.

Test and evaluation has become an essential and integral part of the systems acquisition process. We recognize that there are other improvements which can and should be made to enhance test and evaluation effectiveness.

One such improvement is to begin limited operational testing earlier in the acquisition cycle of a weapon system. In the early 1970's almost exclusive emphasis was placed on conducting Operational Test and Evaluation just prior to the major production decision. This was because of the large funding commitment resulting from the production decision. Lately we have realized that earlier Operational Test and Evaluation may be needed for four reasons:

- In considering systems concepts during the early stages of a program, we need to know if these concepts are technically viable or tactically sound.

- To have any impact on the research and development phase of a program, Operational Test and Evaluation should be conducted much earlier than just before the major production decision since by then most research and development resources will have been expended.

- Cost and schedule impact can be minimized by early detection and correction of operational deficiencies.

- A recent trend toward initiating limited production before the major production decision has resulted in significant commitments of procurement funds prior to the major production decision.

For these reasons, we need to place more emphasis on early operational testing. While we cannot expect as much realism in early operational testing as can be achieved at a later program stage, the information obtained will provide significant improvements in the quality of inputs to decisions that are made early in the acquisition process.

Another improvement of considerable benefit is to relate test and evaluation efforts to the actual risks involved in a particular acquisition program. The link between testing and program risk is obvious. For the zero risk case we would probably need to do little or no test and evaluation to assist the decision process. For the high risk case, where we know little about possible decision outcomes, we might want to do a great deal of test and evaluation to enhance the probability of making the "correct" decision.

The implications of this relationship are clear. We need to expend more effort in defining the various technical, operational, cost and other risks that exist within an acquisition program. We need to encourage decision makers to define the level of risk that can be tolerated at various decision points. Given these inputs, we can "tailor" a test and evaluation program to better fit the risk situation that exists at a particular decision point. The result will be a more efficient and more economical test and evaluation program.

Another important function test and evaluation can perform is to provide an adequate transfer of "lessons learned" to the system design and engineering community. Often a deficiency discovered during test and evaluation on one program is nearly identical to that discovered on past or present programs. Test and evaluation can play a significant role in reducing this iteration. An example is provided by a current test and evaluation program in which we are modifying military specifications and handbooks and publishing appropriate design guides to insure that design deficiencies discovered in electromagnetic susceptibility testing of one system are not repeated in other systems. Be expanding such efforts to capitalize on lessons learned during testing, the test and evaluation community can improve efficiency and the end products of the weapon systems acquisition process.

PAYOFF

In summary, test and evaluation has become firmly established as an integral part of the weapon systems acquisition process. We in the test and evaluation community are now concentrating on improving the effectiveness of test and evaluation and adapting it to changing acquisition strategies. The ultimate payoff of these efforts will be maximum military capability for our limited defense dollars.

Lieutenant General Walter E. Lotz, Jr., USA (Ret) is the former Deputy Director, Office of the Director of Defense Research and Engineering (Test and Evaluation). The Office is now that of the Undersecretary of Defense, Research and Engineering (Test and Evaluation). Lieutenant General Lotz was appointed to this position shortly after his retirement from the US Army (1975). Prior Army career assignments include service as Commanding General, US Army Strategic Communication Command; Commanding General, US Army Electronics Command; and Deputy Director General, NATO Integrated Communications Systems Management Agency.

Lieutenant General Lotz holds a BS degree from the US Military Academy, an MSEE from the University of Illinois, and a PhD in Physics from the University of Virginia.

In March 1977 Lieutenant General Lotz received the Secretary of Defense Meritorious Civilian Service Award. Lieutenant General Lotz is a fellow of the Institute of Electrical and Electronic Engineers, and is a member of Sigma XI (national honorary research society).

STREAMLINING ARMY TESTING

by

Major General Patrick W. Powers, Department of the Army

The Army is ever changing. World War II veterans express shock at the differences between the brown shoe Army and the All Volunteer Force. Usually, caustic comments refer to the modern barracks, higher pay, better living conditions and other facets of the "new" Army. But, similar observations might be made when these veterans look at the changes in the way the Army equips modern volunteers and makes sure Army materiel is the best that Defense dollars can buy.

Materiel test and evaluation formerly was simple and less complicated. A weapon or piece of equipment was examined by the service boards for engineering quality, fielded for experimental purposes and accepted or rejected largely on the field commander's opinion of how the item performed while being handled by the troops.

Today, the process is more intricate. Several Army commands enter the test and evaluation picture in a variety of phases. Decisions involve the highest levels of management including Congress and the White House. It would seem that development and deployment of materiel would be cumbersome, more costly and, understandably, take longer. But should it?

TRANSITION

The old proverb, "The more things change, the more they remain the same," is true here. While there are many more players in the materiel life cycle now, the basic two, the Army and industry, remain—as does the traditional and common goal to cut production costs. Working together on new approaches to acquiring materiel, there is hope that these two partners can achieve some mutually beneficial inroads toward shorter and less expensive development cycles, while maintaining high standards of quality.

The standard manner in which the Army formerly conducted test and evaluation had to change. Reduced budgets and manpower was one reason. Another reason was that too much duplication was taking place, both among facilities and testers. Example: it took 13 years to field the Redeye Air Defense Missile. Much of that time was used to put the item through a series of sequential tests, making sure it would do the intended job. That approach is no longer affordable.

Today, new requirements, burgeoning costs and new technology require shorter development times. With high costs, layers of decisionmakers and a complex test sequence, there is an urgent need to streamline. Testing is a productive area in which to start.

Testing has become the key to the development process that accelerates as testing becomes more objective and uniform, and as schedules more completely integrate the efforts of the many participants. We need an approach that insures quality, economy and flexibility. We believe we are getting closer to it. After a difficult metamorphosis, the Army test community has emerged somewhat streamlined in its test and evaluation posture. This is what we look for:

- Shorter tests.
- Schedules where efforts can be combined and outlays reduced.

- Ways testers can work with industry and developers to further reduce expenditures.

Before explaining our present approach, let's look at how test and evaluation evolved.

HISTORY

Prior to 1924, when the Army issued regulations making test and evaluation integral to the procurement of weapons and equipment, Army-wide purchasing decisions were based on an officer board's experience with a weapon or second-hand reports from the field. From the Colonial period on, soldiers often purchased weapons on their own.

Following the 1924 decision and until World War II, the combat arms (infantry, cavalry, field artillery, coast artillery) requested development of new equipment and determined characteristics. Supply arms and services (engineer, signal, ordnance, chemical, quartermaster, medical) were charged with conducting development projects within their respective spheres.

Engineering tests, conducted by the developing agency, determined engineering and scientific factors. Service tests, conducted by the using agency, determined suitability for field use. The Secretary of War usually accepted or rejected the item based on the user's evaluation.

In 1940, with the establishment of a research and development section in logistics, each arm and service performed its own test and evaluation. Service tests were assigned to a board closely connected with the respective service school.

Under direct control of its respective arm of the service, each board was concerned only with equipment used by that arm. A central authority did not exist, and the evaluation by one board was not binding on the others.

In World War II, testing was accelerated and six new boards were established. At one point, because of heightened demands for equipment, service tests were conducted during production, and modifications were incorporated on the assembly line. The New Development Division, established in 1943 to supervise testing of new weapons and equipment, sent teams overseas to introduce equipment and observe equipment performance in battle. Theater commanders still made the final decision, however.

During the 1950's, new equipment became obsolete almost as soon as it became standard. It was obvious that pre-World War II philosophy and procedures would have to be adapted to the rampant technology of the postwar world or the United States would fall behind in modernization.

In 1962 the Department of the Army activated the US Army Materiel Command (AMC) for weapons and equipment development and procurement. The Army Materiel Command assigned testing responsibilities to the Commanding General, US Army Test and Evaluation Command (TECOM). This latter Command assumed responsibility for the test and evaluation missions formerly assigned to the technical services.

As the Army's principal materiel testing organization, Test and Evaluation Command was assigned the basic mission of providing decisionmakers with unbiased independent appraisals of Army materiel. The Army chartered the Test and Evaluation Command to reduce the time frame between design and production, and eliminate duplication of effort through integrated testing and better coordination.

To accomplish this charter, Test and Evaluation Command structured its operation around two traditional test types: engineering and service. The engineering tests revealed technical flaws in design that should be eliminated by the producer. The Army Test and Evaluation Command then placed the materiel in the hands of military personnel for service testing under realistic operational conditions. These tests determined the equipment's suitability for Army use and release for production. During early production, initial production tests determined if deficiencies found during previous testing were corrected. This cleared the way for full production of the item.

In 1970, the Presidential Blue Ribbon Defense Panel recommended that an operational assessment, independent of the materiel developer, be conducted prior to a major production purchase. Because of Test and Evaluation Command's affiliation with Army Materiel Command developers, criticism arose when the Test and Evaluation Command expanded its service test to include an operational exercise.

As a result, the Operational Test and Evaluation Agency (OTEA) was established (1972). The Operational Test and Evaluation Agency was given the management responsibility for all operational testing. The Operational Test and Evaluation Agency shares the management responsibility with the US Army Training and Doctrine Command (TRADOC) which was formed less than a year later. The Army Test and Evaluation Command continued to perform engineering testing and provided services for the operational testers because much of the capability for operational testing remained at Test and Evaluation Command test boards.

In May 1973, a joint Army Materiel Command/Industry meeting was held in Atlanta, Georgia. The outcome of this conference was a recommendation to simplify test procedures and eliminate duplication of effort.

This recommendation was followed by a study released by the Army Materiel Acquisition Review Committee in October 1974. This study boosted the trend toward simplification of the test process through mutual exchange of test data. Additionally, the Committee study recommended that the Test and Evaluation Command test boards be transferred to Training and Doctrine Command for operational testing, force development testing, and experimentation. The change drew a greater distinction between development and operational testing—the latter superseding service testing.

On January 23, 1976, Army Materiel Command changed its name to the US Army Materiel Development and Readiness Command (DARCOM) to better convey its twofold mission of developing materiel and making certain the materiel is ready for issue. The US Army Materiel Development and Readiness Command has a clear responsibility to develop hardware according to field unit needs, and to be sure the hardware is supportable in the field.

The evolution of Army materiel testing continues: changing, refining, constantly improving. The idea is to benefit from test trials and failures. In Army testing, the past is certainly a prologue to a successful future.

PRESENT DEPARTMENT OF ARMY TEST COMMUNITY

Today, Army tests are of two categories: development and operational. These tests have been known by other terms but still retain the traditional meanings. *Development tests*, performed by scientists and technicians or specially trained soldiers in controlled settings, determine how well the item works. *Operational tests*, accomplished by the ultimate user of the item, determine if a soldier can apply the item in a tactical environment.

Development testing is conducted to demonstrate that the engineering design and development process is complete; that design risks have been minimized; that the test item will meet specifications; and to estimate the system's military utility if made part of the inventory. This testing is accomplished in the factory, laboratory and proving ground.

The US Army Materiel Development and Readiness Command is responsible for most development testing and assigns the bulk of it to the Test and Evaluation Command. Historically, Test and Evaluation Command has provided Army decisionmakers with unbiased, independent testing and reliable test data. Now Test and Evaluation Command continues to make a valuable contribution in the research, development, test and evaluation effort, but assumes a markedly different posture. Emphasis has shifted from independent testing to independent evaluation. Test data is collected from contractors and proponent developers, as well as from tests executed by the Test and Evaluation Command, to support the evaluation.

The other test category, *operational testing*, is conducted to estimate the prospective system's military utility, operational effectiveness and suitability, and need for modifications. Operational testing also provides information on organization, personnel requirements, doctrine and tactics as well as verification of associated operating instructions, publications and handbooks. Operational testing is conducted in as realistic an operational environment as possible by the Operational Test and Evaluation Agency and Training and Doctrine Command. Both answer directly to the Army Chief of Staff.

The Operational Test and Evaluation Agency, with about 120 officers, is responsible for test design of major and designated nonmajor systems and provides an independent evaluation. The Infantry Fighting Vehicle is an example of a major system. The XM711 High Explosive Projectile for the 8-inch howitzer is an example of a nonmajor system. The Operational Test and Evaluation Agency either appoints a test director or tasks Training and Doctrine Command to perform the test. The US Army Forces Command (FORSCOM), along with the Materiel Development and Readiness Command and the Training and Doctrine Command, generally provides the military organization to perform the test.

The Training and Doctrine Command is responsible for individual training, education and combat developments. The Training and Doctrine Command also conducts operational tests of all assigned systems and provides test design and evaluation as directed for nonmajor systems.

To conduct tests, Training and Doctrine Command uses subordinate organizations. These include the Training and Doctrine Command Arms Test Activity, Fort Hood, Texas; the Combat Development Experimentation Command, Fort Ord, California; and seven test boards. The boards are the Airborne, Communications and Electronics Board, Fort Bragg, North Carolina; the Armor and Engineer Board, Fort Knox, Kentucky; the Air Defense Board, Fort Bliss, Texas; the Field Artillery Board, Fort Sill, Oklahoma; the Aviation Test Board, Fort Rucker, Alabama; the Intelligence and Security Board, Fort Huachuca, Arizona; and the Infantry Board, Fort Benning, Georgia. In addition, Training and Doctrine Command is responsible for Force Development Testing and Evaluation, which is user testing to determine user equipment requirements.

TEST AND EVALUATION COMMAND

The Army Test and Evaluation Command, with headquarters at Aberdeen Proving Ground, Maryland, directs and manages nine test agencies within the continental United States, Alaska and Panama. The command controls about 4 million acres of real estate. Acquisition cost of Test and Evaluation Command installations—with 20-million square feet of building floor space, utilities and equipment—approaches one-half billion dollars.

Authorized strength is almost 12,000, including nearly 9,000 civilians. Field organizations are categorized as proving grounds, environmental centers, a national range and the Aircraft Development Test Activity.

Direct management of test operations is done in Headquarters, Test and Evaluation Command, by six materiel test directorates (MTD) and the Test Operations and Policy Office. The latter monitors and coordinates the efforts of the materiel test directorates, records status of all test efforts, supervises workloading of subordinate agencies and develops timely test policy. The materiel test directorates have overall staff responsibility for efficient, economical and effective testing of materiel and are organized so that a single major system usually can be handled by one directorate as opposed to being fragmented among many.

In addition to the materiel test directorates, five other directorates address systems analysis and evaluation, test design and statistical analysis, methodology, human factors engineering and instrumentation, among other considerations.

The Army Test and Evaluation Command's installations and activities look like this:

Aberdeen Proving Ground, Maryland, is the Army's oldest proving ground. All weapons, except nuclear explosives and long-range missiles, can be tested there. Instrumented firing of weapons up to ranges of 42,000 yards is possible for the 100 firing positions available. The post and its waters occupy 82,000 acres. A chief facility is the Munson Test Area containing what has been called the "World's Worst Road" on which wheeled and tracked vehicles are tested.

Cold Regions Test Center, Fort Greely, Alaska, occupies more than 750,000 acres of typical, arctic terrain and is located about 180 miles below the Arctic Circle. Firing ranges, vehicular test courses, improved terrain areas, and maintenance shops are used to perform environmental tests on all types of materiel under extreme natural conditions.

The US Army Aircraft Development Test Activity is a tenant activity at Fort Rucker, Alabama. The Aircraft Development Test Activity has the responsibility to plan, conduct, evaluate and report on government test elements and to monitor, evaluate and report on contractor test elements of aircraft components and aircraft-related support equipment.

The US Army Dugway Proving Ground, Utah, conducts field and laboratory tests to evaluate chemical and radiological defense systems and materiel, as well as biological defense research. Dugway Proving Ground serves as the DOD joint control point for chemical and biological defense data. At this installation, 840,000 acres of Great Salt Lake Desert are used in conducting investigations in biological research and chemical, radiological, meteorological, ecological and epidemiological areas.

The Electronic Proving Ground, Fort Huachuca, Arizona, tests communications and electronics equipment and systems for the military services. The proving ground is situated between Yuma Proving Ground and White Sands Missile Range and provides overlapping, compatible instrumentation facilities for several types of inflight test programs. The clear electromagnetic environment, the excellent climatic conditions, and the freedom from aircraft congestion make this an unusually fine area for electronics testing.

Jefferson Proving Ground, Indiana, is devoted to testing ammunition. This is the Army's smallest Test and Evaluation Command installation, occupying 55,000 acres of southeastern Indiana. The proving ground's firing ranges can accommodate all weapons and weapons systems from grenade launchers and small arms to 240-mm howitzers.

White Sands Missile Range, New Mexico, with 4,000 square miles, is the country's largest military reservation. On July 16, 1945, the first atomic explosion was triggered at Trinity Site in a remote area of the 1-week old post. The range has gained fame for testing the moon lander, setting world records for reception of highly resolved photographs from weather satellites, and for diverse operations from nurturing exotic wild game to using the nation's largest solar furnace that generates up to 5,000 degrees Fahrenheit in environmental experiments. In addition, the White Sands Missile Range conducts tests for the Army, Air Force, Navy, National Aeronautics and Space Administration, Department of Energy and other federal agencies. The mission is to assess the performance of missiles, missile systems, space communications, test methodology and radiation propagation devices. A chief asset of the range is its extremely precise instrumentation. Instrumentation here includes telemetry, photogrammetry, real time data networks, transmission and communications associated with test and evaluation.

Tropic Test Center, Panama Canal Zone, is suited for tropical environment testing of materiel. Located 600 miles north of the equator, this test area has several facilities that provide a large variety of landforms, vegetation, soil, topographic, hydrological and climatic conditions for test and research projects.

Yuma Proving Ground, Arizona, occupies nearly a million acres in the Sonoran Desert. The proving ground's initial mission was desert environment testing. Over the years the mission and function of Yuma Proving Ground have grown owing to post size and the relative freedom from adverse weather interference. Now, like Aberdeen Proving Ground, test and evaluation are conducted at Yuma Proving Ground in a wide range of areas including weapons, munitions and automotive items. The Yuma Proving Ground has a multipurpose space position range designed to meet the need of aircraft armament tests. This installation also has more than a dozen vehicular test courses for automotive testing including sand dunes, rock and gravel courses, and water basins for amphibious vehicles.

CURRENT METHOD OF MATERIEL ACQUISITION

Today, the Army has four basic methods to use in filling its materiel needs. The Army can purchase existing commercial or foreign developed products. This provides a low cost, quick response to certain requirements. Second, the Army can modify commercial or foreign developed items. This method also provides for a quick response and reasonably low cost. Third, the Army can choose to product improve already existing US military equipment to meet new requirements. The fourth method is to initiate a new development program. The advantages in the last option are that the new item can meet exact military requirements: the design and configuration will be Government controlled and logistic supportability will be enhanced.

No matter which of the four options is chosen, the items must be tested to insure that they will work under realistic combat conditions. For the purpose of this discussion, the fourth option or a new development program has been selected.

The Army process for developing and fielding new items of equipment is formalized in a management model called the Life Cycle System Management Model (LCSMM). The Life Cycle System Management Model is divided into four phases. The phases are the program initiation phase, the validation and demonstration phase, the full-scale engineering development phase and the production and deployment phase.

As an item moves from one stage to the next, prototypes become better defined. Each phase includes both development and operational tests. These tests are conducted as early as possible and throughout the materiel acquisition process to reduce acquisition risks and to assess military worth. Schedules provide for accomplishing test and evaluation milestones prior to key decision points.

The latest approach to Army development testing is the Single Integrated Development Test Cycle (SIDTC). The Single Integrated Development Test Cycle aims at a truly integrated development test with everybody on board at the start—the developer, contractor, tester, evaluator and logistician. The Test Cycle embodies the objectives of reducing costs and saving development time and hardware by combining test phases and using contractor test data.

Before the Single Integrated Development Test Cycle, the Army's test cycle was almost entirely sequential. There was duplication and room for criticism. Testing was redundant and, in some instances, not particularly cost effective.

The primary objective of the Single Integrated Development Test Cycle is to eliminate duplication in development testing. The SIDTC emphasizes the total integration of development testing and calls for consideration of combining, and the concurrency of, some phases of development and operational testing. Important contributions have been the introduction of greater precision and discipline into testing and the emphasis placed on independence of evaluation. For SIDTC to work, the entire test and evaluation community needs to coordinate its requirements and programs as early as possible in the development cycle.

There are five blocks for implementing the Single Integrated Development Test Cycle. The blocks represent: coordination of effort, utilization of contractor test data, integration of tests, increased early soldier participation, and emphasis on independence of evaluation.

Consider coordination of effort. Program management responsibility is now focused more than ever on the materiel developer. The materiel developer must coordinate the efforts of all participants and integrate valid requirements into cost effective development. One of the principal means of accomplishing this coordination is for the developer—actually the project manager—to form and chair test integration working groups. The developer, contractor, development and operational testers and evaluators, and the logistic support people are intimately involved in this coordination process. This group must be composed of cooperative individuals with the authorization to make concessions with respect to the parochial interests of the participants. The materiel developer and his test integration working group establish and review the Coordinated Test Program to insure that maximum integration has been effected and that the test design is appropriate and cost effective.

The SIDTC concept requires maximum utilization of contractor test results. In the past, the Army received and utilized contractor data and reviewed contractor test plans to a limited extent. Now the Army is becoming more involved in monitoring contractor testing and using contractor test data. Army access to contractor data is steadily increasing as the Army is kept better informed about current and proposed contractor test planning.

The contractor's share of testing is being expanded, and test resources at contractor disposal are being increased. In effect, the Army's development test capability at proving grounds and ranges at major Army installations is available to the contractor, with reimbursement based on direct costs only. When feasible, industry can conduct contractor tests at Government test sites with contractor personnel. Government personnel can also perform tests on a customer basis.

The reduction of testing time and the integration of developmental effort are basic objectives of SIDTC. The Army has often been criticized for taking too much time in testing. Integration of test phases is one way to achieve production decision points earlier. Concurrent testing between development and operational testing is both feasible and practicable. In the past, operational testing often appeared as a separate block of time at the completion of development testing. Now, the Army must be prepared to realign its test programs to allow for concurrency, when concurrency is feasible, desirable and makes economic sense.

During the materiel acquisition process, items are taken through three stages of development and operational testing. The first development test (DT I) is conducted early during the validation and demonstration phase to demonstrate, fundamentally, that technical risks have been identified and that solutions are in hand. The first operational test (OT I) is conducted during the validation phase to provide an indication of military utility and worth to the user.

Development test II (DT II) tests an engineering development prototype during the full-scale engineering development phase. The DT II provides final technical data for determining a system's readiness for transition into production. Operational test II (OT II), conducted in as realistic an operational environment as possible, tests issues best examined by troop units in controlled field exercises.

Development test III (DT III) is conducted on limited production items to verify the adequacy and quality of materiel when manufactured. The Development Test III is not required if Development Test II is successfully completed and the system progresses directly to full production. Operational Test III tests initial production items to verify that all critical issues have been resolved. When a full-scale production decision is made, the subsequent testing may include production testing and, if necessary, follow-on evaluation testing.

The Army must consider not only the integration of Government and contractor testing, but also the integration of Development Test/Operational Test, especially in Development Test I/Operational Test I and Development Test III/Operational Test III phases of the life cycle. With respect to Development Tests/Operational Tests I and III, the Army must consider also the possibility of a reduction in scope of these programs. This may be possible if the scope of the Development Test II/Operational Test II effort is more comprehensive. The Operational Test and Evaluation Agency is currently investigating how to integrate its requirements for Operational Test I and Operational Test III with appropriate Development Test phases.

Too much modification of requirements could be costly. Reliability, that driving factor behind the duration of most development testing, may come to haunt us in the future. Experience has shown that reliability values reported from the field are usually lower than those predicted by design analysis, or demonstrated by laboratory testing. Since all participants are concerned about reliability, a new Department of Defense approach may help. Efforts are underway to establish a uniform set of reliability terms and definitions that can be tracked through all phases of research, development, test and evaluation. When this is done, the Army/Industry team will have an improved basis upon which to evaluate reliability.

Early user or troop participation in the development cycle, is the fourth major element. This emphasis has already injected troop know-how into some of our top priority weapon systems. Solutions to user equipment problems, surfaced early in the development cycle, insure compatibility of critical man-machine relations and acceptability of the design strategy.

Finally, the new emphasis on independent evaluation, with respect to the developer, is resulting in more efficient test designs and program structures. Additionally, it increases the credibility of the process. The increased technical planning requires intensive early management. It is expected that this early commitment of increased resources will generate substantial savings as the program develops.

The Single Integrated Development Test Cycle has been in full gear for 2 years. The method is continually being fine-tuned to be more responsive to the developer and decisionmakers. Inroads are being made into reducing costs and the length of the equipment acquisition cycle.

It is difficult to draw a comparison between the efficiency of testing as performed prior to SIDTC and testing as performed now. Systems currently under evaluation were largely developed prior to 1975. However, the primary objective of SIDTC was to eliminate unnecessary duplication in development testing. That has been accomplished to a large extent. The total integration of development testing and the combining of some phases of Development Test and Operational Test have received an encouraging start.

The most important contribution of the Single Integrated Development Test Cycle has been to introduce more precision and discipline into testing and to emphasize the independence of evaluation. Critical test issues are defined so that Development Test and Operational Test can be evaluated on a valid basis. Sequential development testing has been significantly reduced. Test integration working groups are coordinating the efforts of all participants and integrating valid requirements into cost effective development.

Despite problems arising from the introduction of the new system, large dollar and materiel savings have been identified in specific development programs. Continuing refinements point to further gains.

For example, in the Ground Emplaced Mine Scattering Systems program, three test integration working group meetings were held to review contractor engineer design and Test and Evaluation Command test programs. Test item reductions of approximately 80 percent were accomplished through the integration of development test, operational test, contractor and developer tests. Some redundant subtests were completely eliminated. The need for more than 1,000 mines was deleted from Test and Evaluation Command development test phases. Further reductions were achieved as a result of changes to testing methods and procedures, and the use of dummy, less costly prototype mines.

The Stinger, which is a shoulder-fired, air defense weapon system, is the planned replacement for the older Redeye. Significant savings in the development testing program were achieved by test integration working group efforts. In terms of development time, Stinger should be fielded in 8 years, as compared to 13 years for Redeye. While approximately 500 missiles were fired in developing and testing the Redeye, only 130 missiles were fired to develop and test the Stinger. The Redeye experienced cost overruns of about 300 percent in research, development, test and evaluation costs. The Stinger had only a 15 percent cost overrun in the same areas.

Another example of SIDTC application occurred in a new area of weapons development. The Roland surface to air missile, a European weapon system, was adapted for US testing and production. Here the emphasis was on increased cooperative testing with the French and Germans at European and US test sites. The thrust was to obtain more usable data per missile from European tests in lieu of tests to be conducted later in the United States. The data is to be obtained by using US telemetry on the European missiles and providing US high-performance aircraft targets.

The Army was interested also in earlier operational tester participation, testing of support equipment and initiation of training. Conversely, the Europeans were more concerned with the interchangeability of parts, interoperability of subsystems and design control of the US and European versions of the system.

A test review team, including US contractors, devised a cooperative test program that was successfully negotiated in principle with the partner countries. The price the Army paid in this case included additional front-end costs, but no increase in the overall program. A formal set of agreements with the Europeans commits the United States to detailed procedures of design control and adherence to the extensive testing support offered by the United States.

The United States learned much in this reverse flow of weapon systems technology from Europe. There is a difference in test philosophy, procedures, standards and techniques. If this sharing is to continue and progress, the United States must be prepared for the ramifications of technology transfer and the impact of the metric system on drawings, tools and instrumentation.

INTERNATIONAL MATERIEL EVALUATION

The United States is interested in sharing development with European countries. In November 1976, as a result of Presidential, Congressional and Department of Defense concern that the Army was not doing enough in this area, a new program was established. The Army Test and Evaluation Command was given the responsibility for evaluating foreign military weapons and equipment for possible US procurement as alternatives to developmental items.

The new mission will be accomplished with a small staff element at Test and Evaluation Command headquarters. The Test Operations and Policy Office will serve as the point of contact.

The new program will involve evaluating foreign systems that meet US inventory requirements. The object in considering foreign systems is to obtain improved capability, decreased costs, earlier operational availability and an optimum degree of NATO standardization and interoperability, within the constraints of existing US laws and regulations. The evaluation process is divided into three phases.

Phase one is the identification of possible foreign systems for evaluation. After a requirement is made known, a foreign capability search will be made. If potential condidates are identified, Test and Evaluation Command will attempt to verify the information that justified the appearance of the name of a foreign system(s) on the candidate list. An inprocess review committee is to decide which items merit a further, indepth evaluation.

In the second phase, more foreign data gathering is necessary. Evaluations to this point usually are made on data existing within US agencies. If the item is in production or deployed, copies of foreign development test reports will be requested. The Test and Evaluation Command will draft an evaluation plan to analyze and validate the data gathered. From this, a preliminary evaluation is to be drafted. Again, the inprocess review committee should decide either that further effort will be expended in the program or that the foreign alternative is not viable.

In the third phase, a limited test program is to provide the information not available through analysis of existing data. Funding will be established and the test integration working group process should determine the extent of testing to provide a final evaluation. At the end of this stage, regardless of the inprocess review committee's decision, the Test and Evaluation Command role is complete. Further effort follows normal materiel acquisition guidelines.

Currently, Data Exchange Agreements in testing and evaluation exist among the United States and some foreign countries. The new International Materiel Evaluation Program will utilize these agreements during the first phase of the program to increase information exchange and coordination.

Most of this exchange now is accomplished after each country has spent considerable funds on independent research and development projects. This new program should help eliminate duplication in research and development areas. Hopefully, the countries will share expertise and funds. The push for more integration and coordination with the Army's research and development sector thus spreads into development areas shared with foreign governments.

DARCOM TEST FACILITIES REGISTER

With the implementation of the Single Integrated Development Test Cycle, industry has a much greater role and responsibility in development testing. Industry must recognize that adequate test facilities are an essential part of cost effective testing. This is not to say that industry should begin developing extensive ranges, test facilities or instrumentation, but that industry should make maximum use of what is available, especially government facilities.

For its part of the program, Test and Evaluation Command is taking concerted actions to use all valid data from contractors. Furthermore, we are counting on industry for improved test data. There remain situations where contractor test facilities are inadequate. This often results in data that can not be used and necessitates additional costs for repeating the test at a government facility.

The use of Test and Evaluation Command facilities by industry is not new. At White Sands Missile Range, Test and Evaluation Command has conducted contractor testing for many years. Missile firing tests conducted by contractor crews are a common occurrence. At Aberdeen Proving Ground, Test and Evaluation Command conducted contractor testing of the two US XM1 prototypes to the full satisfaction of both contractors.

The Single Integrated Development Test Cycle has emphasized and formalized the process of contractor testing at government facilities. A new Army Materiel Development and Readiness Command regulation, 70–61 (Apr 77), will assist contractors in using government facilities. The new regulation defines test facilities as an inclusive term covering real estate, airspace, buildings, grounds, instrumentation, vehicles, equipment and personnel. But, the regulation applies only to those cases where a government contract is involved. Where a government contract is not in being, a separate regulation covers the use of Test and Evaluation Command test facilities by private industry.

The utilization policy expressed in the new regulation addresses three main issues. The first states Test and Evaluation Command test facilities will be made available for use by contractors when such arrangements prove to be cost effective to the government. The second emphasizes that provisions for use of Test and Evaluation Command test facilities will be specifically included with applicable requests for proposals, invitations for bid and government contracts under negotiation. The third issue stresses that priorities for test facility use by contractors will be equivalent to priorities afforded respective Army materiel development and production programs.

The new regulation simplifies and defines the requirements that contractors must follow to use Test and Evaluation Command facilities. The regulation should benefit both industry and the US Army. However, it is not enough to know that government facilities are available to contractors. Industry must know what facilities exist.

The Test and Evaluation Command has prepared a two-volume DARCOM Test Facilities Register to ease identification and selection of test capabilities. Volume I, listing 29 Army Materiel Development and Readiness Command installations and activities, is for general reference and describes major facilities worth $50,000 or more. Each facility is described as to what it is and what it can do. Environmental features and constraints also appear.

Volume I was published in May 1976. The Test and Evaluation Command is updating that volume to include annexes that describe 40 other Department of Defense test facilities as well as 60 contractor facilities that have provided test support to the Army during the last 5 years.

Volume II is a computer listing of instrumentation and equipment (20,000 pieces each worth $3,000 or more) designed to invite comparison of various options.

For the first time, all users in the research and development community now have available a complete statement of test instrumentation and test facilities. The Register provides a single-source, clearinghouse service. By using the Register, unnecessary duplication of facilities with the Army can be avoided. Program dollars, through the use of existing resources, can be concentrated in hardware development and not be diverted to buy already available instrumentation.

The Register will permit more intelligent planning and programming of instrumentation, improvement and modernization funds. Shortfalls in resources can be recognized earlier, before test facilities are needed for immediate project support.

Use of the Register should facilitate cooperation between Army and non-Army testers—other Department of Defense agencies and industry. The Register should allow for cost reduction and production of more reliable data. Even when development testing occurs outside Test and Evaluation Command, that command maintains responsibility for test execution. When any test facility is used, Test and Evaluation Command must be assured that the testing is conducted under conditions that meet Army requirements.

OBSERVATIONS

Today's testing methods are more effective than those of 20 years ago. As weapon systems and equipment become more sophisticated and complex, test planning and procedures become more complicated.

The Single Integrated Development Test Cycle, with its thrust for test integration and Army/Industry coordination, has been in operation for 2 years. Improvements in the system are measurable. But, there has been a price to pay for these innovations. Additional costs were often incurred early and management risks were increased. Overall, some time and money savings have resulted.

The Single Integrated Development Test Cycle is not perfect. As the policy is used, areas for improvement become noticeable. Industry must provide the Army with reliable data. More usable contractor test data would reduce government testing time and scope. Theoretically, if all requirements in the development test phase could be met by the contractor, using contractor or Army facilities, some of the ultimate objectives of the Single Integrated Development Test Cycle would be guaranteed.

On the operational testing side, hardware requirements and logistics support have been strong points, but training and personnel support areas often have been troublesome.

As weapon system complexity increases, maintenance becomes more difficult. Also it is more difficult to train soldiers to use the complex systems. Therefore, at the front end of the development cycle, user requirements must be specific so training packages can be complete to ready soldiers for operational testing. By the time of Operational Test I, troops should be trained to use the equipment so that the United States Army Training and Doctrine Command can determine how well the system measures up to desired performance levels. In that way, trainers know what kind of simulators will be necessary for training. Furthermore, contractors know earlier if the equipment must be simplified because it is too complex for the skill level of the soldier who is going to operate it. The most reliable system is never effective if a soldier can not use it in the operational environment.

With SIDTC, the time allotted for developing instrumentation and facilities for testing has been shortened. Previously, when testing was sequential, new instrumentation development could be postponed. Now, with several phases starting at once, instrumentation must be ready from the outset. Early in the life cycle, test facilities and instrumentation requirements must be identified. Then, with the use of the DARCOM Test Facilities Register, a determination can be made about what requirements must be developed.

In the final analysis, to reduce the duplication and unnecessary redundancy still in the materiel acquisition cycle, the cooperation between the Army and industry must improve. The potential to solve the problems associated with growing costs and rampant technology does exist. Ever-increasing efforts to pool data, expertise and facilities must continue. The Army continues to change, but we appear to be on the right track for improved testing of weapons systems.

●

Major General Patrick W. Powers is the Commander, US Army Test and Evaluation Command headquartered at Aberdeen Proving Ground, Maryland. He is a graduate of the US Military Academy. Major General Powers received his MS, Engineering, from the University of Southern California. He has attended the Command and General Staff College and the Naval War College.

Other publications authored by Major General Powers include a book titled, *A Guide to National Defense*, and numerous military and technical articles about artillery, military tactics and rocket aerodynamics and propulsion systems.

IMPROVING AIR FORCE INDEPENDENT OPERATIONAL TESTING

by

Major General Howard W. Leaf
Commander, Air Force Test and Evaluation Center
Kirtland Air Force Base, New Mexico

The Air Force Test and Evaluation Center manages the Air Force's Operational Test and Evaluation program and independently evaluates and reports on the operational test and evaluation of new Air Force weapon systems in compliance with Air Force Regulations AFR 23–36[1] and AFR 80–14.[2] The Air Force Test and Evaluation Center was established on 1 January 1975 to fulfill Department of Defense and Congressional desires that each service branch have an agency separate and distinct from the developing and using commands to conduct operational testing.

In this article the initiatives taken to improve operational testing are addressed and examples of the steps the Air Force Test and Evaluation Center is taking to put these initiatives into practice are provided. Before examining these improvements to operational test and evaluation, a brief history of Air Force Test and Evaluation Center involvement in operational testing is appropriate.

INTRODUCTION

Operational test and evaluation has received so much attention during the past 3 or 4 years that it is sometimes thought to be a distinctively new concept. It is not. Within the evolution of American Air Power there is a long history of both formal and informal test and evaluation. In fact, I would suggest that most, if not all, technological improvements have resulted from either a formal or informal evaluation of the technology's *operational* effectiveness and suitability. We improved our devices when they did not completely satisfy their intended use. Until recently, however, this evaluation process tended to occur too late in the life cycle of systems. In the case of technology of weapon systems especially, we waited for training or combat usage to show us the flaws.

Military aviation got its start in the United States, for example, not on *30 July 1909*, at Fort Myer, Virginia, when the Wright Brothers successfully demonstrated their Wright Flyer, but rather sometime later when Captain Benjamin D. Foulois who, in the course of teaching himself to fly, discovered the airplane's useful capabilities and its operational weaknesses. Hard landings often dumped the 135-pound Foulois to the ground. To rectify the problem, Foulois mounted a metal tractor seat on the airplane. When that proved less than satisfactory, Captain Foulois looped his Sam Browne belt through the tractor seat and strapped it tight around his lap. Contoured seats and safety belts have been a part of the aviation scene ever since.

More than 30 years later, early American participation in the European air war proved the need for long range escort fighters to accompany strategic bombers to the target area. Without fighter escort, bomber losses were prohibitive and deep penetration of the enemy homeland was risky and ineffective. As developed, produced, and deployed, the P-38, P-51, and P-47 fighters of the time did not have the endurance to go as far as was needed. The fighter planes were modified in England to carry belly-mounted auxiliary gas tanks that were fabricated in the field. The deficiency was recognized from operational experience and was corrected by modification in the field after, rather than prior to production.

By the decade of the sixties, the cost of correcting deficiencies, discovered from the operational use of systems after they were deployed, became excessive and resulted in unanticipated increases in life cycle cost. Clearly, operational test and evaluation as it was traditionally performed was taking place too late in the acquisition cycle to control either the cost or the military utility of new weapons.

There was a growing conviction among students of the acquisition system that operational test and evaluation (OT&E) needed to be performed sooner if it were to contribute to the effectiveness of procurement practices. The 1970 report of the Blue Ribbon Defense Panel said, for instance that, "...to be effective, OT&E must be a total process, using all appropriate methods of evaluation, which spans the entire life cycle of a system from *initial requirement* until it is phased out of the operational forces."[3]

Responding to that recommendation, DOD Directive 5000.3, 19 Jan 1973, specified that,

> "...Test and evaluation shall be commenced as early as possible and conducted throughout the system acquisition process as necessary to assist in progressively reducing acquisition risks and in assessing military worth."

and that,

> "...Acquisition schedules will be based, inter alia, upon accomplishing test and evaluation milestones prior to the time that key decisions which would commit significant added resources are to be made."[4]

In short, the basic concept of operational test and evaluation (OT&E) has not changed significantly. The difference now is in what we are doing, and how it is being done, and most significantly, when it is being done. The major objective now is to prove the military utility of new weapons *before*, rather than *after* they are produced.

BEGINNING OF AFTEC

In 1973 the Air Force Chief of Staff directed the establishment of the Air Force Test and Evaluation Center (AFTEC) at Kirtland Air Force Base, New Mexico.[5] The original manning called for 212 authorizations (141 officers, 25 enlisted, 46 civilians). These people were assigned to the AFTEC headquarters staff at Kirtland Air Force Base, New Mexico, and as AFTEC test directors at the field test sites where operational test and evaluation (OT&E) is actually conducted.

In its 4-year existence, AFTEC has become increasingly involved in planning, conducting, and reporting on operational test and evaluation for new weapon systems. Some examples are the F-15 and the A-10 aircraft, the Cobra Dane radar, and the Air Force Satellite Communications System. In this time, AFTEC has taken several initiatives and is working with Headquarters United States Air Force to improve procedures for conducting and independently reporting on operational test and evaluation. The primary areas include: providing operational test and evaluation in as realistic an operational environment as possible; establishing levels of performance for measuring a weapon system's effectiveness; and assisting in developing and formalizing procedures for a new deficiency reporting system for the Air Force.

The Air Force Test and Evaluation Center headquarters manages the operational test and evaluation on new major weapon systems to include: day-to-day activities of ongoing programs, provisioning for resources for the conduct of operational test and evaluation, and preparation of test plans and reports. The headquarters staff also supports the test directors at the field test sites. At the field test sites, the field test teams, under the operational control of an AFTEC test director, execute the tests and prepare the test reports. The field test team consists of people from the major air commands (MAJCOM) who will operate and support the system when it is deployed.

For example, an AFTEC operating location at Edwards Air Force Base, California, consists of separate test teams for the F-16 aircraft and the advanced medium short take-off and landing transport. Each of these test teams includes operations-oriented personnel from the using command and personnel from the supporting commands (e.g., the YC-14 test team consists of Military Airlift Command, Air Force Logistics Command, and Air Training Command personnel).* In each case, however, the operational test and evaluation purpose is the same: to provide an independent assessment of how well a weapons system works in an operational environment.

THE DT&E/OT&E DISTINCTION

Important to the understanding of independent operational testing is the distinction between development test and evaluation (DT&E) and operational test and evaluation (OT&E). The fact that development test and evaluation and operational test and evaluation necessarily examine the same features of a system is evident in the basic definitions of DOD Directive 5000.3. How can development test and evaluation "demonstrate that design is complete" without also estimating "the need for modifications" (an operational test and evaluation function)? The answer is it cannot. Why then are development test and evaluation and operational test and evaluation being asked similar questions? The reason is that their viewpoints are completely different.

Development test and evaluation is the responsibility of Air Force Systems Command whose viewpoint is primarily technical. Primary responsibilities are to answer the questions: (1) does the system meet engineering specifications; and (2) is the system sufficiently developed to enter production? These issues are generally satisfied by testing in a *controlled environment* that minimizes the chance that unknown (or unmeasured) variables will affect system performance. The testing is conducted by technical personnel skilled at fine-tuning to maximize performance.

Operational test and evaluation is primarily the responsibility of AFTEC and the using MAJCOM whose viewpoint is operator-oriented. Operational test and evaluation issues are: (1) does the system operate effectively in its intended operational environment; and (2) can "blue-suit" personnel maintain the system as envisioned in the maintenance concept. These responsibilities are best satisfied in an uncontrolled environment where many operational variables may affect the system performance, e.g., operator training and skill level, compatibility and interoperability with other systems and simulated threat environment. Testing is conducted by the operators and maintenance personnel who will use the system in the field.

An example from the E-3A Airborne Warning and Control System (AWACS) program might illustrate the difference between a developer's viewpoint and an operator's concern. The developer might satisfy his specifications if the AWACS surveillance radar functions properly and gives a specific detection range at a specific altitude (a test environment). The operational tester, however, would also want to know how well a radar scope operator and the radar (under the same conditions) follow maneuvering tracks in a hostile electronics countermeasures environment after the AWACS has been aloft and operating its surveillance radar for a number of hours (an operational environment).

---

*In this regard, AFTEC differs from the Navy's Operational Test and Evaluation Force (OPTEVFOR). The OPTEVFOR consists of a headquarters in Norfolk, Virginia, and test elements that conduct OT&E on weapons systems OPTEVFOR actually owns and operates. The Army's Operational Test and Evaluation Agency, OTEA, in Falls Church, Virginia, does business in much the same way as does AFTEC. The Operational Test and Evaluation Agency designs and conducts operational tests for major and Category 1 nonmajor systems; and selectively reviews, approves, and monitors operational tests for nonmajor system testing conducted by other Army testing agencies.

Such an example illustrates the difference between development test and evaluation concerns. But, the example does not mean that the developer and operational tester are opposed to each other. Rather, development test and evaluation and operational test and evaluation complement each other and form one path toward the same goal—getting an effective weapon system for the Air Force. In fact, the Defense Science Board recently reaffirmed the close relationship.

The Defense Science Board encourages a close interaction, "...particularly feedback from the operational test and evaluation to the developer. Interaction among development test and evaluation, operational test and evaluation, and close contact with the user pay very important dividends in terms of money, time, and operational suitability."[6] Thus, AFTEC will continue to support testing that gives both the user and developer required information and, at the same time, yields valuable operational test and evaluation data that is separately analyzed and reported by the AFTEC test team.

However, as indicated earlier, the development test and evaluation tester and the operational test and evaluation tester have different viewpoints of a system. This difference obligates AFTEC, as the Air Force independent operational test and evaluation agency, to conduct additional, dedicated operational test and evaluation. Such testing gives a better estimate of the full picture of a weapon system's operational capability. Recent findings by other government agencies make it clear that such dedicated testing is a keystone in the weapon system acquisition cycle before the Production and Deployment Decision. In keeping with its charter, AFTEC independently reports its operational test and evaluation findings directly to the Air Force Chief of Staff.

IMPROVING OT&E

Thus AFTEC was created to manage the Air Force's operational test and evaluation program. To fulfill this charter, AFTEC embarked upon a number of initiatives to more closely relate operational requirements to the acquisition of weapon systems and to improve the operational test and evaluation process. The most important of these initiatives deserve some elaboration: early operational test and evaluation involvement in the acquisition process; operational concept; real-world testing; levels of system performance; and deficiency reporting.

EARLY OT&E INVOLVEMENT IN THE ACQUISITION PROCESS

As part of its effort to improve operational testing, AFTEC now gets involved as early as possible in the acquisition process, as far back as Milestone 0, Program Initiation. These early phases of the acquisition process have traditionally been the exclusive realm of *development engineers*. Why, then, is AFTEC concerned about early involvement? Well, for several important reasons. First, so that developers and contractors can learn very early of specific operational considerations against which the system will be tested. Second, so that contractors and developers will have a much clearer idea of the resources needed to conduct operational test and evaluation. Finally, AFTEC can plan early for realistic tests that will provide answers to important operational questions voiced during the decisionmaking process, including those asked by members of the Defense Systems Acquisition Review Council (DSARC).

The MX strategic missile is an excellent example of this early involvement. The MX is still in the Demonstration and Validation Phase. Nonetheless, AFTEC has operational test and evaluation test personnel working with the System Program Office at Norton Air Force Base, California. Even though actual MX launches will probably not take place until early 1980, AFTEC involvement now will mean that operational considerations are integrated into the development and subsequent testing of the system. The AFTEC contingent will form the nucleus of the future test team that will conduct operational test and evaluation on full-scale MX missiles.

To identify future programs for early AFTEC involvement, a Planning Division has been established at Headquarters, AFTEC. This division works with the appropriate System Program Office and the Air Staff to insure that operational test and evaluation requirements are included in early program documentation and to make sure that an operational or real-world flavor is included in the test program from the very beginning.

OPERATIONAL CONCEPT

Closely related to the early start of operational test and evaluation is the requirement for an operational concept for a new weapon system. The operational concept describes how the weapon system will be employed by the operator. We have been working closely with Headquarters, United States Air Force in writing an Air Force-wide operational concept regulation. Basically, this regulation will require the using major air command to clearly define early in the acquisition cycle the employment, deployment, and support concepts for a system.

The operational concept will be a living document. As tactics or technology change, the concept will be updated. The document will provide a basis for system test and evaluation. In addition, it will provide for such items as the posturing of combat forces and standards for basing forces and equipment. In short, this user-written operational concept will be a basis for establishing specific operational objectives and evaluation criteria. Hence, AFTEC will be able to test against these criteria and estimate the effectiveness of the system early in the acquisition cycle.

REAL-WORLD TESTING

Early in this article, mention was made of *development test and evaluation* and *operational test and evaluation*,and the differences between them. The skilled engineers and the controlled conditions of development testing must eventually give way to blue-suit maintenance and testing under operational conditions whether the system evaluated be a radar, a communications network, or a fighter aircraft. In other words, development test and evaluation, conducted in a controlled environment, must give way to operational test and evaluation conducted in an operational environment. The ultimate test of a weapon system's operational effectiveness and suitability must be in the operational or real-world environment.

Two examples of real-world testing and a discussion of some steps AFTEC has taken to increase its interface with the users of new weapon systems will illustrate the point. First, in November 1976, AFTEC, working with Tactical Air Command, Strategic Air Command, and Aerospace Defense Command, tested the E-3A Airborne Warning and Control System in the largest opposing air operation (in terms of time and space density) ever conducted during peacetime. Over 400 aircraft from airbases in the western United States flew over the Edwards/Nellis range (located in Southern California and Nevada) in a simulated air battle between Warsaw Pact nations and NATO. A production E-3A was tested as it performed its full command, control, and surveillance functions during the exercise. Only an actual combat situation could have been more realistic.

A second example, the May 1977 F-15 European Tactical Test, shows that operational testing is not necessarily a lengthy process. The Air Force Chief of Staff wanted an operational evaluation of the F-15 in Europe. In less than 2 months, AFTEC, working closely with Tactical Air Command and United States Air Forces in Europe, planned, flew, and reported on the F-15 in Europe. Eight specific missions were flown against a simulated Warsaw Pact threat. This real-world operational test and evaluation yielded valuable information, within peacetime limitations, about the F-15, its fire control system and weapons, and the aircraft effectiveness in the European environment.

To increase the opportunities for such real-world testing, AFTEC has established two detachments, Detachment #1 at Ramstein Air Base, Germany, and Detachment #2 at Eglin Air Force Base, Florida. The Ramstein detachment will accomplish two goals: (1) future operational tests, like the F-15 test, will be planned and coordinated with US forces in Europe; and, (2) operational test and evaluation conducted in the US will always have a "think Europe" flavor because the AFTEC European detachment will review AFTEC Operational Test and Evaluation plans to assure realism. The Eglin detachment will work closely with the Armament Development Test Center and the Tactical Air Warfare Center. In particular, AFTEC will establish liaison with the Armament Development Test Center and the Tactical Air Warfare Center in the areas of operational test and evaluation reliability, maintainability, and supportability for weapon systems.

The concern for meaningful operational test and evaluation in simulated combat situations will become increasingly important in the future. In the 1974–77 time frame, AFTEC was primarily testing new types of hardware (e.g., the F-15, the A-10, and the AIM-7F missile) that are coming into the Air Force inventory. The next several years of operational test and evaluation will continue such testing but there will be more concentration on operational test scenarios (such as the Airborne Warning and Control System tactical test) that permit testing of refinements and enhancements to these systems.

LEVELS OF SYSTEM PERFORMANCE

In the past, AFTEC was all too frequently asked, "What criteria did you test the system against?" In some cases, AFTEC had to admit that evaluations were too subjective in nature. What was missing were clearly-defined operational test and evaluation criteria against which to test new weapon systems. The Air Force Test and Evaluation Center looked at program documentation for each new system (e.g., Required Operational Capability, Decision Coordinating Paper, Program Management Directive, and operationally significant values contained in contractual specifications, etc.). However, these documents are issue-oriented or engineering-oriented and do not always yield meaningful operational test and evaluation criteria.

The Air Force Test and Evaluation Center then started working with the using major air command of each new weapon system to spell out realistic evaluation criteria for specific test objectives. But, before specific criteria could be worked for each system, AFTEC and the major air command had to agree on a standardized terminology. The three key terms, "*thresholds,*" "*standards,*" and "*goals,*" are currently being defined and used in on-going programs.

Thresholds are quantitative or qualitative minimum-essential levels of performance/capability that permit mission accomplishment. These levels are based on:

- operational and maintenance concepts;
- the threat estimate;
- operationally significant performance contained in documents such as the Decision Coordinating Paper (DCP), Program Management Document (PMD), Test and Evaluation Master Plan (TEMP), etc.; and
- the capabilities of existing systems (when a valid comparison can be made).

Standards are quantitative or qualitative levels of performance/capability that will satisfy the operational requirements established at Milestone II for a fully operational system. These levels are based on:

- the threat estimate;
- operational and maintenance concepts that are approved at Milestone II; and
- other program documentation when appropriate.

Goals are quantitative or qualitative levels of performance/capability that will enhance the system or are new requirements identified after Milestone II. These levels are based on:

- operationally significant performance contained in documents such as the DCP, PMD, TEMP, etc.;

- identification of an increase in the threat estimate; and

- changes in the operational and maintenance concepts formulated at Milestone II.

Further, goals can be identified as additional system attributes or logistics support approaches not contracted for that, if achievable at a reasonable cost, will result in significant improvement to the system.

Performance below the thresholds is deficient. This requires management action to raise the level of performance to permit mission accomplishment.

An example of how thresholds, standards, and goals were used as test criteria on the F-4G Wild Weasel will illustrate the value of this new approach to operational test and evaluation. One important Wild Weasel operational performance parameter was emitter detection. The F-4G contract specification called for detection of any given number of threat radars within "x" number of seconds—but failed to spell out what percentage of time the F-4G should achieve this mark. The Air Force Test and Evaluation Command, working closely with the Tactical Air Command, looked at the threat, the expected combat scenario, and how the F-4G system operated. Based on this analysis and the current operational requirement of the F-4G, AFTEC and Tactical Air Command established thresholds, standards, and goals for the Weasel's radar detection rate. Thus, based on the operational test, the F-4G crew knew how to grade the system performance. For example, if the system had exhibited only the threshold performance (detection of a certain percentage of all emitter signals in "x" number of seconds), an aircrew would know there were emitters they were not detecting and fast movement was needed to avoid, say, a simulated surface-to-air missile threat. If the system failed to reach the threshold, it would be rated deficient, and the deficient hardware or tactics would have to be corrected.

This evaluation criteria approach to operational test and evaluation is not exactly new. Thresholds, standards, and goals were contained in the contract with the Wright Brothers for the first military airplane procured by the United States government in 1908.* The government wanted a heavier than air machine that could reach 40 miles per hour in still air conditions. The "*threshold*"speed was 36 miles per hour, the "*standard*"was 40 miles per hour, and the "*goal*" (the attainment of which would result in a bonus payment) was 44 miles per hour.

DEFICIENCY REPORTING

When a piece of hardware fails to perform satisfactorily during the test program, a deficiency report is written by the test team. The Air Force Test and Evaluation Command has been working with the major air commands and the Air Staff to put deficiency reporting in tune with the peculiarities of Air Force test and evaluation. This has resulted in a section in the new Technical Order 00–35D-54, The USAF Material Deficiency Reporting and Investigating System, which lays out separate deficiency reporting procedures to be followed during test and evaluation. As part of the implementation of this system, deficiencies reported to the System Program Office are now "prioritized," first by the urgency of the need for the correction (e.g., mission essential or mission degrading), and, then, by periodic rank ordering. Efforts are aimed at focusing the attention of Program Managers on the key deficiencies that must be corrected to make the weapon system effective in the operational environment. Deficiency "prioritization" is definitely not a new concept, but is receiving added emphasis today.

The Air Force Test and Evaluation Command is also placing a great deal of emphasis on tracking the status of deficiency corrections. Deficiency reports are tracked by the AFTEC test team from the time of discovery of the deficiency through determination of the proper

*Terence R. St. Louis, AFTEC historian, has a copy of the contract, dated 10 Feb 1908, between the Signal Corps, United States Army, and the Wright Brothers. The contract specifies minimum, acceptable, and superior performance goals—and appropriate payments thereto.

corrective action by the development engineers to physical verification of the "fix" by the test team or the using command. The purpose is to provide using command and AFTEC managers an up-to-date, accurate status accounting of deficiency corrective actions. This will help identify important corrective actions that may require using command or AFTEC management attention for timely resolution.

The recently concluded A-10 follow-on test and evaluation was used as a test case for the new Deficiency Report section of Technical Order 00–35D-54. The results were that deficiency reports were standardized and were "tracked" more easily from originator to the System Program Office responsible for correcting the deficiencies. Also, because the deficiency reports were in order of priority, managers focused attention on the key deficiencies right away.

OTHER AREAS

The Air Force Test and Evaluation Command is seeking to improve operational test and evaluation in several other areas. As weapon systems have become increasingly complex, so, too, have the related subsystems. In this regard, we are now directly involved in the *operational test and evaluation* of *computer software* and *simulator programs*--the former as they affect computer controlled programs, weapons, and command control communications, the latter because of the increased emphasis being given simulators to augment training. The Air Force Test and Evaluation Command is applying a computer simulation model, the Logistics Composite Model, to determine maintenance manpower requirements for new aircraft systems. The model "predicts" resources (people, parts, facilities) needed to support a weapon system over a given period of time and the impact of resource shortages on the operational status of a flying unit. Finally, AFTEC is developing an OT&E Management Document that will provide guidance on how to plan, conduct, and report on operational test and evaluation. It will serve to standardize operational test and evaluation throughout the Air Force.

SUMMARY

In summary, then, AFTEC is a young, independent Air Force agency that is in dynamic evolution in its efforts to improve Air Force *operational test and evaluation.* The Air Force Test and Evaluation Command is pursuing its charter on two fronts. First, it is actively seeking early consideration of operational test and evaluation requirements in the acquisition cycle even before prototype hardware is built. Second, as the lead Air Force operational test and evaluation agency, AFTEC is continually pursuing new initiatives to improve Air Force operational test and evaluation. These all add up to what AFTEC feels is the bottom line—to ensure the US Air Force is getting the most effective weapon systems possible.

REFERENCES

1. Dept of Air Force, "Air Force Test and Evaluation Center," AFR 23–36, HQ USAF, Washington, DC, 19 Jul 76.

2. ______, "Test and Evaluation," AFR 80–14, HQ USAF, Washington, DC, 19 Jul 76.

3. Dept of Defense, "Report of the Blue Ribbon Defense Panel to the President and the Secretary of Defense on the Department of Defense," Washington, DC, 1 Jul 70, p 88.

4. ______, "Test and Evaluation," Department of Defense Directive 5000.3, 19 Jan 73, p 2.

5. Terence R. St. Louis, *History of the Air Force Test and Evaluation Center, 1 January – 31 December 1975*, Air Force Test and Evaluation Center, Kirtland AFB, New Mexico, Sep 76, Chap 1.

6. Defense Science Board, "Report of Task Force on Test and Evaluation Policies," Office of the Director of Defense Research and Engineering, Washington, DC, 17 Feb 77, p 7.

Major General Howard W. Leaf is the Commander of AFTEC (the Air Force Test and Evaluation Center), a separate operating agency, located at Kirtland Air Force Base, New Mexico. Major General Leaf became Commander of AFTEC on 1 October 1976. His organization is charged with the management of Operational Test and Evaluation for the US Air Force.

Major General Leaf has served as a test pilot for the Air Proving Ground Command; worked as a geophysicist in the Air Force Office of Scientific Research; twice worked in the Office of the Deputy Chief of Staff, Research and Development; and has served as Commander of the 333d Tactical Fighter Squadron, Takhli, the 1st Tactical Fighter Wing, McDill Air Force Base, FL, and the 366th Tactical Fighter Wing, Mountain Home Air Force Base, ID. He has served as Assistant Deputy and Deputy Chief of Staff for Requirements, Tactical Air Command. Major General Leaf is a command pilot with 4500 hours, including 321 combat missions and nearly 600 combat hours.

Major General Leaf holds a BS in Geophysical Engineering from the Colorado School of Mines and a MS, Geophysics, from St. Louis University. He is a graduate of the Air Command and Staff College and the Industrial College of the Armed Forces.

OPERATIONAL TEST AND EVALUATION OF SHIPS: POLICY AND PRACTICE

by

Commander Ian E. M. Donovan, Department of Navy
and
Lieutenant Commander Thomas A. Fitzgibbons, Department of Navy

The primary reason for conducting operational test and evaluation (OT&E) in major ship acquisition programs is to reduce *operational* risks *rapidly* and minimize the need for modification. The second reason for conducting operational test and evaluation is to gain insights into tactical employment of the new warship. This is particularly important when new warfare capabilities are introduced, as in NATO hydrofoil and AEGIS ships. In this paper the authors discuss ship operational test and evaluation, the policy behind it, and the ways in which it is carried out in practice.*

MODERN WARSHIP COMPLEXITY

A modern warship—an aircraft carrier, a cruiser, a submarine—is the most complex, self-contained, highly automated and rugged assembly of equipment constructed by man. She must be capable of steaming and fighting independently or in coordination with other forces, including Allied Forces. Whether on or below the surface of the ocean, she must be at home in the environment, even when that environment is turbulent, as it frequently is. She must carry with her the men and stores to operate, maintain and support her weapons, sensors, propulsion machinery and auxiliary systems, far from shorebased support. When fuel, ammunition, or food are needed, she must be capable of receiving them at sea, while maintaining her battle readiness. She must be able to communicate classified and unclassified information over vast distances and she must be able to detect, identify, locate and attack a variety of air, surface and subsurface threats before she herself is attacked. For these reasons, a modern warship is an integration of practically all military/technical disciplines, and incorporates a complexity unthinkable 20 or 30 years ago. While this complexity is the source of her capabilities, it also causes many of her difficulties. Complexity increases the time required for ship design and construction; it magnifies reliability and maintainability requirements; and it escalates the need for, and cost of, test and evaluation.

POLICY FOR OT&E IN SHIP ACQUISITION PROGRAMS

Warship complexity is recognized in Department of Defense (DOD) and Navy acquisition directives. The test and evaluation policy contained in these directives is a codification of lessons learned in the prior shipbuilding programs, incorporating the best elements of successful ship acquisitions and applying them to future programs.

*Much of the material about the practice of ship operational test and evaluation has been taken from official COMOPTEVFOR evaluation reports to the CNO.

Department of Defense Directive 5000.1, "Major System Acquisitions," is the cornerstone of the policy, stipulating that "...Programs shall be structured and resources allocated to ensure that the successful demonstration of program objectives is the pacing activity...," and requiring that test and evaluation begin as early as possible.

Within the Navy, DOD Directive 5000.1 was implemented by SECNAV Instruction 5000.1, "System Acquisition in the Department of the Navy." In this Instruction, the Commander, Operational Test and Evaluation Force (COMOPTEVFOR), under the command of the Chief of Naval Operations (CNO), is assigned responsibilities as the Navy's independent test agency for the required operational test and evaluation (OT&E).

A Memorandum, subject: "Ship Program Procedures," issued by the Deputy Secretary of Defense on 30 September 1975 sets forth the *unique sequence of program events and milestones* for structuring major ship acquisition programs to comply with DOD Directive 5000.1. The memorandum establishes the standard procedure for communication of plans and decisions about the Navy's overall ship design, development, and acquisition programs.* The Navy initiates the communication at the beginning of each fiscal year with an estimate of all POM (Program Objectives Memorandum) ship development and acquisition programs, and the key dates (including operational test and evaluation events), preliminary cost estimates, recommended milestones, and, if necessary, rationale for variance from the prescribed program events shown below:

Ship Acquisition Program Events

Item	Normally authorized at Milestone
Initiate preliminary design	I
Initiate contract design; long lead time procurement for lead ship	II
Lead-ship detailed design and construction; Follow-ship long lead time procurement	III
Follow-ship construction	IIIA

If a prototype ship is to be developed, typical program events are:

Prototype ship detailed design and construction	IIA
Follow-ship construction (if any)	III

*"System Acquisition in the Department of the Navy," (draft) SECNAVINST 5000.1A, directs that ship programs designated as major be structured in accordance with the 30 Sep 75 DOD/Navy agreement, as well as DOD Directive 5000.1.

Department of Defense Directive 5000.3, "Test and Evaluation," specifies two fundamental types of operational test and evaluation: initial operational test and evaluation that precedes Milestone III, and follow-on operational test and evaluation that succeeds Milestone III. Department of Defense Directive 5000.3 specifies a test and evaluation policy for ships that is different from other system acquisitions. Recognition of ship complexity, the long construction times, and the need for a different management approach dictated the policy. Provision is made for progressive steps—development, test, engineering, integration, and more test—of complex combat systems at a land-based test site, before the major production decision for the ship class. For advanced nonnuclear ship propulsion systems, adequate test and evaluation of prototypes is required prior to the first major production decision. (Test and evaluation of Navy nuclear propulsion plants is accomplished in accordance with methods in use by the Department of Energy). Follow-on test and evaluation on the lead ship of a new class is prescribed for specified systems and equipments, and a new concept is introduced: "...if required, full ship operational evaluation to the degree feasible."[1] Within the Navy DOD Directive 5000.3 is implemented through OPNAV Instruction 3960.10, "Test and Evaluation."

PLANNING FOR OT&E IN SHIP ACQUISITION PROGRAMS

In the last 5 years, the Navy has invested heavily in technical and operational testing with emphasis on structuring milestone-oriented ship acquisition programs. Individual shipboard equipments are being subjected to technical evaluation by the developing Agency and extensive operational evaluation by COMOPTEVFOR before receiving approval for service use. The approval for service use process has been strengthened and directly tied to operational test and evaluation. Achievement of approval for service use is now an integral step in selecting new equipments for incorporation into warship designs.*

Land-based test sites, useful for both development test and evaluation and operational test and evaluation on new ship propulsion systems and combat systems, are an important engineering tool in many ship acquisition programs. The lead ship of a new class now is being subjected to far more comprehensive operational test and evaluation after delivery to the government.

The top level management tool used for planning the various phases and types of test and evaluation is the Test and Evaluation Master Plan (TEMP). The TEMP is the single Navy document that integrates development test and evaluation, operational test and evaluation, and production acceptance test and evaluation for a ship acquisition program. In accordance with the Navy governing directive for TEMP, OPNAVINST 3960.10, planning starts early. The TEMP, at least in outline, should be drafted before Milestone I, and have received approval from the Chief of Naval Operations by Milestone II.

The TEMP planning and coordination process is long and iterative—it must be, because of the uniqueness and complexity of the acquisition process for a major new class of warship. A ship TEMP concentrates on ship-related issues—design, engineering, and integration of the nonnuclear propulsion system and the combat system, including the ship's operational performance in specific warfare missions. The ship TEMP identifies only the separate equipment and subsystem development programs and the respective TEMP for these items. The ship TEMP then can be focused on relevant ship-level issues. Reference is

*OPNAVINST 4720.9D, page 4, sets forth the concept of approval for service use and criteria for decisionmaking. Approval for service use is "...that determination made by the Chief of Naval Operations, or other delegated authority, that new systems or equipments or significant alterations to existing systems or equipments have undergone appropriate test and evaluation, to the extent that there has been: (1) demonstrated reliable performance, in accordance with design specifications, in the intended or existing operational environment; (2) demonstrated ability to be operated and maintained by personnel with the level of skill anticipated to be available under Navy service conditions; and (3) sufficient evidence that the equipment can be supported logistically in a deployed status."

made to the component subsystem-level TEMP for subsystem-level details.

The TEMP is prepared jointly by the Ship Acquisition Project Manager and the Commander, Operational Test and Evaluation Force (COMOPTEVFOR). The TEMP is approved by the Chief of Naval Operations. The TEMP specifies what testing will be done when, and what resources are necessary to carry out the tests. Approval of the TEMP constitutes Chief of Naval Operation direction to conduct the test and evaluation program defined therein, including the commitment of fleet services to support TEMP. In addition to a separate status listing of all related subsystem TEMP, the ship TEMP includes a separate "Fleet Introduction Schedule." This schedule highlights the many tests, trials, and training actions that must be accomplished on the lead ship before it can be released for unrestricted service. The schedule provides a focal point for the ship designers, builders, testers, trainers, and users to begin coordinating the most critical and hectic phase in a new warship's life, her fleet debut. The "Fleet Introduction Schedule" and the considerations addressed in its preparation are vital to conducting orderly operational test and evaluation on the lead ship.

The Commander, Operational Test and Evaluation Force organizes for ship operational test and evaluation in a manner that parallels the ship acquisition structure. The separate nature of ship sensor, weapon, and propulsion machinery development programs, and the management need for a perspective of the total warship acquisition program are recognized. As a matter of routine, operational test directors from the COMOPTEVFOR staff divisions (Underseas, Air, Command and Control, and Surface Warfare) are assigned to plan, conduct, and report on operational test and evaluation for individual equipment. An additional operational test director from the Ship Evaluation Division of the staff is assigned to plan, conduct, and report on overall ship operational test and evaluation. Concern is with the *ship acquisition program* and *its objectives.* The operational test director oversees the work of the operational test directors who are evaluating individual systems that will eventually be integrated into the new ship design. The ship operational test director coordinates with the "ship team" to ensure that all relevant ship and equipment operational issues are addressed at the proper level in a TEMP. This procedure is not easy, but it is essential, because the saving in expensive resources used in operational test and evaluation can be significant. For example, a single firing of an advanced surface-to-air missile at a drone aircraft can generate data to meet the needs of the operational test director evaluating missile, fire control, combat, electronic warfare, and data link systems, while also serving the ship evaluation operational test director.

The operational test director is responsible for preparing the input of the Commander, Operational Test and Evaluation Force, to the TEMP. Specifically, this test director is expected to:

- Define areas of operational risk for critical test and evaluation issues.

- Delineate the performance parameters by which operational effectiveness and operational suitability* will be assessed through OT&E before each milestone.

- Prepare the "OT&E Outline," that includes for each milestone/test phase, the operational test and evaluation objectives, events/scope of testing/basic scenarios, and scope of the operational test and evaluation effort.

- Determine the resources needed for operational test and evaluation and resource criticality to the testing.

TYPES OF SHIP OT&E

Ship operational test and evaluation varies according to the type of ship development and acquisition program and according to when

*Operational suitability includes reliability, maintainability, availability, logistics supportability, compatibility, interoperability, and training requirements.

(during the development and acquisition program) it is conducted.

There are two primary types of ship acquisition programs:

1. Lead-ship, follow-ship program

2. Prototype ship program

In a lead-ship, follow-ship program, the acquisition strategy is structured to permit adequate test and evaluation of new subsystems ashore and at sea in surrogate ships, prior to release of funds for the first follow-ship construction (Milestone IIIA). This strategy represents a compromise between early operational test and evaluation on a prototype, and late operational test and evaluation on the first production ship, when follow ships will be delivered in rapid succession. The lead-ship, follow-ship program structure provides funding for the lead ship in year "1", no funding in year "2", and follow-ship funding in years "3" and beyond. This schedule permits the lead ship to undergo continuing phases of operational test and evaluation to and through fleet introduction, by which time all ships of the class may be under contract. While some concurrency of development and production exists, there is time for lessons learned and necessary modifications identified in operational test and evaluation and incorporated in the lead ship, to be fed back into follow ships while these ships are under construction. Examples of lead-ship follow-ship acquisitions are the FFG 7 and DDG 47 classes, each having a funding profile of 1, 0, X, X, X over the first 5 years. This program structure is applicable only for evolutionary ship designs with long construction times, i.e., typical displacement-hull warships.

The prototype ship acquisition program is one that includes significant research and development funding for a prototype, usually a new design incorporating major technical advancements not proved in nonnuclear propulsion of hull design. Examples of such innovative and unique ship types are hydrofoils, air cushion vehicles, surface effect and small waterplane-area twin-hull ships. In these and other advanced concepts in hull form and propulsion the Navy hopes for a quantum improvement in performance across the surface of the ocean (instead of plowing through the sea with a displacement hull). In this type of ship acquisition program, the unknowns are many and the issues go beyond a particular design. The entire operational concept and the military potential of the concept are scrutinized. The principal means for such a determination is initial operational test and evaluation of the prototype warship at sea before Milestone III, the first major production decision for follow-ship acquisition.

Usually the prototype with weaponry will undergo a full ship operational evaluation that evaluates the ship, not specific equipments. The individual equipments will have been subjected to prior operational evaluation within individual development programs. The ship OPEVAL objective is to see how well the ship *performs her missions,* not how well a particular system works. Indeed, the prototype *ship operational evaluation* goes further and *challenges the need for any planned follow-on acquisition.* For example, an operational evaluation was conducted at sea by Commander, Operational Test and Evaluation Force on the NATO Patrol Missile Hydrofoil. From this testing not only the operational effectiveness and operational suitability of the PHM advanced-design propulsion and control systems, struts, and foils were evaluated, but also evaluated was the operational concept of a high speed, missile-firing fleet ship.*

The OPEVAL, a term unique to the Navy, is the final phase of a system's initial operational test and evaluation (OT-III), and takes place *immediately* before the first major production decision. The test period is called OT-III to indicate that it precedes Milestone

*In the 10 years before PHM, the Navy designed, constructed, and operated four hydrofoil ships (USS HIGH POINT, USS PLAINVIEW, USS TUCUMCARI, USS FLAGSTAFF). The PHM class is the first USN hydrofoil constructed as a fleet warship with adequate firepower and command, control, and communications capability to permit operational risk, as well as technical risk, to be addressed thoroughly through test and evaluation.

III. The OPEVAL always requires operation and maintenance by fleet-type user personnel with the test system in a configuration as near the production configuration as possible. The evaluation is usually structured to stress the system in scenario-type tests, at sea, with all the vagaries of weather, threat, and operators.

Less visible than OPEVAL is early initial operational test and evaluation, OT-I and OT-II, that occur before acquisition Milestones I and II, respectively. The tests before Milestone I are usually addressed to concepts and feasibility. The extensive at-sea tests in the early Seventies of a conceptual sea control ship is an example of OT-I. In these tests, the USS GUAM (LPH 9) was designated an interim sea control ship for test purposes and new antisubmarine warfare task force tactics were evaluated. The tests before Milestone II represent the more critical early initial operational test and evaluation period because it holds the greatest potential to influence future fleet equipment from the operational point-of-view. These tests must use fleet-type personnel for hands-on operation of the equipment, but may use developing agency/contractor personnel for maintenance and support. Pre-Milestone II testing is concerned with the *operational direction* the weapon system design is taking, and its *potential* to be operationally effective and operationally suitable in the fleet. Follow-on operational test and evaluation (OT-IV and OT-V) occurs after Milestone III, and is directed to actual demonstration of operational performance and tactics. The difference is primarily that production systems are tested, and in OT-V, all resources are budgeted by the Fleet Commander instead of the acquisition manager.

INITIAL OPERATIONAL TEST AND EVALUATION AND RISK REDUCTION

Initial operational test and evaluation has as its objective an actual demonstration that operational risks have been reduced to an acceptable level before the system is placed in serial production. On a lesser system, one not so complex as a ship, OPEVAL would be conducted on a mature or nearly mature design, with the expectation that the production unit would not differ significantly from the model tested. Typical examples include a new shipboard radar, propulsion engine, or fire control system. In conventional warship acquisition, the situation is quite different. Since the ship does not exist at Milestone III or IIIA, initial operational test and evaluation must be suboptimized by looking at parts of the eventual total ship and estimating potential operational effectiveness and suitability. Consequently, the decisionmaker in reviewing the results of initial operational test and evaluation is concerned less with the details of design, and more with the *direction the design is taking* and the *track record of the developer* in meeting the specific milestone technical and operational performance objectives. The "earlier-than-usual" decisions required in shipbuilding programs are inherently risky. In addition to increased risk, the probability increases that the design or operational concept will include not only the known-unknowns, but also disquieting unknown-unknowns—*surprises*!

The decisionmaker is interested in all risks—technical and operational performance, cost, and schedule. Initial operational test and evaluation is concentrated only on demonstrated operational performance. The concept of technical risk is generally well understood, and is assessed by addressing the questions:

- Does the system meet the required technical characteristics and specifications?
- If not, what is yet to be accomplished?
- Is the remaining engineering development within the state-of-the-art?
- Can the system be produced by industry?

The concept of operational risk, while not nearly so well understood, is receiving increased attention. Operational risk is the risk that the warship might not be militarily useful once put to sea, achievement of stated requirements or specifications notwithstanding. This could be the result of the wrong specifications or technology applications having been invoked, lack of knowledge or training in the tactical use of the ship, or perhaps an inadequate assessment or oversimplified assumptions about the environment in which the ship would operate. Failure could result from the unknown-unknowns in areas never addressed in the design or requirements process. Consideration of these operational risks is of overriding importance, particularly in development of an entirely new ship-type like a hydrofoil or surface effect ship, because the ship will not be militarily useful without development of supporting systems, nor will it be useful without tactics and procedures for use in combat under nonbenign conditions.

These operational risks, to a greater or lesser degree, exist in all shipbuilding programs, but particularly in advanced technology areas, for example, when prototyping is necessary. The initial operational test and evaluation necessary to highlight the degree of achievement of performance objectives relative to the remaining operational risk, has become a fact of life taking several forms.

Ship Design

For initial operational test and evaluation during the ship design phase, operational *evaluation* plays a far greater role than operational *testing*. The role of the Commander, Operational Test and Evaluation Force is to continuously evaluate the design as it matures, and carry the fleet operator's message to the design engineer. The operational test director does not drive the design in any way, but rather is available to transfer firsthand knowledge and at-sea experience to assist the designer in understanding how the ship and its equipments will be used. Through participation in design reviews, and evaluation of compartment arrangements, equipment mock-ups, and the output of computer simulation models (or any valid design and engineering tools), the operational test director draws on his skills, experience, and ability to conceptualize ultimate fleet employment of the ship to make analyses and draw conclusions.

The discourse is two-way. It is not, during the ship design and engineering phase, a simple evaluator/evaluated relationship. The operational test director attempts, during the reviews, to gain a fuller understanding of the design. This understanding cannot come from a reading of specifications and drawings alone. It is equally important that the operational test director hear the specification and employment concept, and learn exactly what it is that the designer is trying to accomplish, and what assumptions are implicit but perhaps not obvious in the design. With this understanding, the operational test director can proceed to construct meaningful and realistic tests—tests that will allow the ship design to prove itself capable of meeting the letter and spirit of the original ship acquisition program operational requirements as approved by the Chief of Naval Operations.

When reviewing the ship design and formulating test plans, the operational test director is able to challenge the design. He can ask "what if" questions to define operational risks clearly and to surface operational issues that might need further attention. It is this *evaluation, rather than test,* that constitutes the important portion of initial operational test and evaluation in the ship design phase.

Individual Systems

Parallel to the ship design phase, individual subsystems identified for incorporation into the new warship (guns, radars, engines, aircraft catapults, missiles, etc.) undergo separate initial operational test and evaluation within the individual development programs. This evaluation culminates in individual OPEVALS and Milestones III, prior to the ship acquisition Milestone III. The final phase OPEVAL usually is the most rigorous phase of initial operational test and evaluation, planned and conducted by COMOPTEVFOR using fleet personnel at sea in threat-oriented scenarios. The test results, properly evaluated, can affect not only the future of the subsystem, but also the ship for which the system is intended. The Commander, Operational Test and Evaluation Force's report of an OPEVAL to the Chief of Naval Operations includes an assessment of the weapon system's operational effectiveness and operational suitability, together with an approval for service use recommendation. If a subsystem represents a significant capability within the ship, its OPEVAL outcome could become an issue in the ship production decision.

Land Based Testing

Given that ship design and individual subsystems have been properly evaluated, the next step in ship acquisition initial operational test and evaluation is to assess the amalgamation of individual equipments into a higher level combat system or propulsion system at a land based test site. Land based test sites are not the rule. When complexity or technical advancement warrants, a LBTS is constructed. The site is to duplicate and simulate as many elements of the selected subsystem's planned installation as practicable, and necessary, to reduce acquisition program risk before Milestone IIIA.

As an example of how a land based test site might be used, consider the propulsion system for a new warship class. The propulsion system test site ashore is usually incomplete in that it lacks the ship's hull and propeller. What usually is installed is the power train, from prime mover through propeller shafting, and the control system. A water brake or other device serves as a load on the system to simulate ocean action on the ship and propeller. With these limitations to a performance evaluation, initial operational test and evaluation leans heavily toward assessing the adequacy of component integration and evaluation of operational suitability issues such as reliability, maintainability, and compatibility, where early problem detection can result in significant cost savings. While effectiveness issues do exist in the prime mover and propulsion train, these issues are for the most part technical in nature, such as horsepower developed or fuel consumption rates. The real operational effectiveness issue for new propulsion systems is how well the control system performs when it is required to orchestrate other propulsion system elements during repeated maneuvers matched to the operating profile of a warship at sea.

Although not present at the land based test site, the propeller will, when appropriate, undergo separate at-sea operational test and evaluation on a surrogate ship. To support use of gas turbine propulsion in the DD 963 and FFG 7 classes, it was necessary for the Navy to develop a new propeller capable of transmitting thrust in the 40,000 shaft horsepower range, yet capable of reversing pitch (gas turbines are unidirectional). Two competing designs were put to sea in USS BARBEY (FF 1088) and USS PATTERSON (FF 1061) for combined *development* and *operational* test and evaluation. The failures that occurred during this combined test and evaluation led to additional research and development in propeller design and control. Further engineering ultimately permitted the introduction of gas turbine propulsion to the fleet.[2]

Despite propulsion train testing at the land based test site and propeller testing at sea, several parts of the propulsion/mobility system are neither integrated nor tested prior to construction of the lead ship. The land based test site may not include, or exactly duplicate, auxiliary and support systems directly related to the propulsion train as it will be installed in the ship. These support systems might include: lube oil, feed water, electrical, and compressed air. Such systems are not easily defined or bounded and often do not go through the standard operational test and evaluation/approval for service use process. These are systems of the "as required" type, usually contractor-furnished, and unique to each ship class. While these systems escape initial operational test and evaluation at the land based test site, they cannot be overlooked. Failure of any of these systems can result in loss of warship mobility. The Navy requires that these systems meet the objectives of the approval for service use process, even though furnished by the shipbuilder as an integral part of the shipbuilding contract.[3]

Department of Defense Directive 5000.3 is specific about the prototyping of new propulsion systems, although a specific statement is not included that requires a propulsion system land-based test site. The Navy addresses this need in Navy directives.[4] Recent experience in the NATO patrol missile hydrofoil program, which had both a propulsion land based test site and a prototype ship, suggests that where an RDT&E,N-funded* prototype ship exists, it may be more cost-effective to use the prototype ship for propulsion system testing rather than a land based test site. In the patrol missile hydrofoil case, all operational test and evaluation data were collected during prototype testing at sea in 1976. Operational test and evaluation was not conducted at the land based test site.[5]

The combat system test site ashore differs significantly from the propulsion land based test site. While also lacking the ship and ocean environment, it can be used for operational performance evaluation through the use of computer simulation methods. Installed at the combat system land based test site are a combination of warship equipments and simulators. These items represent the ship's sensor, weapon, and information processing subsystems integrated through interfaces, e.g., digital computer techniques. The site must include a physical and functional replication of the ship's combat information center if it is to be of significant value for initial operational test and evaluation. Additionally, there will be some type of test control center from which the COMOPTEVFOR operational test director and his evaluation team can stimulate the ship's combat information center with a wartime scenario by introducing targets, communications, and bits of intelligence to which the combat system must react. Thus much of the combat system capability can be evaluated including equipment, computer programs, and manning.

The goal of combat system land-based operational test and evaluation is an evaluation of the intended ship's man/machine interface. Combat information center is simply the location where the man/machine interface takes place; and "combat system integration"—which is tested at the site—is simply the means by which it is done. This man/machine interface is the central focus of initial operational test and evaluation at the combat system land-based test site. On the "man" side of the equation, the goal of the test site is achieved by:

- deciding upon the stationing of personnel;
- deciding upon radio/interior communications/voice communication patterns;
- developing a "combat system doctrine"; and
- training.

The "man" side of the equation can be adjusted after ship construction.

On the "machine" side of the equation, the goal is achieved by numbers, location, capability, and interconnection of equipments. The "machine" side (whether equipment or computer programs) is relatively expensive to change after ship construction. Thus a goal for operational testing at the combat system test site is to bring the "machine" side of the interface to an adequate capability level before ship construction. In this way subsequent post-construction fine-tuning can be done on the "man" side of the equation.

Initial operational test and evaluation at the FFG 7 combat system land based test site resulted in major equipment and configuration changes in the combat information center to make the design operationally effective for one of the ship's major mission areas.

The combat system land based test site is not without limitations. Some key system elements cannot function at a shore site. Sonars, underwater listening devices, towed arrays, missile, gun, torpedo, and chaff firings must be simulated. Likewise, surface search radars, active electronic countermeasures, navigation systems, etc., cannot or are not allowed to be operated ashore in the same manner as at sea. Outside of the combat information center and other replicated operating spaces, the support

*Research Development Test and Evaluation money appropriated for Navy use.

systems, such as electrical power, cooling water, and compressed gas systems, usually bear little resemblance to the intended ship configuration.

The emphasis here on the shortcomings of land-based testing is not intended to degrade the importance of these sites. Land based sites serve as a valuable engineering tool in the early reduction of risks, technical and operational. The intent is to point out the need for operational testing at sea. The combat system and propulsion system are not two independent blocks that can be stacked one above the other on a hull without further integration, and then be put to sea as a warship. Ship integration, propulsion or combat system, is not accomplished by summing the individual elements. *The elements are not independent.* The elements can complement each other and they can interfere with each other. At sea, system integration with the hull is stressed in operational testing and attempts are made to measure system performance, at least in qualitative terms. This ship integration is something that cannot be reproduced ashore at a test site.

The actual warship capability may be more or less than the sum of the capabilities of all the elements. We do look for something extra as a result of the integration. Perhaps it is a faster reaction time or fineness of control, but whatever we seek we find that a major portion of the integration effort cannot be accomplished during the ship design phase and must, of necessity, be undertaken concurrent with the lead ship's construction and fleet introduction. To make this apparent concurrence in design and acquisition work productively, the Ship Acquisition Program Manager must include initial operational test and evaluation as an important tool in pacing program progress.

A complete understanding of the purpose of any test site is essential. The land-based test site must allow not only the design and engineering effort, but also the test and evaluation to demonstrate that the integration has been satisfactorily achieved. Inherent in the capability to simulate portions of the system is the pitfall of unrealistic simulation. Accordingly, results of land-based test site initial operational test and evaluation, and particularly combat system initial operational test and evaluation, are heavily dependent on the validity with which the behavior of the missing equipment, ship, and its external environment (threat and natural) are replicated and perceived by operators participating in the tests.

A given combat system land based test site may provide adequate simulation for the designer and engineer to demonstrate that the ship's combat system functions as specified, in a technically acceptable manner, and that the known-unknown risks have been reduced. This is usually demonstrated by playing a preprogrammed scenario to stimulate the combat information center. The ship's crew responds to the air, surface, or subsurface threats. The threats are presented so that tactical decision-making is simplistic and sequential. Own ship combat information center is, normally, the victor. Engagements are not constrained by realistic limitations in the ship's weapon load as to type and quantity, fuel remaining on board, heavy weather, radio communications and intelligence failures, etc. Operator frailties, particularly under stress, are discounted as notests. If meaningful initial operational test and evaluation is to be conducted at this combat system land based test site, additional expenditures are necessary to simulate an enemy that fights back, i.e., that can be maneuvered in response to own ship combat information center actions.

Operational test and evaluation is concerned with the interoperability of the ship combat information center during command, control, and communications interfaces with other ships and aircraft in the simulated task force. Tests highlight the ship's combat information center arrangement, equipment, operator tools, and interconnections—all things that can be changed before the ship is actually constructed. These items are costly to change once installed. What is *not* of concern in this operational testing is who wins or loses the simulated combat engagement because that was addressed before the extensive Navy commitment to a test site and a shipbuilding program. The question to be answered by this operational testing is what must be modified to increase the warship's operational effectiveness and operational suitability? From operational test and evaluation should come identification of some unknown-unknowns, and reduction of operational risk.

All types of initial operational test and evaluation have had as an objective the demonstration of program objective achievement by Milestone IIIA. At this point the key decision is made to commit dollars to follow-ship production, that is, award of construction contract. Test and evaluation, both operational and technical, must demonstrate system capabilities sufficient to give the decisionmaker a clear understanding of the risk involved. It is not necessary that all risk be eliminated. It is necessary that risk be reduced to an acceptable and manageable level, given the planned shipbuilding schedule; and where risk still remains, that there exist a firm plan for its resolution, to include a fallback position should resolution fail.

FOLLOW-ON TEST AND EVALUATION

The culmination of initial operational test and evaluation at the land based test site and subsystem OPEVALS occurs immediately before the Milestone IIIA decision. Deficiencies are identified, and fixes are developed. The ship contracting process can proceed after Milestone IIIA while land based test site and OPEVAL deficiencies undergo an interactive process of fix and retest until equipment and systems are operationally effective and suitable to receive unconditional approval for service use.

After the ship Milestone IIIA decision and contract award for the first follow-ship, *at least 2 years remain* before the lead ship is delivered and put to sea. This time is not lost, because implicit in Navy ship acquisition program structure and supporting test and evaluation policy is the knowledge that engineering and testing will continue. It is partially in recognition of this planned test and evaluation for risk reduction, as documented in the TEMP, that the follow-ship decision can be made well before the lead ship is launched.

The long design and construction time for a conventional warship permits the Navy to capitalize on the opportunity to make changes in equipment design, operating compartment arrangement, and computer program configuration, as well as adjustments to manning. The changes and adjustments are further tested at sea on surrogate ships and ashore at the land based test site. The Navy's newest class of frigates (FFG 7), for example, passed Milestone IIIA in late 1975. Based on initial operational test and evaluation results, the combat information center was almost completely rearranged, new equipment added, and computer programs modified. Eighteen months later, this redesigned combat system was subjected to follow-on operational test and evaluation at the land based test site by COMOPTEVFOR, *several months before the lead ship* was scheduled to be delivered. Similar follow-on operational test and evaluation was conducted at the propulsion system land based test site.

Delivery of the lead ship by the builder is followed by rigorous Acceptance Trials by the President, Board of Inspection and Survey. This is followed by extensive tests, trials, and training under direction of the Chief of Naval Material and the Fleet Commander. Next, selected ship subsystems and equipments undergo further operational test and evaluation to verify correction of remaining OPEVAL deficiencies, complete the ASU process where necessary, and complete system integration. For both the propulsion and combat systems, the lead ship is really the first time that everything is integrated completely, placed in the at-sea environment with a trained crew, and tested in mission-and-threat-representative scenarios with live weapon firings.

The usual fleet introduction period for a new warship is about 11 months, during which time tests, trials, training, and a shipyard availability occur. Squeezed into this period are 3 to 6 weeks of follow-on test and evaluation by COMOPTEVFOR on propulsion and combat system elements (where necessary to complete the approval for service use process), and on the integration of these elements. Lead-ship test and evaluation concentrates on the remaining operational effectiveness and operational suitability issues.

WHOLE-SHIP OPERATIONAL TEST AND EVALUATION

In addition to the continuing phases of operational test and evaluation on the lead ship of a new class at sea, the operational test and evaluation policy of DOD Directive 5000.3 includes a phrase that allows "...*if required*, full ship operational evaluation to the degree feasible." *This is a new concept to the Navy.* Notice that this policy permits the decisionmaker considerable latitude for defining need and scope of "operational test and evaluation of a ship as a whole."

The Navy has studied this new concept in great detail for two ship classes, USS VIRGINIA (CGN 38) and USS OLIVER H. PERRY (FFG 7), both typical displacement-hull warships. *The concept developed by* COMOPTEVFOR was to evaluate a *ship*, not specific equipments. The individual equipments in the ship would have been evaluated previously, singly and as elements of an integrated combat or propulsion system.

NOTE

The objective of doing operational test and evaluation on an entire ship is to evaluate *how well the ship performs her missions*, not how well a particular system works. Since a ship's mission necessitates accomplishment of a wide variety of tasks *simultaneously* in a combat environment, the operational test and evaluation must take place with the ship in an operational environment.

The approach to whole-ship operational test and evaluation in a lead-ship, follow-ship acquisition program in this concept is two-phased. *Phase I* is the *preliminary* phase, to provide an understanding of the ship's performance with normal manning, under fleet control, in an operational environment. Phase I would be conducted sequentially through a series of tests and evaluations to determine:

- **Individual system technical performance.**
- **Technical capability of systems in integrated operation.**
- **System operational effectiveness afloat.**
- **System operational suitability afloat.**

***Phase II*, the *operational* phase, would determine whole-ship operational effectiveness and operational suitability in a threat environment. This phase would ensure that:**

- **All significant aspects of the ship's mission performance are evaluated.**
- **Any potential problem areas (e.g., systems interfaces) have been identified, and the problem impact on mission performance has been assessed.**

***To avoid duplication of testing* between phases, every effort would be made to conduct Phase I tests on the lead ship under conditions such that repetitive Phase II tests could be avoided. Phase II would cover only that essential test and evaluation not previously accomplished.**

The preceding discussion of the concept deals with a *short-range* evaluation of the ship, i.e., how effective is she with current sensors, weapons, and computer programs? There is another important aspect of a new warship, and that is her *long-range* effectiveness. How well will the ship do over a 30-year life? This type of evaluation involves growth potential, ease of update, modularity, space and weight reservations, excess power and water, excess crew space, etc., as

well as durability and maintainability of the basic ship systems. While Phase II operational testing would concentrate on short-range aspects of the ship, COMOPTEVFOR would provide both a short-range and a long-range assessment.

As a result of Milestone II for the FFG 7 (formerly Patrol Frigate PF) program in 1972, the Deputy Secretary of Defense stated the general purpose of whole-ship operational test and evaluation in the decision memorandum for Milestone II.

> "...Also, it may be desirable that a period for operational test and evaluation of the lead ship, prior to that ship's full release to normal Fleet usage, be allocated to OPTEVFOR. The purpose of this testing would be to determine the PF's expected operational effectiveness in its expected roles and the need for any early modification to follow ships. Should such modifications be required, a later DSARC would have to determine the relative merits of opening existing contracts to change by change order procedures or making modifications after acceptance from the shipbuilder."[6]

In simple terms, the reasons for doing whole-ship operational test and evaluation are:

1. to find out, at the earliest possible time, what the ship capabilities in mission terms really are,

2. to discover any weaknesses the ship may have—again, as soon as possible—so we can either correct them or learn how to live with them,

3. to confirm correction of discrepancies noted on previous trials/tests,

4. to identify key logistics considerations (e.g., repair parts, maintenance support, training) which should be modified or established as soon as possible to ensure that the fleet can operate, maintain, and support the ship,

5. to provide early feedback to OPNAV/NAVMAT, to assist in work on future ships and systems, and

6. to develop initial procedures/tactics that specify how to best employ the ship.

These early insights on tactical employment are a by-product of the operational appraisal—turned over to Fleet Commanders with the ship, for further development.

The usual reason for conducting an OPEVAL—to verify that the system is ready for production—does *not* apply in the case of a conventional, displacement-hull ship. By the time the lead ship has completed an operational appraisal, all other ships of the class should have been contracted for and several will be well along in construction.

The above discussion has addressed whole-ship operational test and evaluation using a typical displacement-hull warship, after all lower levels of test and evaluation are complete—a process that may take many months, even years. Contrast this with the prototype ship acquisition program where a ship OPEVAL is conducted *before* Milestone III.

The decision for both acquisition programs studied to date by the Navy (USS VIRGINIA and USS OLIVER H. PERRY) was not to require operational evaluation of the lead ship (whole-ship operational test and evaluation). Given the policy in DOD Directive 5000.3, it can be expected that the requirement for whole-ship operational test and evaluation in future ship acquisition programs will be addressed on a case-by-case basis.

Why was the decision made not to require whole ship operational test and evaluation on the CGN 38 and FFG 7 classes? Certainly cost was an important factor. In both cases studied, the minimum out-of-pocket dollar cost to the Navy to do whole ship operational test and evaluation was estimated to be less than one half of one percent of total program cost, depending on costing assumptions. However, if even this modern cost becomes an unfunded requirement the Ship Acquisition Program Manager has another significant management problem to solve! Two other factors probably had greater influence on the decision, the comprehensive development test and evaluation and operational test and evaluation plans outlined in the CGN 38 and FFG 7 TEMPs (and currently being executed), and the fact that the result of whole-ship operational test and evaluation, if it were conducted, probably would not have significant impact on changing the many ships of the class already contracted for without incurring exorbitant cost and schedule penalties. In addition, the many weeks that a ship is dedicated to whole-ship testing, it is not available to the Fleet Commander for filling forward deployment commitments.

SUMMARY

Ship test and evaluation policy recognizes that *ships are different* from most other weapon system acquisition—the difference being that the long design, engineering, and construction period for a major warship will normally preclude testing the lead-ship in a class prior to the decision to proceed with follow-ships. The purpose of ship test and evaluation, at least for conventional, displacement-hull warships, it to *reduce risks rapidly*, and minimize the need for modification to follow-ships before deployment. The policy presumes a commitment to the ship acquisition program, and that Milestone IIIA will occur, once sufficient performance has been demonstrated to reduce technical and operational risks to a manageable level. For ship combat systems, the policy is detailed, considered, and flexible.

The thrust of the policy for *new ship types*, those incorporating major technical advancements not earlier proved in hull and nonnuclear propulsion design, appears to focus test and evaluation on technical advancement. This policy could be interpreted to mean that an advanced technology, high-speed prototype platform (not a warship) capable of zipping about the ocean's surface, but lacking a warship's combat system, is adequate to address all the *operational risks* before Milestone III. It is not. The final sentence in the directive on the subject of ship propulsion system innovations clearly implies that the policy is concerned with acquisition of warships, not just high performance platforms:

> "Adequate test and evaluation on such prototype [ship] will be completed prior to the first major production decision on follow ships."[1]

In addition to directing a program of early, at-sea follow-on operational test and evaluation in the lead ship for specified systems and equipments on all new ship classes, the policy goes one step further in regard to conventional, displacement-hull warships. The whole-ship operational test and evaluation concept "...if required, full ship operational evaluation to the degree feasible..." seems incongruous with the policy premise of rapidly reducing risks so as to minimize the need to modify follow-ships in a class, at least until we learn how to shrink the many years needed to design, engineer, and construct the lead warship.

Almost certainly, additional knowledge about the new warship's capability to serve the Navy mission in the long-term and short-term could be generated by carrying out the whole-ship operational test and evaluation. Perhaps the few remaining operational unknown-unknowns might be gleaned, but these should have been ferreted out through vigorous operational test and evaluation on the combat and propulsion systems in the normal lead-ship operational test and evaluation shortly after delivery. What probably cannot be obtained by COMOPTEVFOR without the rigor of a whole-ship operational evaluation are solid answers to long-term evaluation questions, or warship performance in a vast range of fleet/threat scenarios and ocean environments. Given the 10 or more years from initiation of a major warship acquisition program until results of a whole-ship operational test and evaluation are available to the Chief of Naval Operations, the value of the policy as a useful management tool for rapid risk reduction during acquisition seems questionable. The resources expended in such a whole-ship operational evaluation might be better spent in careful engineering and test and evaluation before Milestone IIIA.

PRACTICE

When the new ship test and evaluation policy was published in early 1973, the Navy had already structured two major ship acquisition initiatives to comply fully with this policy.* The patrol missile hydrofoil program was to use a prototype ship (PHM 1) for operational evaluation before the first major production decision at Milestone III. The FFG 7 program had both a propulsion system and a combat system land-based test site under construction, as well as extensive planning underway for component equipment and subsystem OPEVALS on surrogate ships. Timing of the operational test and evaluation for over nine major systems and their integration was orchestrated to ensure that comprehensive evaluations would be available to the Chief of Naval Operations at the critical Milestone IIIA decision for the follow-ships.

Even though the production PHM and FFG 7 class warships are still years away from deployed fleet service, it is fair to say that operational test and evaluation has shown a return on the investment. Final judgment on the value of initial operational test and evaluation conducted as part of the PHM and FFG 7 programs cannot be rendered until the ships have been through several years of fleet service. Major warship acquisition programs initiated after PHM and FFG 7 programs are structured using a similar demonstrated-performance/milestone strategy.

Whereas the FFG 7 program faced a single critical test and evaluation milestone, the new AEGIS ship acquisition programs are scheduled for a sequence of four significant engineering/test events (Operational Test IIIA, B, C, and D), starting with basic weapon testing at sea and culminating in operational test and evaluation of the integrated combat system at the land based test site before the full ship production release milestone. Follow-on operational test and evaluation on the class lead-ship should complete the operational test and evaluation process.

Prior to the "try before buy" policy, a Ship Acquisition Program Manager's performance and motivation had been measured by the ship he delivered, *after* he delivered it. Typical management performance measures would be unit cost for each ship constructed; whether scheduled contract delivery dates were met; and, how many design and production deficiency items were included in the report by the President, Board of Inspection and Survey, after a brief Acceptance Trial at the time of ship delivery (years after the design decisions). An operationally-oriented, independent evaluation of the ship's potential effectiveness, based on demonstrated performance to date, was not available to senior Navy decisionmakers at intermediate points in the acquisition program. Operational risks identified in the ship design and engineering process were addressed through a few formal boards and official correspondence by the Chief of Naval Operations, the Naval Material Command, and the fleet and type commanders when they were not overly committed with their primary mission of fleet operation, readiness training, and maintenance.

*PHM class and FFG 7 class ship acquisition programs were structured under the new test and evaluation policy. Five ship class acquisition programs (CVN 68, SSN 688, DD 963, LHA 1, and CGN 38), all past Milestone IIIA, were designated by the Chief of Naval Operations to comply with the new test and evaluation policy contained in DOD Directive 5000.3 for ships, with minimum cost and schedule impact. Compliance consisted of "back-fitting" the new policy to include follow-on test and evaluation where a need existed, e.g., operational risk, complete approval for service use process, demonstrated operational effectiveness and operational suitability of integrated combat or propulsion systems. Five very different follow-on test and evaluation programs resulted, ranging in scope from a few COMOPTEVFOR officers witnessing the lead ship's Acceptance Trials and shakedown training evolutions, to minioperational evaluations of new systems and extensive combat system integration tests.

Today, the operational effectiveness, operational suitability, need for change, and preliminary tactics for new ship classes are the direct concern of COMOPTEVFOR throughout the ship design, engineering, and fleet introduction phases of the acquisition process. More importantly, the COMOPTEVFOR independent assessment of operational performance is available to the Chief of Naval Operations when program milestones dictate. The actual planning, conducting, and reporting operational test and evaluation is the function of an operational test director and his team of officers and senior petty officers who address the adequacy of demonstrated warship performance. The ship operational test director is responsible for drafting the operational test and evaluation plan for integration with development test and evaluation and production acceptance test and evaluation in the TEMP, the Chief of Naval Operations-approved "contract" negotiated between COMOPTEVFOR and Ship Acquisition Manager. Operational test and evaluation is a planned element in the structure of ship acquisition programs.

For the record, FFG 7 class passed Milestone IIIA in 1975 without the need for a formal meeting of the Defense Systems Acquisition Review Council because the Navy had done its homework in operational test and evaluation. Production of the NATO Patrol Missile Hydrofoil class was authorized in 1977. When PHM 1 entered the fleet in the same year, the President, Board of Inspection and Survey reported to the Chief of Naval Operations that PEGASUS, (PHM I), was one of the best, if not the best trial the Board had completed.

"...THE BOARD RECOGNIZES THAT THE SHIPS HAD BEEN IN SERVICE FOR SOME TIME BEFORE PRESENTATION, HOWEVER THE EXCELLENT ENGINEERING OF HER PROPULSION SYSTEM DEMONSTRATED SUPERB RELIABILITY THROUGH HER TRIAL WITH NO MAJOR OR SIGNIFICANT BREAKDOWNS OR FAILURES. THE COMBAT SYSTEM ALSO PERFORMED AS DESIGNED AND ADVERTISED. THE SHIP, AS DESIGNED AND SUPPORTED BY EITHER A MOTHER SHIP OR ADVANCED BASE WITH ITS LOW PROPULSION SIGNATURE, HIGH SPEED AND MANEUVERABILITY, ARMED WITH A HARPOON MISSILE SYSTEM, HAS A FORMIDABLE CAPABILITY FOR SURVEILLANCE AND MONITORING NARROW SEA PASSAGES AND ANTI-SHIP ACTION WHERE A LARGER UNIT'S CAPABILITIES MIGHT BE INAPPROPRIATE OR WASTED..."[7]

It is the application of operational test and evaluation and the involvement of COMOPTEVFOR in ship acquisition programs *before* and *after* Milestone III that is new in the Navy response to DOD Directive 5000.3.

REFERENCES

1. Dept of Defense, "Test and Evaluation," Directive, Number 5000.3, 19 Jan 73, p 5.

2. Dept of Navy, COMOPTEVFOR to CNO (serial 912), Ltr, subject: "Second Partial Report on CNO Project 179–2-OT-III, Partial Operational Test and Evaluation of the FFG-7 Propulsion System," 3 Sep 76.

3. ________, "Approval of Systems and Equipments for Service Use," OPNAVINST 4720.9D: Para 12, "Special Considerations," 23 Aug 74, p 11.

4. ________, "Land Based Test Site (LBTS) Selection Policy," NAVMATINST 3960.8, 1 Sep 76.

5. Cdr Karl M. Duff, USN, LCDR Henry Schmidt, Jr., USN, and LCDR Michael R. Terry, USN, "The NATO Patrol Missile Hydrofoil (PHM) Ship and Weapons Systems Technical Evaluation Program," Technical Paper, Preprint No. 76–848, presented at Third Advanced Marine Vehicles Conference sponsored by American Institute of Aeronautics and Astronautics, Inc., and Society of Naval Architects and Marine Engineers, Arlington, VA, 20–22 Sep 76.

6. Dept of Navy, COMOPTEVFOR to CNO (OP-03), Ltr, subject: "Operational Appraisal Plan for FFG-7," 30 Oct 75.

7. ________, President, Board of Inspection and Survey (PRESINSURV) to Chief of Naval Operations, message, (131748Z) Jun 77.

Commander Ian E. M. Donovan is assigned to the Ship Evaluation Division at COMOPTEVFOR as the Operational Test Coordinator for USS VIRGINIA (CGN 38) ship class and the new AEGIS cruiser class (CGN 42). Sea experience of Commander Donovan includes deck, gunnery, engineering and operations duties in USS CONE (DD 866), USS SELLSTROM (DER 255), USS FORRESTAL (CVA 59), and USS STICKELL (DD 888). He also served in USS BIDDLE (CG 34) and USS SELLSTROM as Executive Officer. For 13 months in Vietnam he was Force Plans Officer on the staff of COMNAVFORV/CHNAVADVGRP MACV.

Commander Donovan is a graduate of the DESLANT Engineering School, Defense Intelligence School, and School of Naval Command and Staff, Newport, RI. While on the staff at the US Naval War College, Cdr Donovan was Director, International Law Study, a Research Program Officer, and Director, Senior Officer Executive Management Course. At the NROTC Unit, University of Louisville, KY, he instructed midshipmen in naval engineering and management.

Commander Donovan received his commission through the Regular NROTC program when he graduated from The Ohio State University with a BS in Business Administration (1955). He received his MS degree, International Affairs, from The George Washington University.

Lieutenant Commander Thomas A. Fitzgibbons is Engineer Officer, USS TRIPOLI (LPH 2). He recently completed a 2-year tour at COMOPTEVFOR where he was the Operational Test Director for DD 963 class destroyers. Prior assignments include service as Chief Engineer with the commissioning crew, USS FANNING (FF 1076) and a tour of duty in Vietnam as a Riverine Warfare Advisor. In his first tour of duty he served as Main Propulsion Assistant and Damage Control Assistant in USS SOUTHERLAND (DD 743).

Lieutenant Commander Fitzgibbons graduated from the US Naval Academy, class of 1966. He later completed Destroyer School (Department Head Course). LCDR Fitzgibbons received a masters degree in Weapon Systems Acquisition Management from the Naval Postgraduate School.

DEFENSE PROCUREMENT:

THE BRITISH CONNECTION

by

Captain T. H. Sherman, United States Navy

For many years the United States has encouraged standardization of weapons systems among our Allies, principally within the North Atlantic Treaty Organization (NATO). However, an examination of the weapons system inventory of the United States reveals few foreign systems. Conversely, US equipment frequently appears in foreign inventories. The situation has fostered the opinion, held by many abroad, that the United States supports standardization only so long as it applies to purchases of US defense equipment. While US Allies acknowledge that the US Services have been limited in purchasing equipments abroad by the "Buy American Act," (47 Stat. 1530; 41 U.S. Code) the perception is that the US Armed Services have a reluctance to purchase any systems "not invented here."

A MEMORANDUM OF UNDERSTANDING

Standardization within NATO has been a goal for at least three recent Secretaries of Defense. Secretary Schlesinger made many pronouncements of this goal and began negotiations with the United Kingdom. In September 1975, Secretary Schlesinger concluded a Memorandum of Understanding (MOU) with his British counterpart regarding cooperation in reciprocal defense procurement. The primary purpose of the agreement is to promote greater United States—United Kingdom (UK) cooperation in Research and Development production and procurement to enhance NATO rationalization and standardization, and to achieve the greatest NATO capability at the lowest possible cost. With certain exceptions, the MOU permits the industry of each country to compete in the defense market of the other on an equal basis. An ultimate result of the agreement will be a long term equitable balance in defense trade. As a government department head, with authority delegated by Congress, the Secretary of Defense has exempted defense items from the restrictions of the Buy American Legislation for the purposes of this agreement. The opening of this two-way street with Britain implies a commitment by the United States to allow foreign suppliers to bid in the United States defense market on equal footing with United States industry. Procurement from other European Allies with whom the US does not have a similar MOU is handled on a case-by-case basis. In these cases when NATO standardization will be enhanced, the US may waive the "Buy American" price differentials.

The MOU is broad in scope and describes the principles to be followed by the United States and Great Britain in reciprocal procurement. Included is a provision for the development of test and evaluation procedures by the Office of the Secretary of Defense (OSD) in cooperation with the United Kingdom. The Director of Defense Research and Engineering is responsible for research, development, test and evaluation related to bilateral arrangements for joint military development. Subsequent implementing instructions to the Services address the issue of test and evaluation (T&E) in providing for the consideration of UK items which have been tested and accepted by the UK for Service use. An appraisal of the sufficiency of UK testing for US purposes is to be on a case-by-case basis, and, where retesting is necessary, United States test and evaluation standards, policies, and procedures will apply. The negotiations concerning the test and evaluation aspects of the MOU are currently under discussion and questions concerning these negotiations have not been fully resolved.

The question of acceptability for Service use is a most important aspect of the US/UK memorandum and the implementing instructions. For example, United States' Services have definite test and evaluation disciplines to which new hardware is subjected. Included are development and operational testing, both of which ultimately lead to a judgment on item acceptability for Service use. One asks the question, "How closely do UK test and evaluation disciplines parallel those of the US? Is it possible for the United States to relate the various phases of UK testing to US standards? If so, should an equipment or system which has matured in the UK, and which has been in the field with a UK Service, be subjected to the full gamut of US testing or can the process be abbreviated?" Recently there have been examples of mature UK systems having been disqualified by US procurement activities based on inadequate testing (in US eyes) when in fact extensive testing and Service use had been accomplished in the United Kingdom. Most probably this disqualification occurred as the result of ignorance of British testing and its applicability to the US procurement process. How to reconcile these inconsistencies and implement the MOU in a timely manner is a current task of the Deputy Director, Defense Research and Engineering (Test and Evaluation). To determine the range of this effort an examination of the DOD test philosophy and definition as stated in DOD Instruction 5000.3 is in order.

A new system under development undergoes many tests, from subsystem to system, either in a laboratory environment or in the field. This Development Test and Evaluation (DT&E) is an integral part of advanced and engineering development. The developer utilizes this testing to validate design, sound out the envelope of capability and demonstrate reliability growth. At some point the developer and the contractor reach a level of confidence that the system can be operated and maintained by US soldiers, sailors, or airmen (without outside help) and is ready for the field. At that time, the Materiel Command certifies that the system is ready for a military operational determination of system acceptability for Service use.

With development certification achieved, the system is passed to the cognizant Service's independent test agency for Operational Test and Evaluation (OT&E). These tests are planned and conducted by the agency, although the actual evaluations may be performed by representative field units under cognizant agency control. The tests are made without contractor support as the field unit evaluates the system for the Service Chief. Test conditions are selected to be as realistic as possible, including stress environment such as would be experienced in wartime. Thus, the developing agency and the contractor are taken out of the loop and the system undergoes a graduation exercise conducted by a stern judge. Note that while this judge is in uniform, he is not serving as the ultimate user, the troops in the field. This independence from both developer and user in testing and reporting is most important. Independent test reports concerning the operational effectiveness and suitability of the system go directly to the respective Service Chief of Staff and do not imply the urgency to field the hardware that the user might feel nor do the reports reflect the advocacy of the developer. In short, the independent tester is the referee between the developer and user.

The test and evaluation procedures followed by the Services involve extensive resources of money, time and personnel. As experience is gained, more efficient means of accomplishing test and evaluation goals are sought and implemented. For example, while the independent tester has been, in the past, relatively noninvolved in developmental testing, valuable resources can be saved if the independent tester participates in and keeps well informed of the developer's testing. In this manner, redundant testing is avoided during Operational Test and Evaluation and the developer can better prepare the system for the operational tests that will be imposed.

Finally, if testing has been successful and other requirements of need, resources and risk reduction have been met, the system may be released for production. Even then, testing may continue after production to proof design changes and backfits required to correct previous test deficiencies or, when appropriate, to contribute to modernization of fielded equipments.

Discussions with the Ministry of Defence were begun in Fall 1976 to determine the applicability of United Kingdom testing to United States procurement policies. These discussions were informal and were intended to review the problem and to outline a course of action for the future. First, it was agreed that negotiations would be concerned only with complete defense systems as opposed to a full shopping list of subsystems and components. While United Kingdom defense production quantities are much less than those of the United States, it should be possible to relate the US concept of a "major system" to that of the United Kingdom. Second, it was apparent that each side should become familiar with the test and evaluation policies and procedures of the other in some detail. Relevant Office of the Secretary of Defense/Ministry of Defence (OSD/MOD) documentation was exchanged for further study. Reciprocal visits to test and evaluation agencies were proposed to facilitate agreement on terminology, procedures, equivalence, and perhaps a Test and Evaluation annex to the MOU.

DIFFERENCES

On the US side it was apparent from the documentation that two basic differences in test and evaluation philosophy set the United States and the United Kingdom apart, aside from the expected differences in nomenclature.

1. The UK does not have a counterpart to the US independent Service test agencies (Commander, Operational Test and Evaluation Force (COMOPTEVFOR), Operational Test and Evaluation Agency (OTEA), and Air Force Test and Evaluation Center (AFTEC)). All testing is the responsibility of the developer. The United States policy stresses the independence of the operational tester from the developer and user. The British Army more nearly approaches the US concept of the independent tester by utilization of special boards that observe and report on trials.

2. Most operational testing in the UK Services (by US definition) is performed *after* a decision is made to go into production.

To place these differences in proper perspective, note that in Britain the procurement process is extremely complex and demanding of a Program Manager. Procurement milestones are more numerous and often more stringent than those imposed by the US Department of Defense on the three US Services. Further, British defense resources are severely limited in both people and money. For these reasons the British feel that the UK cannot afford a separate test and evaluation hierarchy and count on the integrity of the present system to ensure adequate test and evaluation. Since the UK production runs are orders of magnitude smaller than those of the US, fewer test articles are available. Consequently United Kingdom test programs are less ambitious than those of the United States. Operational testing must await a commitment to production to obtain test articles. The British Operational testing is comparable to US Follow-on Test and Evaluation (FOTE). Feedback from FOTE occurs and since the United Kingdom inventory is small, fewer items require backfitting when discrepancies are discovered by testing. The British find this method of defense system test and evaluation to be cost effective and within resources.

The first formal meeting in the negotiations consisted of a visit by a representative of DD(T&E) to Britain for the purpose of seeing first-hand the test and evaluation activities of the United Kingdom Ministry of Defence. The purpose was to determine the similarities and differences between the United States and the United Kingdom systems. On completion of the visit, a meeting was held to explore ways to reconcile existing differences that might hinder reciprocal purchasings. The conclusions reached were:

- **In the United Kingdom, test and evaluation is an integral part of project development.**

- **The United Kingdom project directors/managers have full responsibility for all aspects of test and evaluation during development and for recommending system readiness for Service use. As a consequence, United Kingdom test and evaluation (UK T&E), prior to production release, would be considered in the context of DODD 5000.3 as limited to Development Test and Evaluation.**

- **Operational testing is usually conducted after the production decision is made by the UK Service involved. Particular groups in the three Services are not involved in Operational Test and Evaluation on all projects. Participation in such testing is generally on a case-by-case basis.**

- **Rigorous adherence to common test and evaluation procedures does not exist among the UK Services; in fact, procedures may vary among projects within a single Service.**

Formal presentation of these conclusions was made in a joint paper, agreed upon between the British Ministry of Defence and the United States DD(T&E). The paper has been submitted to higher authority in both countries. The intent is that this joint paper be the basis for a Test and Evaluation Annex to the MOU.

The identification of differences only surfaces potential problem areas. It is necessary that these differences be evaluated in terms of the limitations, if any, that the differences may impose on reciprocal procurement. Solutions must be found to accommodate the intent of the MOU. In certain situations the US Services might be required to adjust United States procedures to accommodate a United Kingdom system being considered. In other situations an adjustment in United Kingdom testing should be made. It is evident that precise rules and procedures cannot be laid down for all systems. Each prospective foreign procurement should be evaluated on a case-by-case basis. These preliminary meetings aired the problem and indicated possible routes by which US/UK test and evaluation activities can be made compatible. For example, some progress was made in regard to test and evaluation reporting. Owing to resource limitations, the United Kingdom feels that a reorganization within the Ministry of Defence to include an independent Test and Evaluation Agency as in the United States is not practicable. Further, the British belief is that individual UK Service test and evaluation should continue in its present form. However, it was agreed that the United Kingdom should establish a single point-of-contact, a project manager, for test and evaluation matters relevant to reciprocal procurement. The project manager will be responsible for all liaison between the United States and the United Kingdom Services for reporting test and evaluation status and results. In the interest of standardization and for the purpose of facilitating this exchange, the United Kingdom will provide test and evaluation data in the format of the existing United States Test and Evaluation Master Plan (the TEMP). In this manner, the United States counterpart should be able to correctly relate United Kingdom testing to United States requirements. At present it is extremely difficult for an American to correlate United States testing requirements with those of the United Kingdom. The TEMP format should provide an "audit trail" for US personnel use in following UK test and evaluation and to evaluate format adequacy for US purposes. An essential part of reciprocal procurement of newly emergent systems appears to be early involvement by both seller and purchaser. For example, the test and evaluation plan for a system under development in Britain may fully satisfy UK test and evaluation requirements but fall short of certain special requirements of the United States. The United States may require further testing in unusual environments of weather, sea state, electronic countermeasures, etc. It may be possible for the United States purchaser to specify additional requirements early enough in the United Kingdom development cycle so that these requirements may be satisfied in conjunction with United Kingdom testing. At least, the United States purchaser will be well aware of those areas where additional United States testing will be required to qualify the system for United States Service use.

It would seem that the more common case regarding cooperation in reciprocal defense procurement and the one easiest to deal with, would be a United Kingdom system in production which is already Service approved. With the test and evaluation data in TEMP format, the United States purchaser can readily determine those areas, if any, that will require additional testing by the United States.

Given that a United Kingdom system meets, or can meet by additional testing, United States requirements for procurement, a final hurdle must be achieved on this side of the Atlantic. The United States must obtain sufficient test articles for the US designated independent service test agency to conduct some degree of Follow-on Operational Testing to establish operational effectiveness and operational suitability, including interoperability with other US systems. Most probably this Follow-on Test and Evaluation will reveal areas of deficiency. Hence a feed-back loop for corrections by the United Kingdom supplier is essential. The extent of this Follow-on Test and Evaluation will vary dependent upon how early in the United Kingdom development cycle the United States service test agency becomes involved. Advice given by the United States service test agency to the United Kingdom developer should make a significant reduction in the need for redundant testing.

As a means of familiarizing United Kingdom counterparts with the unique US concept of independent operational testing, a team of Ministry of Defence officers and civilians was invited to the United States in June 1977. The team members visited and were briefed by each of the US Service independent test agencies. In addition the team members witnessed operational testing in the field performed by each agency. Although limited in scope by current test and evaluation schedules, the team members saw testing of the Army's Light Air Cushion Vehicle (LACV) at Fort Story, Virginia; the Navy's Patrol Frigate propulsion test site at Philadelphia and the Land Based Combat System Test Site at Islip, New York, followed by Air Force testing of the F-16 at Edwards Air Force Base, California.

Based on this experience, the United Kingdom team members felt that much of the United Kingdom follow-on operational testing parallels that of the United States and that in large degree United Kingdom testing should satisfy United States requirements. To this end, the British extended an invitation for a comparable US team to visit the United Kingdom in Autumn 1977 for visits to United Kingdom test activities and meeting with the three Service counterparts. These visits serve the purpose of gaining appreciation of each other's scope of operation and testing philosophy while at the same time establishing individual points-of-contact for future liaison and coordination on test and evaluation matters (as these matters relate to reciprocal procurement).

The negotiations should culminate in the Test and Evaluation Annex to the MOU that will contain jointly agreed upon policies and procedures. The policies and procedures will most likely appear as broad guidelines for each service to apply on a case-by-case basis rather than as specific procedures. In this manner, adjustments can be made to accommodate the wide variance of test and evaluation in the United Kingdom Services and systems.

SUMMARY

Although the idea of military standardization among US Allies has been with us for a long time, the United States has only recently taken the lead to make it a reality. The MOU with Britain should usher more United Kingdom systems into the United States defense market on equal footing with US suppliers and be a forerunner for agreements with other nations. The MOU with Britain will not open a floodgate of candidates for the US market; there are simply not that many systems in question. The MOU will ensure, however, that the "not invented here" philosophy will not prevent an opportunity for United States forces to obtain superior United Kingdom defense products which may be available. Further, the MOU should eventually enable an equitable balance in defense purchasing between the two nations.

Captain Thomas H. Sherman, USN, is a military Staff Assistant to the Undersecretary of Defense Research and Engineering. A native of Baltimore, MD, Captain Sherman enlisted in the United States Navy in 1944 and graduated from the US Naval Academy in 1950. His commands include BARBET (AMS-41), McCLOY (DE-1038), BARRY (DD-933), and the Destroyer Development Group. Shore duty has been principally in the Washington, DC area with the Office of Naval Research; Bureau of Naval Personnel; Office, Chief of Naval Operations; PM-4; and, the Naval Research Laboratory.

Captain Sherman holds a Master of Science degree from the Naval Postgraduate School and has attended the Naval War College and Harvard Business School.

OPTIMIZING SPACE VEHICLE TEST PROGRAMS

by

Dr. George J. Gleghorn
Assistant General Manager for Engineering
Space Systems Division
TRW Defense and Space Systems Group

Spacecraft have an important and significant role in US national defense. Spacecraft now provide surveillance, navigation, weather prediction, and communications capabilities. The testing and evaluation of spacecraft involves difficulties not usually encountered in test programs of other defense systems such as missiles, aircraft, armored vehicles, naval vessels or ordnance.

To date space systems testing has been an expensive activity. The high premium placed on "success the first time" is reflected in high cost. Space flight and operation of remote sensing and communications equipment in space are new undertakings. In a typical program, the spacecraft test and evaluation activities require from 7 months to more than 1 year and absorb about 10 percent of the total program funding for development and acquisition.

CONSTRAINTS ON THE SPACE VEHICLE

Inaccessibility

The outstanding characteristic of past and present day spacecraft is the current lack of a capability to perform maintenance while the spacecraft is in orbit. This factor drives the design in the direction of never fail and away from free access, preventive maintenance, and the ease of repair and replacement concepts found in typical aircraft, computer and weapon systems. Once a spacecraft has lifted off the launch pad, it is no longer accessible for failure diagnosis, maintenance or repair. Even in manned spacecraft there are severe limitations on the capability for performing the visual inspections and manual repair that is taken for granted with most other defense systems.

Not only is repair or maintenance impossible, but also any direct observation, performance evaluation or change in operating mode not designed into the spacecraft is also impossible. The only access to the spacecraft after liftoff is through telemetry and the radio frequency command link. The integrity of radio communications with the spacecraft is of vital importance. The use of this communications link in the ground test program follows from the basic objective of the test program, which is to demonstrate that the spacecraft is capable of operating as desired in space.

Delivery Cost

Each spacecraft requires a delivery system that places the spacecraft into the proper orbit. The cost of the nonrecoverable boost vehicle and the launch support costs are large, that is, never less than from $8 to $10 million and ranging to upwards of $50 million. Hence, delivery of extra spacecraft to replace a spacecraft that has failed in orbit generally represents an excessive economic burden. A lower cost approach is to provide sufficient testing and "burn-in" prior to launch to eliminate all weak components which are prone to early failure. Thus the high delivery cost provides strong motivation for an effective and exhaustive ground test and evaluation program. Despite the fact that use of the space shuttle will reduce the delivery cost, especially for low earth orbit missions, it still remains costly enough to sustain the requirement for an adequate test program.

Space Environment

The unique characteristics of the space environment and the intense vibration, acoustic, and gravitational stresses imposed on the spacecraft during boost have a significant impact on the spacecraft configuration and the verification test program. Equipment must work in a permanent, hard vacuum that has a strong effect on design. One cannot depend on corrective airflow for temperature control. Rather the designs depend upon analysis of heat conduction and radiation paths and careful control of heat loss to space by means of insulation and surface treatment of radiating panels.

Hard vacuum also has its effect on lubrication design. Lubricants must be sealed in, or special vacuum lubricants must be used. Hard vacuum has its effect on radio frequency design. Often transmitters must be able to radiate at all pressures from atmosphere to full vacuum. The potential of corona discharge must be considered in this full regime.

Furthermore, equipment must work in a free fall, zero-G environment. This has benefits as well as disadvantages: For example, a structure for booms and appendages can be made very light as it need not support weight. Conversely dependency cannot be placed on gravity as in ground operations. Example: fuel must be forced through the engines, and mechanisms must be positively activated by springs or motors.

Other environmental factors that severely impact the design of a spacecraft include the charged particle fluxes encountered in the radiation belts and the special environment produced by remote nuclear detonations in the atmosphere or in space. Rapid recovery from transient weapons effects and immunity from permanent damage are required characteristics that can be achieved only through special design techniques. The design characteristics must be validated through ground testing that attempts to simulate the weapons effects.

The natural radiation environment is still being explored by a variety of scientific satellites and space probes. The effects caused by the charged particle streams during geomagnetic substorms have been particularly troublesome to certain satellites in geostationary orbits. Special design precautions are used to prevent surface arc discharges because of the environment encountered during substorms.

UNIQUE CHARACTERISTICS

The severe constraints placed on satellites because of remote positioning, the space environment and the difficulty of getting satellites into space lead to a number of unique characteristics and practices which are not found in combination in any other defense system.

Remote Monitoring and Control

The inaccessibility of the spacecraft requires that all payload generated data and all diagnostic measurements be transmitted to earth for interpretation and evaluation over a suitable radio frequency downlink. In addition an uplink is provided to transmit commands from earth to the spacecraft. Thus every spacecraft contains a Tracking, Telemetry and Command (TT&C) subsystem that also can be used to great advantage for ground checkout of the spacecraft. In typical military spacecraft the TT&C subsystem will accommodate more than 1000 different encoded commands and provide several hundred diagnostic performance measurements and status signals. Although this is adequate to control the spacecraft from the ground under normal unfailed conditions and to diagnose failures for which work arounds have been designed into the spacecraft, it generally is not sufficient to permit acceptance testing of the assembled spacecraft prior to launch. Additional hardline test points and external stimulators must be provided for ground testing.

Component Redundancy

Active or standby component redundancy has been one of the most potent methods for achieving a high probability of mission success after delivery of a satellite into orbit. Electronic components and many mechanical components usually have a similar redundant component aboard that takes over automatically or can be engaged by ground command if the prime component fails. In many instances there is more than one redundant component or there is an alternate method of executing the full mission or a reduced mission. This is how the requirement for "graceful degradation" has been implemented. The ground test program must verify that failure of a component can be detected, that failures are not propagated to unfailed elements, and that the telemetry and command subsystem is capable of providing continuing operation by means of the proper redundant component.

A difficulty in considering such redundant designs and long design lives is that one cannot directly verify the design by test. It is difficult to determine, objectively, how to speed up the testing. Life testing can be done on certain components of the system, but economics and program priorities dictate that an entire satellite cannot be tested for years prior to the time the decision to go into production is made. Thus, another set of extrapolations must be depended upon.

The result is that reliability is designed into the satellite by calculation, by incorporating redundancy, by depending upon piece part reliability and by exhaustive testing of each part. It is common practice to do qualification testing of each manufacturing lot of parts. In addition, extensive burn-in and screening of parts for parameter drift may be required. Expensive as this treatment of the problem is, it is cheaper than is part failure after installation.

The strategy of using redundancy to provide mission reliability for long periods of time, such as 7 to 10 years, has not been used to this extent on other defense systems where the unit cost is low or where maintenance and repair are possible.

Physical Characteristics

The boost vehicle imposes severe weight and volume constraints on the space vehicle it places into orbit. Our largest military boost vehicle, the Titan III, can place about 3200 pounds into geostationary orbit at a cost in excess of $40 million. During boost the spacecraft must fit under the aerodynamic shroud and hence must fit into a cylindrical volume 10 feet in diameter. The cheaper Atlas/Centaur boost vehicle can boost about 4200 pounds into a transfer orbit at a cost of more than $20 million. With an additional propulsion motor that fires at the apogee of the transfer orbit, the Atlas/Centaur provides the capability to put a 2000 pound spacecraft into geostationary orbit.

Designing for these boosters requires that large spacecraft antennas, solar arrays, and other appendages be folded to fit into the booster shroud and be deployed sometime after the shroud is jettisoned and the high-G portion of the boost phase is completed. The basic point is that the test and evaluation program must be designed to verify the validity and reliability of the deployment process under close simulation of the zero-gravity conditions.

STRATEGIES FOR SPACECRAFT TESTING AND EVALUATION

In view of the operating constraints and the unique combination of design features, what then makes sense for test and evaluation of a satellite program? In practice, the program for each satellite development is tailored to the state of development of the hardware involved. Such a program generally consists of four parts: development, qualification, acceptance, and prelaunch validation testing.

A uniform set of definitions and typical requirements for each part of the test program are contained in MIL-STD-1540A, "Test Requirements for Space Vehicles." This standard has been prepared to serve as a guideline document for procuring agencies and spacecraft contractors.

Development testing depends heavily on the type of hardware, parts, and components to be used and the variety and novelty of operational requirements. A development test program may include life tests, design margin tests, failure mode tests, and all sorts of design verification tests. Extensive testing of engineering and developmental models is done to verify concepts and performance within constraints. Of particular importance is the verification of structural integrity, mechanisms, and thermal system design. Wherever possible development tests are performed at the lowest level of assembly where results are meaningful and applicable to the performance of the assembly as part of the spacecraft. Thus antennas, for example, can be tested without being mounted to the spacecraft, solar array designs can be validated by measuring the electrical output of the representative modular sections and by checking structural and thermal performance on a mockup that uses simulated solar cells. Stability of the attitude control subsystem can be analyzed by using a single axis rate table and simulators instead of a complete spacecraft and elaborate 3-axis test equipment. Preliminary thermal system design verification is obtained from an analytical model that relies on measured values of emittance and absorption coefficients and estimates of heat loads. This approach, that includes component testing, mathematical modeling based on key measurements, extensive breadboard testing, and subsystem testing, provides confidence that subsequent qualification testing of a complete spacecraft will not require extensive retest and redesign. The approach saves schedule time because development and test of different subsystems can proceed in parallel.

Life testing is done usually only for those devices which are expected to have a wearout mechanism: bearings, lubrication systems, tape recorders, vacuum tubes (travelling wave tube amplifiers are an example), solar cells, and the like. Often the life testing is accelerated when a life dependence of some parameter is known—but for some devices acceleration is not feasible and dependence must be placed on extrapolating trends from realtime life testing. Such is the case for travelling wave tubes and for bearings exposed to vacuum conditions.

It is a common practice to devote one vehicle to ***qualification testing***. That is where, in the test program, the functional performance and the margin of safety against environmental exposures are demonstrated. Typical exposures are acoustic and vibration extremes (to simulate the rocket powered flight regime) and space simulation (to simulate the vacuum and thermal interface conditions encountered in orbit).

Space simulation testing takes many days to complete because the thermal time constants are long. This means that this type of testing constitutes the most severe test, that is, continuous operation without access for vehicle repair, adjustment or maintenance. Recognition of the usefulness of such a test in elimination of flawed equipment has led some programs to require long term operation (say 30 days) with no out-of-tolerance performance as a final acceptance test.

Acceptance testing for each satellite is done to confirm that workmanship is adequate and that the equipment functions, in detail, as expected. Acceptance tests include end-to-end functional tests that exercise every subsystem; an acoustic or vibration test; a thermal vacuum test; and, a high pressure test of any spacecraft pneumatic or hydraulic lines or pressurization equipment. After acceptance testing, the spacecraft is ready for shipment to the launch site.

Prelaunch validation tests are conducted at the launch site to demonstrate

- that the vehicle was not damaged in shipping,
- that the spacecraft was properly mated to the launch vehicle, and

- that the spacecraft can be controlled from the ground, which generally means from the Air Force Satellite Control Facility (AFSCF) and the launch base facilities.

The above test program, in conjunction with strict configuration control, is part of the logical sequence of test and evaluation that demonstrates each satellite will perform its mission in space.

Qualification and acceptance testing of space vehicles requires some elaborate and unique test facilities and instrumentation. A typical thermal vacuum test setup is shown in Figure 1. The FLTSATCOM spacecraft is shown (without solar arrays) in the 30 foot spherical space simulation chamber as is a portion of the ground support equipment needed to operate the spacecraft. Thermal instrumentation, data processing equipment and chamber controls are not shown. A photograph of a 22 x 40 foot bottom loading space environmental chamber is shown in Figure 2. An acoustic test facility that can provide a sound pressure of 120 to 154 decibels is shown in Figure 3.

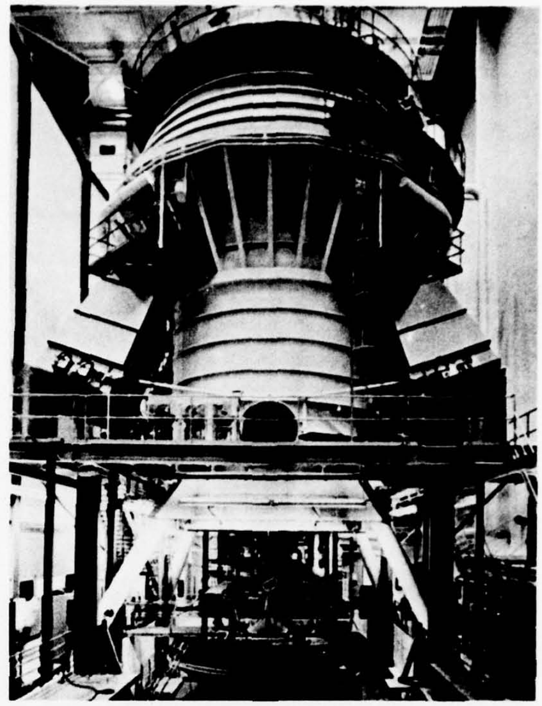


Figure 2. Environmental Test Chamber

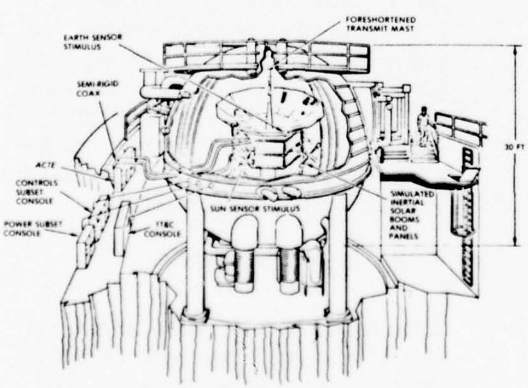


Figure 1. Thermal Vacuum Test Setup

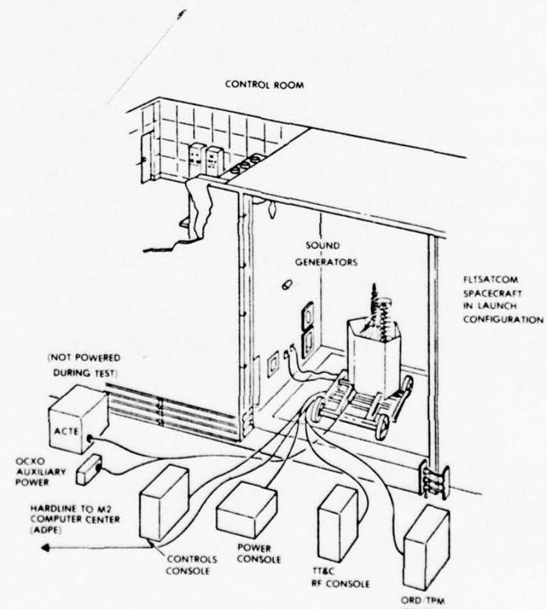


Figure 3. Acoustic Test Configuration

Figure 4 shows a deployment test facility for large structural elements such as solar arrays.

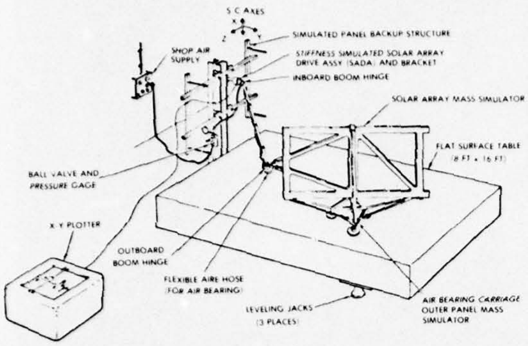


Figure 4. Deployment Test Setup

Functional performance testing of a spacecraft makes use of the onboard telemetry and command subsystem and hardline connections to critical test points. Because of severe weight limitations, built-in test equipment and monitors that do not serve a necessary function after lift-off are minimal. Most electrical components have a test connector that allows checkout using special component test equipment and a nonflight test harness during spacecraft assembly. The spacecraft has a set of test connectors that are accessible during onstand checkout. These connectors bring out critical test points (typically 300 for the FLTSATCOM spacecraft) and provide for connection of the ground power supply that charges the spacecraft batteries and powers the spacecraft bus. In addition, two long wires that are interlaced with the spacecraft harness are used to detect any abnormal electrical noise owing to electromagnetic interference or switching transients.

During thermal/vacuum testing, nonflight test thermocouples, thermistors and radiometers are located throughout the spacecraft and chamber to verify the thermal design under a variety of operating conditions.

Optical stimulators for checkout of the attitude control subsystem are mounted to simulate the earth limb, sun and star signals that serve as attitude control references.

Radio frequency measurements are made through use of a variety of receiving antennas, coaxial cables and waveguide connections to the spacecraft.

Comprehensive functional spacecraft checkout requires that every operating mode and every component, switch and redundancy path be exercised to verify integrity. In a typical spacecraft there are more than 1000 possible ground commands, a similar number of telemetry measurements and several hundred temperatures to be monitored during ground testing. Thus the use of computers for data processing and command sequencing is essential to achieve repeatable test conditions and economy of schedule.

The set of electrical ground support equipments for functional checkout of a typical communication satellite such as FLTSAT consists of the following:

- ***Electrical Power Test Set***

  The Electrical Power Test Set provides electrical power, battery charge control, simulation of solar array output and display of critical spacecraft voltages and currents.

- ***Attitude/Velocity Control Test Set***

  The Attitude/Velocity Control Test Set provides artificial stimulation of the spacecraft's attitude control sensors. This test set also monitors response of reaction wheels, thrusters and solar array drive motors through telemetry or hardline connections for each operating mode of the attitude and velocity control subsystem.

- ***TT&C Test Equipment***

  The Tracking, Telemetry and Command Radio Frequency Console, through coax lines, transmits command uplink signals and receives telemetry downlink signals to the spacecraft.

- ***TT&C Test Set***

  An additional test set contains the digital data processing encrypting and recording equipment for encoding and decoding telemetry and command data.

- ***Automatic Communications Test Equipment***

  The Automatic Communications Test Equipment performs all communications payload tests. Commercial RF test equipment is used. This equipment is controlled by a minicomputer with a capability of storing test sequences for about 2000 separate tests, using over 500 commands. Tabulation and plotting of test results in near real time allows for assessing progress as the test proceeds and for flexibility of testing.

- ***Automatic Data Processing Equipment***

  The Automatic Data Processing Equipment consists of a small real time digital computer with 64000 byte memory capacity and is used to process and display measurements, provide out-of-limits warnings, generate commands and command sequences, store calibration data and provide personnel access for control of the test.

- ***Thermal Test Equipment***

  The Thermal Test Equipment controls special heat sources used during thermal vacuum testing, multiplexes test thermocouple or thermistor outputs and contains a separate computer that processes and displays thermal measurements.

Techniques for reducing the cost and duration of the test program have been developed over the years as automatic data processing equipment has become available and the technology of spacecraft design has matured. Computer hardware, peripheral equipment and computer software have been designed so that payload tests and certain spacecraft subsystem tests can be carried out in parallel. Commonly used computers and software languages minimize startup costs. Use of centralized computers for the testing of different spacecraft in the same assembly and test area have been considered. The fact that spacecraft can be expected to become more complicated while at the same time requiring longer checkout and burn-in testing will require further automation of test procedures.

MANAGING T&E PROGRAMS

Specialized or custom design, low production rates, high reliability and other unique aspects of space vehicle programs impose requirements for significant management trade off among cost, schedule and technical risk in the planning and implementation of test and evaluation programs. Although the test and evaluation experience gained over the past two decades now is reflected in a number of directives, standards, and specifications, there is a flexibility within this framework to custom design each test program, to make appropriate cost, schedule and risk trades to achieve maximum confidence that the design and workmanship will meet mission requirements.

Several significant considerations that can affect cost, schedule, or performance risk are discussed as examples of tradeoffs in the test program.

- **To be effective in influencing design so that the vehicle can be adequately tested, and to reduce schedule, one should commit to and start implementation of a test philosophy and selection of test equipment very early in the program. However, the earlier the design of test equipment is begun, the more vulnerable it is to design changes imposed by changes in design of the vehicle. Cost and schedules are adversely impacted when these changes occur.**

- **The very low number of production articles in each program requires a**

different way of looking at how general purpose facilities, equipment, tooling, etc., will be supplied. Since each program is specifically tailored to its unique mission, the amount of special purpose equipment and specialized computer software tends to be disproportionately high with respect to capital investment and as a percentage of total program costs. The challenge is to increase general purpose capability without jeopardizing the ability to accomplish specialized testing.

- **Since test and evaluation starts at the piece part level and continues through component, subsystem and system level tests, cost and schedule can be reduced by the selective performance of environmental, burn-in, electrical compatibility and other types of tests only at one or two of the assembly stages, rather than at each of the four stages. Such cost and schedule reductions must be accomplished without adding significant risk.**

- **The constraints imposed by the space operating environment and the pressures to keep spacecraft weight down may result in a spacecraft that is relatively fragile, easily damaged, and difficult to maintain because of packing density. To reduce the possibility of damage, safety awareness programs and training need to be continuously worked, protective coverings and devices have to be used for sensitive components, and highly disciplined and procedural test operations are required. These requirements add schedule and cost constraints that must be traded off against potentially serious damage to the test article.**

Although there are many considerations which are amenable to tradeoff, there is a significant body of experience that verifies standardization is a cost effective goal. Thus every test should be thoroughly planned in terms of objectives, data requirements, criteria for success, etc. Step-by-step procedures that call for quality assurance sign-off at all critical steps as the steps are performed are issued to test personnel well in advance of each test. Records are kept of all measurements and all anomalies or "out-of-spec" conditions. The data are evaluated by knowledgeable personnel for tell-tale trends after each major test or environmental exposure to determine any incipient failures that require repair or replacement of equipment prior to launch.

The overall management of test programs, especially for the larger and more complex systems, is enhanced by utilization of earned value and other advanced program management techniques that have been developed and proved over this past decade. The large investment represented by the test items, the many engineering internal and external interfaces involved, and above all the ever present requirement for long term, fault-free reliability demand that detailed cost effective management be exercised at all times. Nothing can be left to chance.

Dr. George J. Gleghorn is the Assistant General Manager for Engineering, Space Systems Division, TRW Defense and Space Systems Group. Dr. Gleghorn joined TRW, Inc., in 1954. He engaged in System Engineering and Technical Direction on the ballistic missile and space booster programs—Atlas, Thor, Titan. From 1958 to the present he has worked in the management of many satellite projects. Early in 1960 he was Program Manager for the Orbiting Geophysical Observatory. Dr. Gleghorn has been in a supervisory management position for such satellites as Pioneer, Vela, Intelsat III, DSCS II, and FleetSatCom. For a 2-year period Dr. Gleghorn was manager of the TRW System Integration and Test Operations.

Prior to his affiliation with TRW, Inc., Dr. Gleghorn worked for several years at Hughes Aircraft and the Jet Propulsion Laboratory. He served 2 years during the Korean War in the US Navy (destroyers).

Dr. George J. Gleghorn received his B.S. from the University of Colorado. He performed his graduate work at Cal Tech.

MANAGEMENT OF MAJOR DOD TEST FACILITIES

by

John W. McCord, Director, Test Resources
Office of the Undersecretary of Defense Research & Engineering

INTRODUCTION

The Department of Defense Major Range and Test Facility Base (MRTFB) comprises 26 test facilities under the management of the Army, Navy and Air Force. See Table I. Within the Office of the Secretary of Defense (OSD), responsibilities for the MRTFB reside with the Deputy Undersecretary of Defense Research and Engineering (Test and Evaluation) (DUSDR&E(T&E)), Office of the Undersecretary of Defense Research & Engineering. Support is provided by the Director for Test Resources. Key responsibilities include

- providing overall policy direction and guidance,
- insuring adequacy to meet present and future requirements, and
- achieving optimum utilization of range assets.

Within each Service, test facility management and administrative activities take place at the headquarters (Army Staff, Chief of Naval Operations, Air Staff) and systems command (Army TECOM, Naval Air Systems Command, and Air Force Systems Command) levels, as well as at the facility command level. The activities include defining specific missions, policies and plans to guide facility development and structuring the facilities to support specific kinds of test activities and internal test programs, joint DOD test programs and, other facility user requirements within the capability of the facilities. The Service must plan and budget for all facility costs and, with a few exceptions, fund all indirect costs, i.e., those costs not directly attributable to a user. The facility commander is required to develop and maintain a master plan for the operation and development of the facility and act as a general "plant manager."

The status of range assets such as real property, data acquisition and processing equipment, and mobile platforms is continually balanced against test support requirements. Because of budget constraints range improvement and modernization programs are driven by the needs rather than the desires of concerned activities. The entire spectrum of test resources of all Services is viewed with the intention of satisfying total Department of Defense (DOD) requirements with minimum duplication. The Deputy Undersecretary of Defense Research and Engineering (Test and Evaluation) (DUSDR&E(T&E)) is the focal point within DOD to resolve major problems concerning encroachment and the availability of adequate land, sea, airspace, and electromagnetic environments to support DOD Testing.

A series of studies aimed at consolidating test functions to effect cost savings was initiated during FY 1977. These actions are supported by recent Congressional guidance and General Accounting Office (GAO) recommendations. Also, in past years savings were realized by closing unneeded facilities, deactivating surplus instrumented ships and aircraft, and retiring

ARMY	
Cold Regions Test Center Fort Greely, Alaska	White Sands Missile Range New Mexico
Tropic Test Center Fort Clayton, Canal Zone	Kwajalein Missile Range Marshall Island, Pacific
Yuma Proving Ground Yuma, Arizona	Electronic Proving Ground Fort Huachuca, Arizona
Jefferson Proving Ground Madison, Indiana	Dugway Proving Ground Salt Lake City, Utah
	Aberdeen Proving Ground Aberdeen, Maryland
NAVY	
Pacific Missile Test Center Point Mugu, California	Naval Air Engineering Center Lakehurst, New Jersey
Atlantic Undersea T&E Center Andros Island, Bahamas	National Parachute Test Range El Centro, California
Naval Air Test Center Patuxent River, Maryland	Naval Weapons Center China Lake, California
Naval Air Propulsion Test Center Trenton, New Jersey	Atlantic Fleet Weapons Training Facility Roosevelt Roads, Puerto Rico
AIR FORCE	
Space & Missile Test Center Vandenberg AFB, California	Flight Test Center Edwards AFB, California
Eastern Test Range Patrick AFB, Florida	Armament Development & Test Center Eglin AFB, Florida
Satellite Control Facility Sunnyvale, California	Air Defense Weapons Center Tyndall AFB, Florida
Tactical Fighter Weapons Center Nellis AFB, Nevada	Arnold Engineering Development Center Tullahoma, Tennessee
	4950th Test Wing Wright-Patterson AFB, Ohio

Table 1, Department of Defense Major Range and Test Facility Bases

obsolete instrumentation. A recent study leading to a system of uniform measures of workload, capacity, and capabilities should allow accurate planning at DOD test facilities. Such a system will assist management levels in analyzing resource utilization and provide a valuable tool to achieve maximum efficiency in the MRTFB. In accomplishing the above functions, close liaison is necessary between key OSD and Service elements as well as with non-Defense agencies such as the Office of Management and Budget (OMB), the National Aeronautics and Space Administration (NASA), the Energy Research and Development Agency (ERDA), the Federal Aviation Administration (FAA), the Department of the Interior (DOI), and the Department of Transportation (DOT).

POLICY AND REVIEW

DOD Directive 3200.11

Defense policy for the use, management and operation of the MRTFB is established by DOD Directive 3200.11. The directive applies to the Office of the Secretary of Defense, the Organization of the Joint Chiefs of Staff, the Military Departments, Unified and Specified Commands, and Defense Agencies. The stated objective of DOD Directive 3200.11 is:

> **"...to insure provision of effective test and operational support by facilitating joint use of the MRTFB, by consolidating and standardizing management responsibility at appropriate levels, and by setting forth uniform operating guidelines."**

The directive designates the 26 major DOD ranges and test facilities that compose the MRTFB. These facilities were selected after a comprehensive DOD Test and Evaluation Facility Base Review accomplished in 1971 to determine which test and evaluation ranges and centers were essential to Defense needs. Some 94 test activities were reviewed following a screening process that considered an initial list of several hundred T&E-related activities. The 26 original MRTFB facilities were ultimately chosen, largely because of their significant test resource assets, multipurpose or multiuser capability and/or unique characteristics or mission.

Under the provisions of DOD Directive 3200.11, the composition of the MRTFB is periodically reviewed to determine if changes should be considered owing to current mission demands or new criteria. Example: in 1976 the Air Force Special Weapons Center, Albuquerque, New Mexico, was deleted from the MRTFB and the 4950th Test Wing, Wright-Patterson Air Force Base, Ohio, was added. As a result of studies recently conducted by OSD and the Military Departments, it is envisioned that additional changes in MRTFB composition will be made within the next few years.

Additionally, the DOD Directive 3200.11 outlines specific responsibilities of the DUSDR&E, Service Secretaries, facility commanders and range users.

The DUSDR&E provides policy direction and planning guidance and insures that the MRTFB is adequate to meet present and future DOD requirements. This is accomplished through annual budget and apportionment reviews, and continuous oversight of range operations and resource needs. The Service Secretaries are designated as management agents and are charged with defining specific facility missions, programming and budgeting, and providing for the acquisition and replacement of range instrumentation. The DOD Directive 3200.11 establishes a Major Range and Test Facility Committee (MRTFC), chaired by the DUSDR&E(T&E) that includes representatives from each Military Department. The Committee provides a vehicle for direct coordination and rapid resolution of key issues affecting the MRTFB.

The major burden of providing test support to DOD and other range users falls upon the shoulders of the facility commander. He is responsible for the actual planning, coordination and conduct of test support and for insuring that all safety requirements are met. Range users must interface with the facility early, provide adequate planning data and fund the direct costs of all test support. The success of the test support mission is predicated upon the timely and thorough information exchange between the range commander and the user.

The DOD Directive 3200.11 also defines the Uniform Funding Policy and provides for consideration of environmental aspects relative to major actions affecting the MRTFB. The directive is presently undergoing review and soon will be revised to incorporate recent organizational and policy changes within OSD and the Military Departments.

Uniform Funding Policy

Nineteen of the 26 major DOD ranges and test facilities, those possessing the greatest potential for multiservice support, operate under a Uniform Funding Policy (UFP). The policy became effective in July 1974. Under the UFP, range users reimburse the facility for test costs directly related to their programs. Other range expenses are funded institutionally by the facility. A uniform system of funding resulted from the findings of the Blue Ribbon Defense Panel following a series of studies. Under prior funding policies, users paid the full cost of testing at some facilities while at other facilities costs were not paid in any amount. At institutionally funded activities there was little incentive for program managers to carry out advanced range support planning. In many instances range selection was based upon economic rather than technical factors. The various funding systems made Service and OSD management difficult. The systems proved inefficient and often obscured the true cost of testing.

A Joint Logistics Commanders panel completed reviews of the UFP in June 1975 and October 1976 and reported that with a few exceptions it is working well. Major problems are not evident—minor problems are being addressed. The policy proved extremely helpful during the justification of annual range budgets and has resulted in better planning and communications within the range support and user communities. The cost of support will continue to vary in accordance with type and location of the operation.

Budget Review

The DUSDR&E(T&E) has three broad areas of budgetary responsibility:

- The Director of Test and Evaluation Defense Appropriation which he manages and controls;
- Military Department test and evaluation support programs for which he has primary OSD cognizance; and
- Military Department major and selected, less-than-major RDT&E programs which he monitors.

The budgetary responsibilities are summarized below.

Director of Test and Evaluation, Defense Appropriation

This separate Defense Appropriation was established by Congress in FY 1973 to support Joint Operational Test and Evaluation (JOT&E). The appropriation provides funds for study efforts associated with the major DOD ranges and test facilities. The DUSDR&E(T&E) manages and controls these funds, including the preparation of the budget request for submission to Congress, testimony before Congress, and execution of the approved budget.

Military Department T&E Support Programs

The DUSDR&E(T&E) reviews MRTFB funding (all appropriations) in detail during annual budget, apportionment and POM reviews. In addition, the Deputy Director has primary cognizance for 18 other RDT&E program elements. The program elements are primarily related to Service test support to include operation and maintenance, overhaul of ships and aircraft, OT&E capabilities, instrumentation, aerial targets, aircraft survivability, project SEMI, and other programs.

Military Department RDT&E Programs

The major responsibility assigned by the Secretary of Defense to DUSDR&E(T&E) is "monitoring closely the test and evaluation planned and conducted by the DOD components for major acquisition programs and for such other programs as he believes necessary." In FY 1978, DUSDR&E(T&E) will monitor a total of 85 Army, Navy, Air Force and Defense Communications Agency major weapon system programs and 15 less-than-major programs.

In FY 1978, budget requests for test and evaluation related activities in all appropriations are an estimated $2.7 billion. The DUSDR&E(T&E) is involved in the complete program budget process. Involvement includes the formulation and dissemination of test and evaluation funding guidance in coordination with the OASD (Comptroller) and active participation in the OSD/OMB budget and apportionment reviews. The Director for Test Resources performs programming and budgeting functions for the DUSDR&E(T&E) and maintains continuous liaison with elements of OMB, OSD, and the Military Departments. Additional budget activities requiring review and coordination include Congressional Data Sheets, RDT&E Descriptive Summaries, Selected Acquisition Reports, Arms Control Impact Statements, GAO Annual Staff Studies, and Congressional testimony and other related actions with test and evaluation impact.

An important OSD level management function is associated with MRTFB funding that in FY 1978 totals some $1.7 billion from several appropriations and includes funds recovered from users for support provided. Each Autumn a comprehensive review is conducted of Service budget submissions for the 26 major DOD ranges and test facilities and the 18 other assigned test and evaluation support programs. The test facility (or support program), systems command and headquarters staffs prepare and present the budget. Representatives from DUSDR&E(T&E), OASD (Comptroller) and fice of Management and Budget perform the review.

Of particular interest at each test range and center are the improvement and modernization program, military construction projects, and manpower and workload data. The validated budgetary requirements of the test facilities and programs are compared with overall fiscal constraints, and, in conjunction with the Services, adjustments and offsets are made. This program is monitored during application for any required modifications necessitated by project or budget changes.

The programs are studied again in conjunction with the DOD Apportionment Review and monitored during the budget year through Service staff elements and status reports. Recommendations are made with respect to instrumentation programs, military construction projects, manning levels and other pertinent areas affecting the MRTFB. The test resource *needs* of the Military Departments are highlighted and priorities are established for the overall test support community to meet these needs within budgetary limitations.

ADEQUACY

Range Assets

A principal responsibility of DOD agencies associated with test resources is to insure that the MRTFB is adequate to meet current and future test and evaluation requirements. This responsibility is carried out by all levels of management to varying degrees and results in:

(1) Continual review of the capabilities and capacities of the MRTFB components to accomplish generic test and evaluation support tasks, e.g., aircraft testing, strategic missile testing, electronic systems testing, etc.

(2) Examination of the types and amounts of test and evaluation workload created by the testing needs of weapon systems programs.

(3) Comparative analysis of the MRTFB capability/capacity versus the type/amount of T&E requirements to derive both qualitative and quantitative deficiencies in the test resource base.

(4) Development of plans and funding approaches to meet validated MRTFB deficiencies at minimum cost to the institutional test support base and to the weapon system programs using this base.

(5) Coordination of test resource development and acquisition projects between the Military Departments to insure that such projects achieve the maximum feasible tri-Service application.

(6) Transfer of test resources among the Services and sponsorship of joint Service development and acquisition efforts, where possible.

(7) Interchange of data associated with all Service programs related to the development of specialized test instrumentation or T&E techniques and methodologies so that the entire community can benefit.

Care is taken to coordinate all efforts with the appropriate agencies outside DOD (e.g., NASA, FAA, DOI). Where required, specialized expertise is obtained from a technical contractor. Activities in various test resource areas illustrate the application of this approach to meeting deficiencies and insuring test resource adequacy.

TSPI Systems

The principal requirement of many test exercises conducted within the MRTFB is the accurate determination of time space-position information (TSPI). The position (and sometimes velocity and acceleration) of a test object (e.g., aircraft, missile, ship) is measured by TSPI instrumentation as a function of time for use in test reconstruction and evaluation, and for range safety. Advances in weapon system capabilities, increasing requirements for more realistic testing/training, and growing constraints imposed by civil encroachment on test ranges have combined to motivate development of new time space-position information systems.

The Extended Area Test System under development at the Pacific Missile Test Center, Pt. Mugu, California, will provide TSPI, communications and target drone control capabilities in the PMTC Outer Sea Test Range up to several hundred nautical miles from land. See Figure 1. The movement of test operations seaward is necessitated by the higher dynamics of current and planned weapon systems and by the encroachment of Outer Continental Shelf oil exploration on traditional PMTC operating areas closer to shore. The Extended Area Test System will accommodate nearly 100 test objects and will use equipment modules aboard test vehicles, processing and relay subsystems aboard special instrumented aircraft, and ground based reference and computation systems.

At the Hill/Wendover/Dugway complex in Utah, a High Accuracy Multiple Object Tracking System is being installed, principally for remotely piloted vehicle testing. This equipment can track more than 20 airborne and surface vehicles simultaneously and display the vehicle positions in real time.

The Air Combat Maneuvering Range and Air Combat Maneuvering Instrumentation are related TSPI systems developed by the Navy and Air Force for use during operational test and evaluation and training exercises. See Figure 2. The systems permit real time evaluation of tactics and simulated weapon firings. Recording, playback and display features allow postflight analysis of maneuvers.

To provide an "at-sea" environment for operational test and evaluation and training, the Navy is developing the Mobile Sea Range. See Figure 3. When fully implemented two mobile sea range systems will furnish a portable, shipboard time space-position capability for Atlantic and Pacific operations. Divorced from direct support by land based elements, the mobile sea range will permit tracking, communications, and related functions for large numbers of airborne, surface and, eventually, subsurface participants.

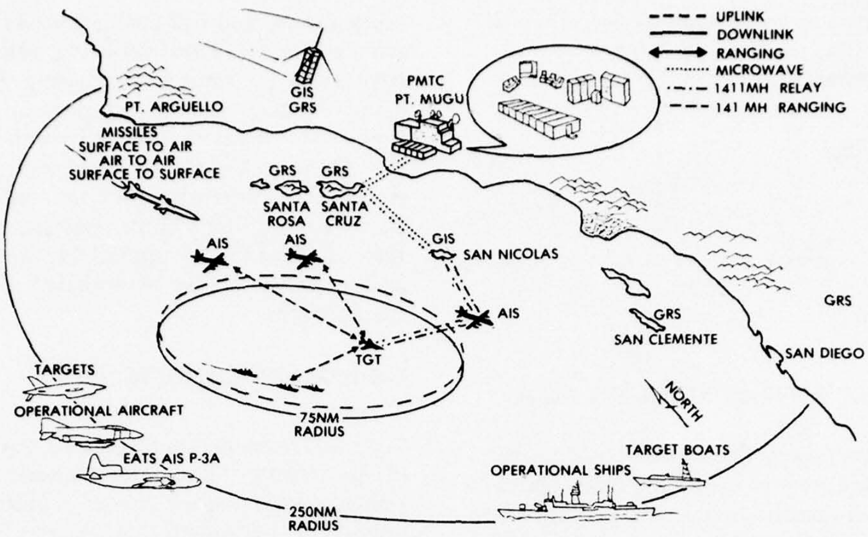


Figure 1. Extended Area Test System

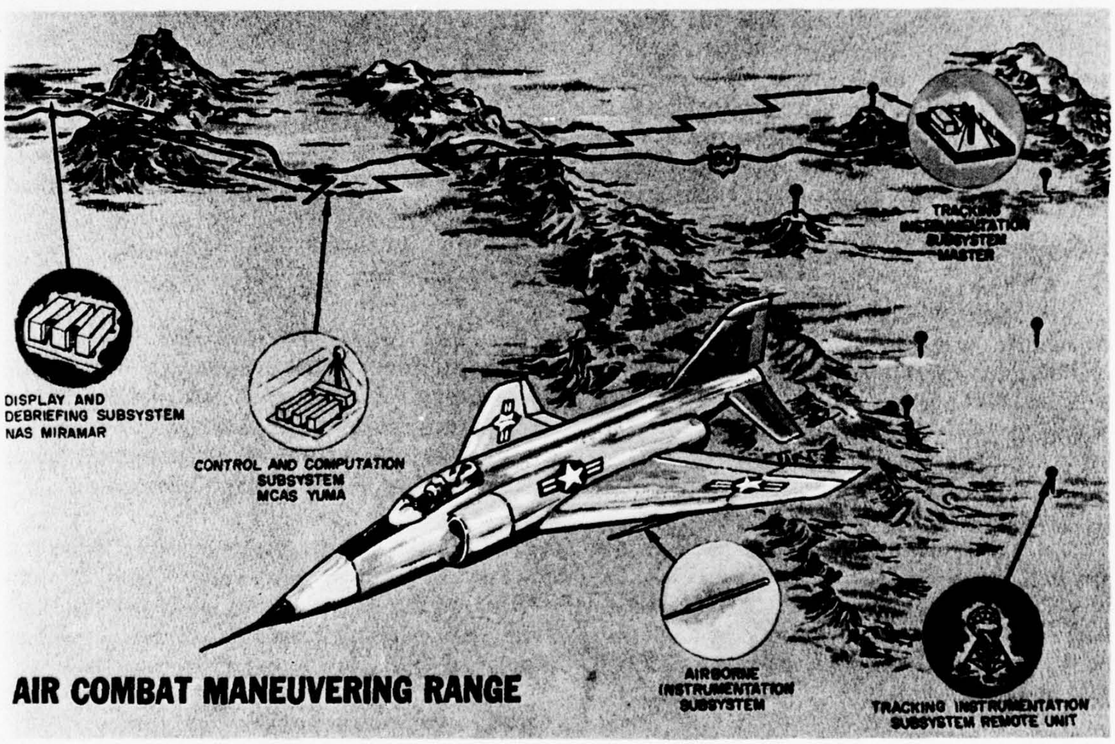


Figure 2. Air Combat Maneuvering Range

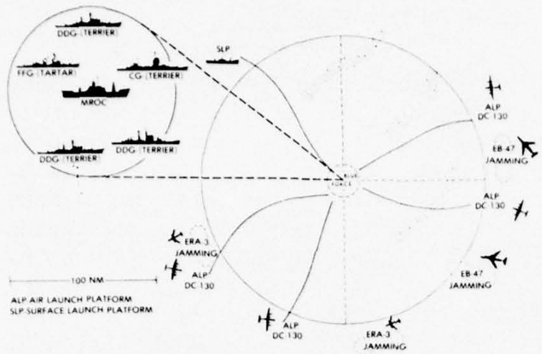


Figure 3. A Typical Mobile Sea Range

To meet special time space-position information (TSPI) requirements the Navy is developing and evaluating for test range purposes a Multiple-Object Instrumentation Radar (MIR) and a dual-frequency tracking radar. The low cost Multiple-Object Instrumentation Radar will employ phased array technology to obtain higher precision simultaneous TSPI on several test vehicles from a single station. The dual frequency tracking radar is designed to solve the difficult problem of accurately tracking a test object at low elevation angles during low-altitude or near-horizon operations. The Wide Area Active Surveillance system provides for TSPI on multiple sea and air targets in addition to performing a surveillance function.

Telemetry and Test Data Processing

Projects are under way to improve range capabilities to acquire, process, record, display, and evaluate telemetry test data. At the Pacific Missile Test Center, a Computer Centralization and Modernization Program wil replace ten obsolescent computers. The replacement will be a single, integrated, primarily real-time system built around a redundant pair of modern central processors. At the Space and Missile Test Center, Vandenberg Air Force Base, California, a Telemetry Integrated Processing System will replace fourteen computers with six telemetry preprocessors connected to a single central processing unit.

The telemetry preprocessors will provide an enhanced capability to process data for missile checkout, calibration, go/no-go decisions, range safety, and test evaluation. At the Naval Air Test Center, Patuxent River, Maryland, the Real-Time Telemetry Processing System is being expanded to accommodate two additional data channels. This expansion will enable the system to handle workload requirements of increasing complexity and amount. The Air Force Flight Test Center, Edwards Air Force Base, California, is upgrading the capability and capacity of the Automated Flight Test Data System.

Underwater Instrumentation

Advances in the range and sophistication of underwater and antisubmarine warfare weapon systems have created requirements for improved instrumentation. At the Navy's Atlantic Undersea Test and Evaluation Center, Andros Island, Bahamas, an Advanced Sonar Calibration Range is being designed and developed. This mobile facility will provide a major capability for underwater acoustical measurement required by modern submarines and surface ships. The Barking Sands Underwater Range Extension project at the Pacific Missile Test Center, Hawaii will increase precise underwater test vehicle tracking and control. This coverage will permit more realistic system testing and antisubmarine warfare training than are possible elsewhere.

Surveillance

Test range operations with high performance aircraft, missiles, and target drones create hazards to personnel and equipment. To maximize the safety of participants and eliminate risks for nonparticipants close surveillance of the range operating areas is maintained.

At the Navy's Atlantic Fleet Weapons Training Facility, Roosevelt Roads, Puerto Rico, testing and training operations are conducted that involve multiple aircraft, ships, and remotely controlled target drones, both airborne and surface. Live air-to-surface, air-to-air and surface-to-air munitions are employed. The location of this facility in an area of dense commercial and pleasure ship and aircraft traffic contributes greatly to the need for range surveillance. The task is currently accomplished using a low data rate air traffic control radar system: A Wide Area Active Surveillance system is proposed. This system would employ phased-array technology and advanced data processing techniques to survey operating areas and provide simultaneous high-accuracy tracking data on multiple test vehicles.

The R-2508 restricted airspace complex over southern California provides an area for operations by several DOD activities. Commercial airline traffic crisscrosses the area. Private aircraft operating under visual flight rule (VFR) conditions can appear anywhere at anytime. To meet the need for surveillance and control of high performance military aircraft, cruise missiles and air-launched ordnance, and to prevent catastrophic interactions with nonparticipating aircraft, a joint Army/Navy/Air Force/FAA R-2508 Enhancement Plan is under implementation.

Simulation

The accurate simulation of the physical environment in which a weapon system must operate is crucial in many test and evaluation exercises. Valid testing of aircraft vulnerability to nuclear blasts must include exposure of avionics/electronic systems to electromagnetic pulse phenomena. To this end, the Air Force is constructing the TRESTLE facility at Kirtland Air Force Base, New Mexico, to produce representative electromagnetic pulse phenomena by nonnuclear means.

Meaningful test and evaluation of electronic countermeasure equipment and tactics requires a realistic simulation of the signals emitted by threat systems. At the Air Force Tactical Fighter Weapons Center and the Naval Weapons Center's ECHO Range, electronically accurate replicas of enemy land and shipborne anti-aircraft missile and gun systems have been constructed to provide an authentic environment. These facilities are expanded and updated based on the most recent intelligence information.

Mobile Platforms

Requirements to provide tracking, telemetry, communications, and related support to strategic and tactical missile test operations over large areas remote from land stations resulted in a fleet of instrumented ships and aircraft. See Figures 4 and 5. The fleet includes the Ec-135 Advanced Range Instrumentation Aircraft, the EC-121 instrumented aircraft, the Advanced Range Instrumentation Ships (USNS ARNOLD and USNS VANDENBERG), and the instrumented ships USNS WHEELING and USNS REDSTONE. These platforms are the objects of continuing improvement programs to overcome capability deficiencies. Obsolete and inefficient ships and aircraft are retired. Instrumented platform needs for the 1980's and beyond are underway.

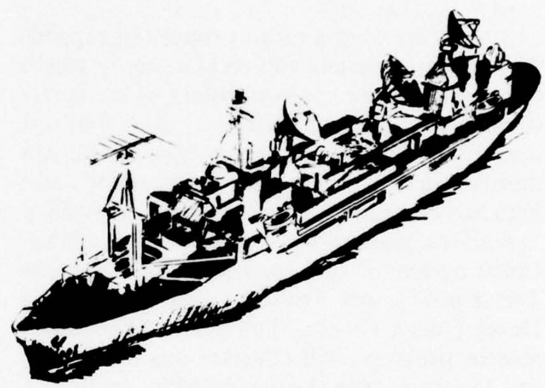


Figure 4. Instrumented Ship

Figure 5. Instrumented Aircraft

Electro-Optical Systems

The rapid development of high energy laser technology has left the DOD unprepared to carry out a complete test program in this area.

To meet this deficiency a tri-Service study panel under DUSDR&E(T&E) guidance was formed to examine joint requirements and potential approaches for developing a High Energy Lasar Systems Test Facility (HELSTF). After examining six alternative sites within the MRTFB the panel selected White Sands Missile Range, New Mexico. Planning for HELSTF is presently underway with maximum use being made of existing buildings and instrumentation.

In addition, to accommodate lower energy lasar seeker tests not appropriate for the large outdoor range environment of HELSTF an Electro-Optical Test Facility is being developed at the Armament Development and Test Center, Eglin Air Force Base, Florida.

Propulsion Systems

New levels of aircraft propulsion capability and the materials and techniques by which they are achieved have outdistanced the corresponding capacity of DOD and industrial test resources. Problems of currently operating US military aircraft such as compressor stalls and high fuel consumption might have been avoided had an adequate test capability been available. Construction of the Aeropropulsion Systems Test Facility, Air Force Arnold Engineering Development Center, Tullahoma, Tennessee, now in progress, will alleviate this deficiency. See Figure 6. The Aeropropulsion Facility, a $437 million military construction project, will permit ground simulation of critical aerodynamic parameters so that propulsion systems can be optimized at the design stage.

Targets and Target Control

The ultimate noncombat test of most weapon systems is an ability to inflict damage on a realistic replica of the design threat. Targets realistic in performance, signature, and tactical operation are an important part of the test resource inventory. Two target systems now in development are typical of required threat simulation capability. One is the Army's STREAKER subscale aerial target that shows potential for tri-Service use. The second is the Anti-Ship Missile Target under development by the Navy. This antiship missile target is to simulate the performance and signature of Soviet antiship missiles.

To address operational deficiencies in target control capability, an Integrated Target Control System was developed to meet tri-Service needs across the family of aerial targets. One of these needs was to follow the DOD direction to abandon the existing drone control frequency band and select another frequency band. For situations where threat formation is significant (i.e., number of threat vehicles and the spacing and formation structure) the Army's Drone Formation Control System, White Sands Missile Range, will provide a multiple drone control capability.

Transfers of Resources

Not all test resource deficiencies are met by new developments and procurements. Often instrumentation excess to needs at one test range is transferred to meet a requirement elsewhere within the MRTFB. Example: a TPQ-18 tracking radar and two splash detection radars made surplus by the disestablishment of the Phoenix Islands tracking station were relocated to Kwajalein Missile Range and are operational. Tracking subsystems of the phased-out NIKE-HERCULES air defense systems are finding applications at several locations within the MRTFB. Other cooperative ventures include the operation by NASA of the Air Force FPS-16 radar at Kokee Park, Hawaii, in support of Navy programs and the use of the NASA tracking facility, Wallops Island, Virginia, by the Naval Air Test Center, Patuxent River, Maryland.

Encroachment

The adequacy of DOD test facilities involves the provision of sufficient land, sea, airspace, and electromagnetic environment in which to conduct test and evaluation. The Department of Defense currently has significant encroachment problems from sources such as energy exploration or exploitation, commercial and private aviation and marine activity, civil demands for access to or joint use of real estate,

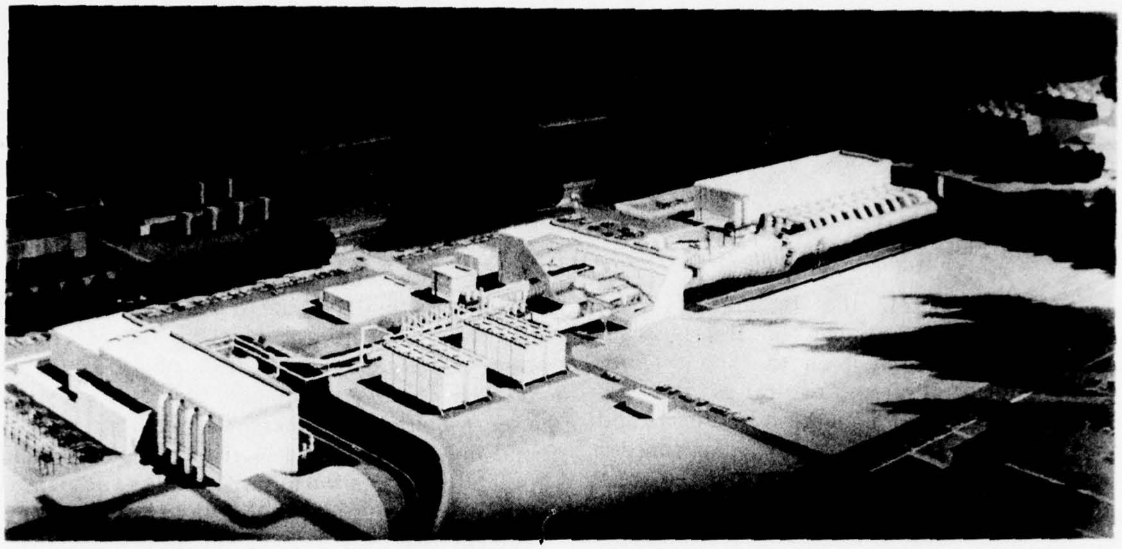


Figure 6. US Air Force Arnold Engineering Development Center

electromagnetic emission, and international pressures (political and financial) on our foreign operating locations. At many facilities, these encroachment areas are present in various combinations further intensifying the problem. Major problems over the past few years have involved airspace usage by both the military and civilian aviation sectors and oil exploration/exploitation activities on the Outer Continental Shelf.

R-2508 Restricted Airspace

A serious airspace encroachment problem involves the R-2508 Restricted Airspace Complex located in the Upper Mojave Desert of California. The complex includes: the Naval Weapons Center, China Lake; the Air Force Flight Test Center, Edwards Air Force Base; George Air Force Base; and the Army's Fort Irwin. The R-2508 Complex is a multibillion dollar tri-Service test complex used by DOD for the advancement of weapons systems technology. The principal DOD activities are aircraft flight testing, RDT&E of weapons, electronic countermeasure and electronic systems testing, training of pilots and tactical air crews, and large scale ground force exercises. The area is roughly 140 miles long by 110 miles wide and consists of a high level restricted area (R-2508) that overlays six independently managed restricted areas where aircraft and other weapons systems are tested and where operational training is conducted. The low level restricted areas are separated by unrestricted airspace within which the flying public may operate. Because of the mountainous terrain general aviation traffic naturally funnels into areas immediately adjacent to high performance, high cost test activities. These geographical constraints and the lack of adequate surveillance facilities create conditions for potential midair collisions. Currently, large areas of the complex have neither radar nor communications coverage of military or civil operations. The air traffic problem is complicated by the fact that commercial operations in and around the R-2508 restricted area are expanding at a rapid rate. With plans for Palmdale, just south of Edwards Air Force Base, to become the major Los Angeles air terminal in the 1980–1990 time period, this expansion will surely continue.

After detailed analysis, it was concluded that expansion of the present radar network and the FAA-operated Radar Approach Control Center at Edwards Air Force Base would rectify the existing situation. A program is in progress to integrate three long range radars, two approach control radars and six gap filler radars. A mosaic (through a computer system and associated software and peripheral equipment) is to provide a capability for real time control of R-2508.

Outer Continental Shelf Oil Exploration

In June 1971, the Bureau of Land Management, Department of Interior, announced an extensive plan to exploit Outer Continental Shelf oil and natural gas reserves. Areas off the coasts of southern California and the Gulf of Mexico were of particular concern to DOD because of extensive use of the areas for testing weapons and training operational forces. In the Gulf of Mexico, DOD missions include gunnery, air-to-air missile firings, bomb drops, high-altitude probes, electronic warfare exercises, and aircraft carrier landing and take-off training. In a 70-mile wide corridor south of Eglin Air Force Base, of the 640,000 acres proposed for lease, DOD negotiated exclusion of 17,000 acres. Off the coast of southern California, DOD missions include air-to-air, air-to-ground, and ground-to-air missile test firings as well as extensive fleet training. Of the 7.7 million acres proposed for lease, DOD negotiated exclusion of 3.4 million acres.

UTILIZATION

Duplication

In addition to insuring the adequacy of the facility base, it is necessary to insure the efficient use of test resources by:

- Authorizing only those test resources that are needed,
- Retiring unused or obsolete assets and,
- Eliminating unnecessary duplications.

In discharging this responsibility, Defense agencies must demand that all facilities are reviewed frequently for workload, capacity, and quality-of-support indicators that signal potential problem areas.

The Military Departments are alert to savings opportunities. The Office of the Secretary of Defense provides the thread that ties all DOD test facilities together. Opportunities for economies through inter-Service use of resources are identified and joint studies and analyses are conducted to evaluate alternative support methods. Funds to maintain and operate the 26 facilities composing the MRTFB represent a significant portion of the Defense budget. The cost of support, the complexity of user requirements, increasing encroachment, and inflation combine to reduce capacity and dollars available for improvements. These factors make it imperative that the extremely important MRTFB support function be carried out in a cost effective manner, and that any funds and effort expended on unwarranted duplication be redirected to overcome program deficiencies.

General Accounting Office Duplication Studies

Improved management of MRTFB functions is of vital interest to the Congress. The General Accounting Office and DOD were directed to take certain actions. In a report entitled, "Review of the Adequacy of Department of Defense Test Resources," (dated April 30, 1975 and submitted to the Senate Appropriations Committee), the General Accounting Office concluded that a "high degree of duplication in test capabilities appears to exist." In May 1975, the Senate Appropriations Committee directed that General Accounting Office study the specific areas of ordnance, propulsion, and underwater testing to determine the extent of duplication and whether such duplication was warranted. The GAO report of March 1, 1976, titled, "Does the Department of Defense Have More Test Capacity Than It Needs?" concluded that potential capacity exceeded workload in the ordnance and engine test cell areas and implied that the workload could be accomplished at fewer facilities. The General Accounting Office report substantiated the need for the existing family of underwater test facilities.

Department of Defense Utilization Studies

The Office of the Secretary of Defense issued directives to the Services in 1976 to review workload, capacity and management of several activities involved in supporting ordnance, propulsion, ballistic missile, parachute and Remotely Piloted Vehicle testing. The intent was to cause consolidation of workload and/or management leading to eventual reduction or closure of specific test support facilities. This action, still in process, is a precursor of further activity of this type. The requirements for test support change with national priorities.

Uniform Measures

The availability of meaningful MRTFB workload and capacity data in uniform units is a prerequisite to the evaluation of efficiency, duplication, and/or under-utilized resources. Existing data has lacked the detail and uniformity needed for significant analysis. In its 1976 study of DOD test facilities, the GAO had difficulty in measuring workload and capacity and concluded that "the measurement of test facility workload and capacity is very complex and no common criteria exists for uniform measurement of test facility capacity and use." The OSD agreed and is working on a system that is to accurately measure test facility workload, capacity and capability in uniform units.

In 1975 an effort was initiated to evaluate techniques for the uniform collection, maintenance, and use of test facility data for the purpose of identifying unwarranted duplication. Early in the program, broader applicability of the data was recognized. The data provided resource identification, capability, workload, and capacity. The resultant system was implemented on a trial basis at three MRTFB facilities (Yuma Proving Ground, Pacific Missile Test Center and Arnold Engineering Development Center) for a 3 month period in 1976. To test the usefulness of the data, the three facilities were asked to submit workload and capacity data in the new format in support of the FY 1978 budget submission. The system is based on using investment resources as the common baseline with data on capability, capacity, workload and utilization being accumulated and reported for each resource. These data can then be aggregated by program supported, time period and event to provide the total usage relative to capacity and application. The Military Departments are now completing the evaluations.

Cost Savings

In the past several years, OSD and the Services have had an active program to retire obsolete instrumentation and equipment, to close facilities that are no longer required, and to reallocate underused resources. From FY 1970 through FY 1975 major resources and facilities were deactivated. The result was an annual savings in FY 1976 and subsequent years of $28.1 million. Specific actions consisted of deactivating seven ships (TWIN FALLS, WATERTOWN, RANGE TRACKER, HUNTSVILLE, SUNNYVALE, LONGVIEW, MERCURY) saving $20.5 million, three test support locations (Eniwetok, Eleuthera, Green River) saving $1.8 million, three major radar installations (AMRAD, HAPDAR, SIMPAR) saving $2.4 million, and releasing 10 test support aircraft saving $3.4 million. In addition, $27 million in cost avoidance was achieved by transferring seven major radars.

The Senate Committee on Appropriations Report 94–446, dated November 6, 1975, directed DOD to effect savings in test ranges and test facilities in FY 1976 and 197T, by consolidations and/or improved management, and to reflect these savings in the FY 1977 budget. Several actions were taken in FY 1976 resulting in annual savings of $15.9 million in FY 1977 and beyond. The Phoenix Islands instrumentation complex was reduced to caretaker status saving $8.7 million; the Air Force Special Weapons Center was deactivated saving $4.5 million, and 11 more test support aircraft were released saving $2.7 million. In addition, $18 million in cost avoidance was achieved by placing the WHEELING and VANDENBERG ships in ready reserve for 14 months ($8.9 million), consolidating the Army Test Boards ($0.7 million) and transferring three more radars ($8.4 million). In FY 1977 six more test support aircraft are being released for an annual saving of $1.8 million, and a reduction of about 500 civilian manpower positions is being achieved among several test facilities, saving $9.6 million.

Since FY 1970 a total annual cost savings of $55.4 million has been realized and is reflected in the FY 1978 budget. Reallocated MRTFB resources have resulted in one time savings of $45.0 million. Cost escalations in personnel, contracts, materials, supplies, and energy have more than offset the savings, making savings visibility difficult. An active program of retiring obsolete or underused resources and closing sites when possible will continue. Actions already accomplished or underway have reduced the facilities to a minimum configuration. To achieve savings in the future, we must look toward significant alternatives such as closure of major facilities. Such actions would allow the Services to strengthen the remaining facilities and provide the resources needed to support the test and evaluation of ongoing and future Defense programs.

Mr. John W. McCord is the Director for Test Resources in the Office of the Undersecretary of Defense Research and Engineering.

The key military assignments of Mr. McCord include: three deployments as an aviator with VP-22 during the Korean War; Assistant Air Operations Officer, Commander, Naval Forces Far East; Program Manager in Operations and Range Development Departments, Pacific Missile Range, CA; Research Associate at the Livermore Radiation Laboratory, CA; and Assistant Director and Technical Assistant for Space and Astronautics, Office of the Chief of Naval Operations.

Mr. McCord, following his naval service, was a technical staff member, Systems Group, TRW, Inc., Vandenberg AFB, CA. Later as an Air Force civilian Mr. McCord served as Technical Assistant to the Deputy Commander for Range Operations at the Space and Missile Test Center, Vandenberg AFB, CA, a position he held until selected for his present position.

Mr. McCord graduated from the US Naval Academy in 1948 and received his wings as a Naval Aviator in 1950. Mr. McCord holds a BS degree in Physics from the US Naval Post Graduate School, Monterey, CA, and an MS in Physics from the University of California at Berkley.

TEST AND EVALUATION OF TRI-TAC EQUIPMENT

by

Lt Col David L. Washington, Department of Air Force

In February 1971, the Joint Tactical Communication (TRI-TAC) Program was chartered primarily to insure:

The necessary degree of interoperability among tactical communications systems and other DOD telecommunications systems.

Timely fielding of new tactical communication equipment required by the Armed Forces to perform their mission reflecting the most effective technologies.

Elimination of duplication in the development of Military Service equipment.[1]

TRI-TAC developments including all trunking, access and switching equipment for mobile and transportable tactical multichannel systems, including associated systems control and technical control facilities, local distribution equipment, voice, teletype, data and ancillary terminal devices and associated communications security equipment.

INTRODUCTION

From the commencement of the TRI-TAC Program, it was envisioned that test and evaluation would be a joint endeavor conducted by the military departments and DOD agencies under the "Lead Service" concept. In recognition of the complexity of such an undertaking, a comprehensive joint test program was established. The program was to provide assurance that the objectives of equipment/system commonalty, compatibility, and interoperability would be maintained in the most cost effective manner. Moreover, the TRI-TAC sponsored equipment, engineered primarily for digital applications, must interoperate with an extensive existing inventory of primarily analog equipment, thus adding to the difficulties of an already complex test and evaluation task.

The joint test and evaluation program is a "new way of doing business" in the test and evaluation arena for tactical communications. The dominant characteristic of a joint test program is its provisions for participating services and agencies to jointly plan and conduct testing with TRI-TAC Office coordination. To achieve program objectives of interoperability and compatibility, it is imperative that a comprehensive joint performance be realized from the military departments and DOD agencies; that is, the TRI-TAC Program participants. Accordingly, an approach encompassing joint definitions of policies and responsibilities has been established to implement the test and evaluation program.[2]

T&E CONCEPTS AND REQUIREMENTS

Primarily, the TRI-TAC Joint Test and Evaluation Program is governed by the DOD process for the acquisition of defense systems. The acquisition program process is divided into four phases. Program decisions are based on evaluated test results for each phase before the decision to proceed to the subsequent phase is made.[3] Each phase has a designated milestone for the decisionmaker and an associated information-bearing testing concept.[4]

- *Milestone 0*—Program initiation (conceptual phase)
- *Milestone I*—Demonstration and validation (validation phase)
- *Milestone II*—Full-scale engineering development (full-scale engineering development phase)
- *Milestone III*—Production/Deployment (production and development phase)

The concept for joint test and evaluation was formulated based on the framework of the policies and procedures established by the Services and Agencies for acquiring heretofore unique equipments. From this framework, the TRI-TAC Office developed common test terms for each test phase to be used in the joint test of TRI-TAC sponsored acquisitions. The following definitions of test periods and associated activities were promulgated as DOD guidance to the Services and Agencies.[5]

Conceptual Phase – The conceptual phase extends from the recognition of a needed operational capability to the program decision (Milestone I) where initiation of the validation phase is authorized. Test planning, if required during the conceptual phase, is the responsibility of the acquisition Service in coordination with other program participants. Test plans are designed to demonstrate the feasibility of a proposed acquisition program providing a desired operational capability or solving a military problem. Tests during this phase are designated as research tests or feasibility tests. Both Government and contractor evaluations of test results support a decision to proceed to the validation phase.

Validation Phase – The validation phase commences after an affirmative program decision (Milestone I) and continues through the ratification decision point, (Milestone II), where full-scale engineering development is authorized. Contractor testing during the validation phase is identified as contractor demonstration. Contractor demonstration plans and procedures are approved by the Service charged with acquisition. Contractor demonstration plans are intended to insure the verification of preliminary design and engineering, and to demonstrate the degree of compliance with specifications and requirements. Government testing in this phase is identified as development suitability tests. Test coordination is the responsibility of the acquiring Service program participants. Development suitability tests are performed to demonstrate that technical and operational risks have been identified and that solutions are achievable.

Full-Scale Engineering Development Phase – The full-scale engineering development phase extends from the ratification decision point (Milestone II) to the production decision point (Milestone III). The full-scale engineering development phase is directed toward confirming that

- engineering is complete,
- previously identified technical uncertainties have been resolved,
- the operational suitability of the item has been determined, and
- realism of the plan has been confirmed.

Tests during this phase are contractor development tests and Government tests consisting of Development Test and Evaluation (developer-oriented) and Initial Operational Test and Evaluation (user-oriented).

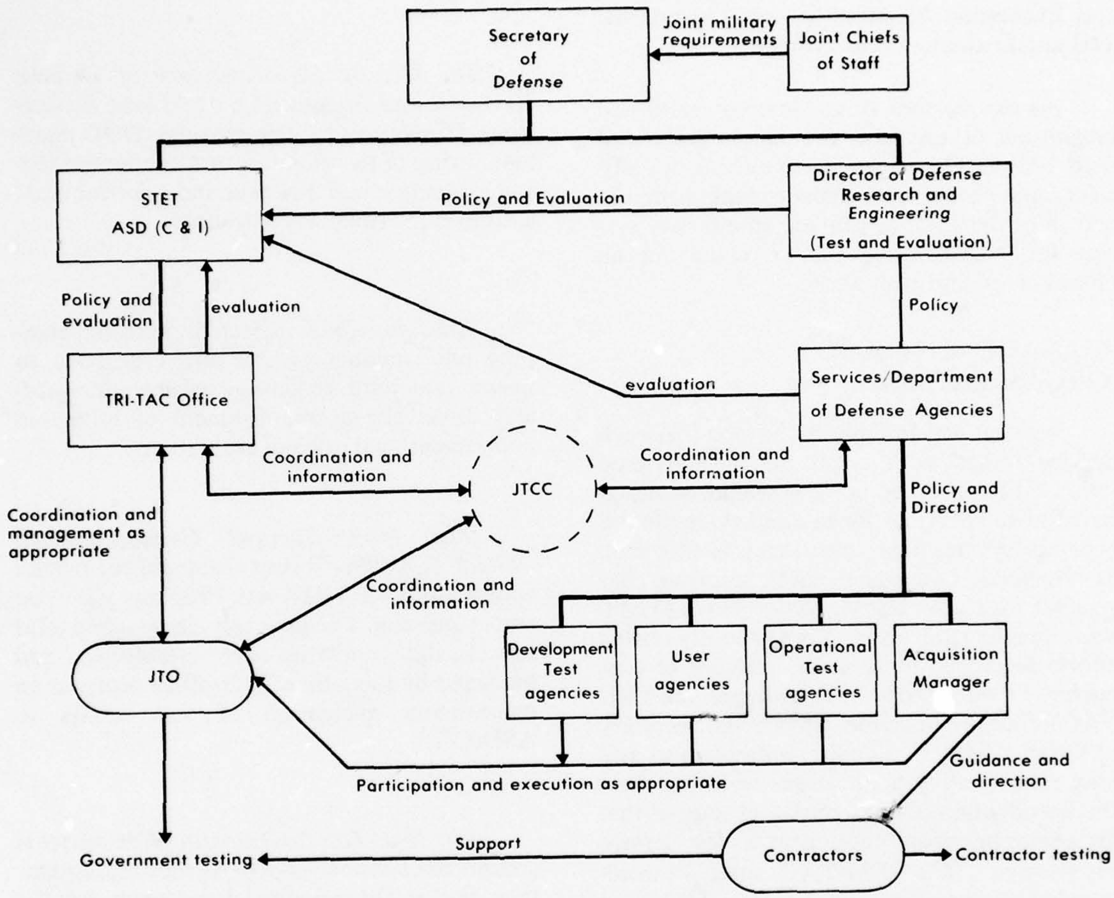


Figure 1. TRI-TAC Program Test and Evaluation Relationships

Production and Deployment Phase – The production and deployment phase commences after the production decision. In this phase, the equipment/system is produced and delivered as an effective operational and supportable commodity for the users. Testing during production is categorized as Production Acceptance Test and Evaluation (Contractor) and Operational Test and Evaluation (Government).

As can be seen from the complexity and magnitude of test and evaluation associated with TRI-TAC sponsored equipment, an efficient and economical management structure had to be designed to provide an effective system for coordinating and/or managing all phases of test and evaluation.

MANAGEMENT AND ORGANIZATION

While it was initially envisioned that each Service tasked with acquiring an individual item of TRI-TAC equipment would be designated "Lead Service" for test and evaluation, it became apparent that considering factors such as timeliness, cost effectiveness, schedule constraints, long term test viability relative to configuration control, and test repeatability testing efforts precluded multiple assignments of concurrent "Lead Service." Accordingly, the TRI-TAC Office provided the Office of the Secretary of Defense with an indepth study of short and long range testing applications resulting from the introduction of modern digital state-of-the-art communications equipment to the Service inventories.[6] The TRI-TAC study recommended the present concept of establishing a permanent TRI-TAC Joint Test Organization and joint test facility for achieving test and evaluation for TRI-TAC sponsored equipment. The recommendation was approved. With the exception of the permanent OSD-cadre of the Joint Test Organization, a special organizational structure is not required to support the TRI-TAC test and evaluation program. However, there are other organizational activities that enter into the test and evaluation arena. See Figure 1. Present responsibilities of the test and evaluation program participants are given here.[2]

The Assistant Secretary of Defense for Communications, Command and Control and Intelligence (ASD (C^3I)) provides DOD staff supervision over the TRI-TAC Program and insures the active participation of relevant DOD components in test and evaluation activities.

The Deputy Undersecretary of Defense Research and Engineering, (Test and Evaluation) DUSDR&E(T&E) provides DOD Staff monitoring of the planning and conduct of test and evaluation and makes an independent evaluation on the adequacy of testing.

The Joint Chiefs of Staff monitor the planning and conduct of test and evaluation to insure that joint requirements are addressed, and direct the accomplishment of follow-on joint operational test and evaluation.

The Joint Tactical Communications (TRI-TAC) Office is responsible for the overall coordination of TRI-TAC Program joint test and evaluation. The joint test element and joint test facility activities are established and managed by this office. This office provides an independent evaluation of test results to ASD(C^3I).

The Joint Test Coordinating Committee is composed of senior Service/Agency representatives and is the principal focal point for the monitoring of all aspects of the test and evaluation program. The Committee arbitrates problems and advises the Director, TRI-TAC Office, on progress of test and evaluation activities.

The Services direct the conduct of development test and evaluation and initial operational test and evaluation; participate actively in all aspects of test and evaluation; and provide an independent evaluation to the ASD(C^3I) and the Joint Chiefs of Staff on the results of testing.

National Security Agency participates in communications security aspects of the test and evaluation program. This Agency independently evaluates test results with respect to the security of the equipment or system under test.

The Defense Communications Agency participates appropriately in all aspects of test and evaluation and independently evaluates test results with respect to the suitability of the equipment under test for use in the Defense Communications Agency.

In support of the test and evaluation effort, a Joint Test Organization was established.[7] The Joint Test Organization consists of the joint test element, facility support element, test support element, Service/Agency test officer for development test and evaluation, Service test officers for initial operational test and evaluation, and liaison officers. See Figure 2. The Joint Test Organization is to serve, primarily as the focal point for the planning, conducting, and reporting of government testing. The Joint Test Organization, with the exception of the joint test element, is composed of resources from the Services/Agencies.

The **Joint Test Element** is a permanent DOD staff reporting directly to the Director, TRI-TAC Office. The joint test element has the responsibility to coordinate the activities of the Joint Test Organization and provides the permanent personnel required to support testing.

The **Facility Support Element** consists of inventory equipment and personnel resources required to maintain the system baseline at the test facility on a continuing basis. The Army will provide the facility support element.

The **Test Support Element** consists of personnel and equipment resources keyed to the execution of a specific test plan for an individual item of equipment. There will be numerous test support elements throughout the test and evaluation program. These elements will be phased in and out of the test facility as required.

The **Service/Agency Test Officers** (development test and evaluation and initial operational test and evaluation officers) are personnel from the Services who act as on-site executive agents for the respective Service test Agencies/Commands on matters relating to test and evaluation. These officers may act as test directors for the individual Service test commands.

Liaison Officers are assigned to the Joint Test Organization as required.

The successful operation of the Joint Test Organization will involve considerable communication, coordination and cooperation among its organizational elements. Primary objectives of the Joint Test Organization are:

- to insure that test and evaluation planning includes sufficient testing of each acquisition item and

- to insure testing in the most economical manner by eliminating duplication and requiring minimum resources to satisfy the desired test objectives.

In addition to the normal preparation of contractor test plans and government development and operational test plans, the Joint Test Organization will prepare an integrated test plan. The integrated test plan will be used primarily as an OSD management-level document based on the development test and evaluation and initial operational test and evaluation plans. The integrated test plan will be used as the DOD-level coordination document for each major acquisition of TRI-TAC sponsored equipment. Moreover, this plan will be the document that identifies overall milestones, schedule and resources; and, accommodates the requirements of other joint test activities conducted simultaneously at the joint test facility. Accordingly, test and evaluation planning must begin early in the acquisition program and encompass both contractor and government test and evaluation.[5] Responsibilities for the specific plans are assigned as outlined here.

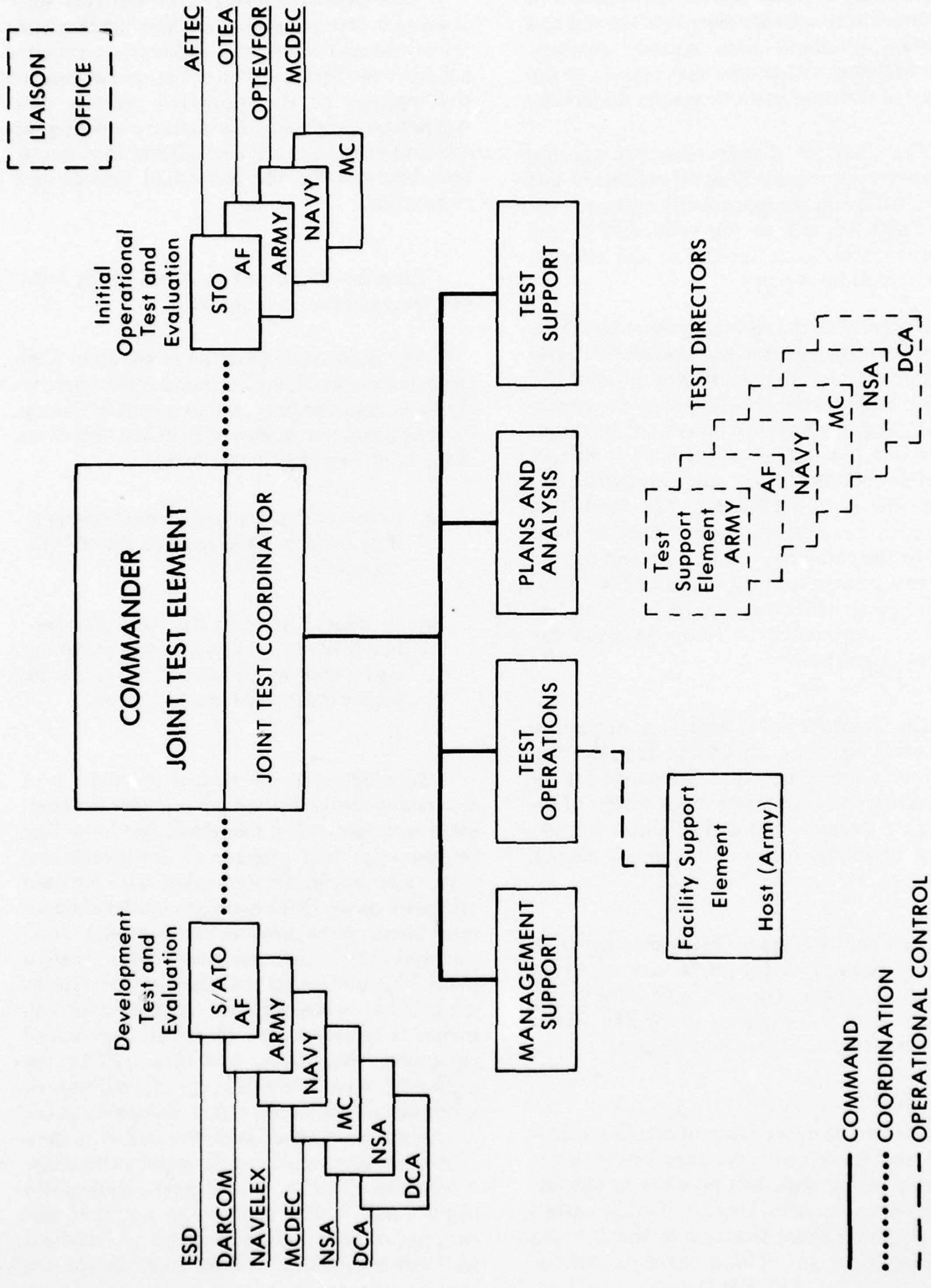


Figure 2. Joint Test Organization Organizational Chart

Contractor Development Test Plans – The contractor prepares the plans and submits them to the acquisition Service/Agency. The acquisition Service/Agency is responsible for coordinating the review of contractor development test plans with the program participants. After review and formal coordination, the plan is approved or disapproved by the acquisition Service/Agency and the contracting officer transmits the decision to the contractor.

Development Test and Evaluation and Initial Operational Test and Evaluation Plans—The acquisition Service/Agency is responsible for preparation and joint coordination of these government plans. Development test and evaluation and initial operational test and evaluation plans will be developed by joint working groups as deemed necessary by the respective testing commands of each acquisition Service. After coordination and concurrence by the program participants, the plans are submitted to the Service/Agency test officer of the Joint Test Organization for distribution within that organization for use as inputs to the integrated test plan. The Director, TRI-TAC Office, reviews the test plans and monitors the progress of all acquisitions to assure adherence to TRI-TAC Program objectives and system design concepts.

Integrated Test Plan—Integrated test plans will be developed by the acquisition Service test officer assisted by the Service test officers of the Joint Test Organization and the joint test element with appropriate inputs from the development test and evaluation and initial operational test and evaluation plans. An integrated test plan will be prepared for each acquisition under test and will address the integration aspects of development test and evaluation and initial operational test and evaluation for that acquisition. Integrated test plans are reviewed and coordinated by the Services/Agencies prior to submission by TRI-TAC to the Undersecretary of Defense Research and Engineering (Test and Evaluation) and the Joint Chiefs of Staff for OSD-level review and coordination.

Production Acceptance Test and Evaluation Plans – The contractor prepares the production acceptance test and evaluation plans and submits them to the acquisition Service/Agency for review and approval. The acquisition Service/Agency is responsible for coordinating the review of these plans with other program participants.

Operational Test and Evaluation Plans—The joint operational test and evaluation plans will be developed by the Services as directed by the Chairman of the Joint Chiefs of Staff. Follow-on operational test and evaluation plans will be developed as directed by the respective Service/Agency. The Deputy Undersecretary of Defense Research and Engineering (Test and Evaluation) will review operational test and evaluation plans to determine plan adequacy. In cases where the facilities of the joint test facilities will be used for operational test and evaluation the Commander of the joint test element will review plans to assure adherence to TRI-TAC Joint Test Program schedules. Joint operational test evaluations will be provided to the TRI-TAC Office indicating the degree to which TRI-TAC sponsored equipment meets joint or Service/Agency operational requirements.

After test plans have been developed (concurrences have been received and the plans completed) the acquisition Service/Agency must prepare for executing those plans. Initial testing is accomplished at the contractor's plant or at government approved facilities. The government will witness all contractor tests to insure that the objectives/requirements identified in the contractor development test plans are satisfied. The acquisition Service/Agency test officers of the Joint Test Organization are responsible for planning the observance of contractor development test by all program participants and insuring that evaluated results are available for use at the next level of test planning. These steps are taken to eliminate redundant testing.

Next, the test items are delivered to the joint test facility where the acquisition

Services/Agencies are responsible for conducting all government test and evaluation on equipment for which they are the acquisition manager. Each acquisition Service/Agency will designate a test director to be responsible for achieving the test objectives with support from the other program participants using the Joint Test Organization elements and the resources at the joint test facility. All resources will be time-phased to insure maximum economy of scale throughout the test period.

The Service/Agency test officers, assisted by the joint test elements, will prepare and submit separate development test and evaluation and initial operational test and evaluation reports for each item under test. Test reports will be provided, for independent evaluation as appropriate, to all program participants, to include the TRI-TAC Office, the Deputy Undersecretary of Defense Research and Engineering (Test and Evaluation), and the Joint Chiefs of Staff.

JOINT TEST SITE AND FACILITIES

The joint test facility is located at Fort Huachuca, Arizona, with an adjunctive Naval Telecommunications System test node at San Diego, California. The centralized test facilities at Fort Huachuca include a central test facility, logistics support test facility, contractor support facility and both local and remote nodes.

Central Test Facility – The central test facility is inclosed by a security fence and consists of a test control center and four collocated test nodes. The test control center consists of operating spaces and facilities required to control testing. The heart of the center is the test control system that is designed to facilitate configuration management at each of the four nodes. The test control system is a semiautomatic switching and patching system that will be used to establish interconnections between various items of equipment both at the central test facility and the nodes. The test control system will interface with the main distribution frame by means of audio cable that carries both analog signals and digital data. The test control system will be transparent to the transmission paths carrying test data.

Three of the four local nodes composing the central test facility are located at hardstand areas approximately 250 feet from the test control center. The fourth node is collocated with the test control center. Each node has lighting and permanent interface facilities to include audio, video, and power distribution systems.

The nodes will be configured as test nodes representing the network configurations required by the various integrated test plans. The separation between nodes is designed primarily to minimize congestion in the control test facility and to reduce the possibility of electromagnetic interference between nodes.

The hardstand and permanent interfaces at each node are to facilitate installation and interconnection of equipment within each node to permit the rapid establishment of various node-to-node trunks through the test control center via the test control system. The disposition and configuration of equipment within individual nodes will be flexible and, except for space limitations, independent of the permanent facilities.

Logistics Support Test Facility – The logistics support test facility will provide a suitable environment for conducting logistics supportability testing, especially at the immediate level of maintenance. Logistics support packages for equipment under test will be stored, used and tested in this facility. The facility will provide administrative, work and storage areas for the acquisition Service logistics support test and evaluation teams. The facility provides for simultaneous accommodation of several different test items during government testing.

Contractor Support Facility – The contract support facility is designed to provide administrative, storage and maintenance work areas in support of contractor items under test.

Radio Relays – The radio relays associated with the joint test facility will serve two different functions. First, the radio relays will permit the establishment of "local node" radio frequency links over distances representative of actual operational conditions when such links are required to provide "electrically realistic" networks. Secondly, during the deployment of equipment to remote operating areas for initial operational test and evaluation, the radio relays will be reoriented as required to provide radio frequency links between the deployed nodes and the central test facility.

Remote Operating Areas – Remote operating areas may be unimproved areas suitable for the deployment of military personnel and equipment to accomplish initial operational test and evaluation, or these areas may be other military installations. In either case, the areas must provide representative environments to simulate expected deployment conditions. The equipment in one or more of the nodes at the central test facility may be disconnected and deployed to one of the designated remote operating areas. The distance of these remote areas from the central test facility will depend upon the desired test scenario.

Naval Telecommunications System Test Node – In support of the joint test and evaluation program the Navy is establishing, as an adjunct to the TRI-TAC joint test facility, a Naval Telecommunications System test node in the San Diego, California area.[7] The objectives of the test node are to provide Navy test and evaluation participation and support to the TRI-TAC Program, to determine the degree of interoperability between the Naval Telecommunications System and subsystems developed under the TRI-TAC Program, and to provide a capability for test and evaluation of unique Naval Telecommunications equipment during the research and development cycle.[8]

As required by approved test plans, test node capabilities will include the ability to duplicate, simulate or otherwise represent end-to-end, any system within the Naval Telecommunications System and will be representative of a mobile fleet unit and/or systems normally found in a Naval Communications Station.

The operation and maintenance of the Naval Telecommunications System test node are the responsibility of the Navy. Personnel utilization for agreed upon test plans will be managed by the Navy in coordination with the other program participants in support of the TRI-TAC Joint Test and Evaluation Program.

ADMINISTRATION AND LOGISTICS

The TRI-TAC Program is unique in that Service resources are pooled for mutual benefit to develop equipment that will satisfy common communications requirements in the tactical arena. No single Service could, practically, afford the investment required to develop the various TRI-TAC sponsored equipment. A significant benefit is realizable because of the life-cycle costs savings inherent in the joint support of these acquisitions after the acquisitions have been fielded.

The responsibilities of the TRI-TAC Program participants for financial management and support of the Joint Test Organization are as follows:

- **TRI-TAC Office** – The TRI-TAC Office, in collaboration with participating DOD components, will determine the resources required to implement the operations of the Joint Test Organization. The office must assure that resource requirements are reflected adequately in the Five-Year-Defense Program and budget documentation prescribed by the Office of the Secretary of Defense for various events in the planning, programming, and budgeting system. In addition, the office must maintain adequate control of all funds made available to support the operations of the joint test element in accordance with current policy and directives.

- **Military Departments/DOD Agencies** – The Military Departments and Agencies will program and budget for resources to support the joint test element in accordance with normal procedures prescribed for the DOD planning, programming and budgeting system. Funds for this support will be apportioned on a 35, 35, 20, 10-percent basis among the Army, Air Force, Navy and the National Security Agency respectively. Joint test element civilian personnel costs will be budgeted by the Army and will be included as part of the apportioned

35 percent.[7] This funding will be provided to the Director, TRI-TAC Office, in a timely manner, consistent with the availability of funds.

The Military Departments and Agencies will program and budget for resources for the test support element and the development test and evaluation and initial operational test and evaluation Service test officers as required.[2] Funds shall be retained by the individual Department/Agency and used to support requirements of the pertinent test planning documents.

The Secretary of the Army will provide logistic and administrative support to all elements of the Joint Test Organization with reimbursement as appropriate. Intra/inter-Service support agreements will be negotiated by the individual Service/Agency with the joint test facility host, Fort Huachuca, as required. In addition, the Secretary of the Army will program funds to operate and maintain the facility support element of the Joint Test Organization.

ADMINISTRATIVE REQUIREMENTS

Prior to the start of development test and evaluation and initial operational test and evaluation the contractor, under the auspices of the acquisition Service for each acquisition item, will install, check out and certify the equipment delivered to the government test site for test. After successful completion of the installation and checkout, the acquisition Service representative for that item shall provide conditional technical acceptance to the contractor. The acquisition Service representative shall insure that all delivery items are complete, to include all test equipment, other supporting equipment, spare assemblies and parts, technical publications to include maintenance and operator's manuals, programmer guides, calibration procedures, test plans, test procedures, data sheets, contractor development test results, baseline drawings, and other contractually required items identified by contractor data requirements lists. Transfer of property control from the acquisition Service to the test directors for the duration of development test and evaluation and initial operational test and evaluation, will be accomplished by utilizing standard military forms/procedures.

Upon completion of development test and evaluation and initial operational test and evaluation, test documentation/assets no longer needed at the test sites for follow-on testing will be returned to the acquisition Service. Designated TRI-TAC sponsored equipment required to form the baseline for other tests will be transferred from the acquisition Services to the facility support element. Property control of test items will be returned to the acquisition Service for disposition when no longer required at the test site.

Physical security for the central test facility and the logistics support test facility will be provided by the joint test element. The acquisition Services will be responsible for providing security for the radio relay sites, remote operating areas, and contractor support facility.

Periodic administrative status reports prepared by the joint test element, and coordinated with the Joint Test Organization will be submitted to the TRI-TAC Office and all test and evaluation participants.

In summary, the test and evaluation for TRI-TAC sponsored equipment must be jointly planned, conducted and reported by the Services and Agencies in coordination with the TRI-TAC Office. Each acquisition must be witnessed/observed by the Services and Agencies during contractor testing to assure that the item is ready for government testing. Each tasked acquisition Service/Agency must direct the conduct of government testing in a joint arena and insure that adequate development test and evaluation and initial operational test and evaluation reports are prepared so that each Service, Agency, the TRI-TAC Office, the Deputy Undersecretary of Defense Research and Engineering (Test and Evaluation) and the Joint Chiefs of Staff can prepare independent evaluations for use by the decisionmakers. The TRI-TAC Office must insure that DOD-level visibility of the test and evaluation program is achieved by timely submission of integrated test plans for each acquisition. Thus, it can be seen that a successful test and evaluation program is dependent upon the communication, cooperation, coordination and support of all TRI-TAC Program participants.

REFERENCES

1. **Dept of Defense, "The Joint Tactical Communications (TRI-TAC) Program," Directive, Number 5148.7, 16 Feb 76.**

2. **________, "Implementing Instructions for Test and Evaluation of TRI-TAC Sponsored Equipment," Instruction, Number 5148.8, 9 Apr 76.**

3. **________, "Acquisition of Major Defense Systems," Directive, Number 5000.1, 18 Jan 77.**

4. **________, "Test and Evaluation," Directive, Number 5000.3, 19 Jan 73.**

5. **________, "TRI-TAC Test and Evaluation Master Plan (TEMP), TTO-ORT-059-77, Joint Tactical Communications Office, Fort Monmouth, NJ, Mar 77.**

6. **________, "TRI-TAC Program Joint Testing," TTO-ORT-034-074, Joint Tactical Communications Office, Fort Monmouth, NJ, Oct 74.**

7. **________, Deputy Secretary of Defense to Joint TRI-TAC Communications Office, Memorandum, subject: "Joint Test Facility and Organization for TRI-TAC Program, 12 Mar 76.**

8. **Dept of Navy, "Master Plan for TRI-TAC/Naval Telecommunications System Test Node," NAVSEC 6178C02 TT4, 6 Feb 76.**

Lt Col David L. Washington, US Air Force, is assigned as an Electronics Engineer with the Test and Analysis Division, Operations Research, Test and Analysis Directorate, Joint Tactical Communications (TRI-TAC) Office, Fort Monmouth, New Jersey. He has served as an exchange Communications Officer with the Navy; Communications Electronics Staff Officer with the 9th Air Force and Chief, Maintenance Division with the 505th Tactical Control Group in Viet Nam.

Lt Col Washington holds a BS in Electrical Engineering from the North Carolina Agricultural and Technical State University, an MS in Electrical Engineering from San Jose College and he is a graduate of the Air Command and Staff College.

HOW MUCH TESTING IS ENOUGH?

by

Howard W. Kreiner, Systems Planning Corp.

When program managers and Service staffs are working with the Office of the Undersecretary of Defense Research and Engineering, a question often heard is, "How much testing is enough"? Test, of course, covers so wide a spectrum of activities that almost everything that is not actual hardware fabrication could be somehow included. In our context the subject is the development and operational testing of total systems, and/or major subsystems. This testing must be planned, conducted, and evaluated before the key decisions to enter full scale development, and later, production of the system can be done.

AN ANSWER

The answer to the question, "How much testing is enough," must take into account the amount of testing that safely can be deferred until

- after the production decision;
- during initial production; and,
- after field deployment of the system.

The simple but unsatisfying answer to the question is: There has been enough testing when you can make a satisfactory estimate of the risks of proceeding into full scale development, or into production, and are able to balance these risks against the effects on US defenses of program cancellation. The answer is unsatisfying. First, a budget cannot be framed from that answer. Second, the Initial Operational Capability (IOC) date cannot be calculated from the answer. Moreover, the answer has a disturbingly subjective tone. The implication is that judgment about the total amount of testing is not made until test results are in hand, rather than at the start of the program.

The nature of the answer, with its high degree of subjectivity, makes it difficult to resist pressures to eliminate or minimize the required predecision test phase.

The pressures that operate against comprehensive testing are obvious and measurable. The longer the period between prototype completion and production initiation, the longer and more costly the R&D phase, including testing, will be. Production facilities and staff require funds for maintenance even if production is not occurring. The inflation meter runs while time passes. Unit costs rise in a manner that currently results in reduced total buys of equipment. And most evident, equipment is delayed in entering service.

In favor of deferred testing with its apparent savings of time and money it is noted that service introduction leads to in-service accumulation of operating hours at a faster rate, lower cost, and with a realism that prototype R&D testing lacks. Service introduction of a system does require consideration insofar as testing is performed to obtain intimate familiarity with the operational capabilities of a system.

But it should not be necessary to argue the merits of a reasonable amount of initial operational system testing. Against all the pressures to reduce Operational Test and Evaluation, there is the balancing risk of committing the Department of Defense to the purchase of unsatisfactory equipment. The reasons for testing have been affirmed at length in the report of the Presidential Blue Ribbon Defense Panel whose members were appointed by the President of the United States. These reasons have been reinforced at intervals by reports of the Defense Science Board Task Force on Test and Evaluation (April 1974 and February 1977). What is required is a clearer definition of the operational meaning of words like "satisfactory," "reasonable," and "risk."

Of these terms, the most difficult word to define is "risk." "Risk" in this context is unmistakably subjective, for the risk is a subjective judgment as to whether or not any of a myriad of unwelcome outcomes will occur in consequence of a decision to proceed. As stated earlier, the counter arguments to proceeding deliberately tend to have more weight. These counter arguments can be made definite and quantitative, i.e., time costs money. The decision to proceed takes one into qualitative judgment that no amount of cosmetic language, such as "subjective probabilities," can turn into quantitative measurement. There are some quantitative aspects: Engine qualification specifications, fatigue test practices, reliability design methods and standards all are risk reducing practices that enable quantitative measurement. But the task of risk *assessment* remains. Without initial operational test and evaluation a method does not exist for ascertaining or for *measuring in advance the probability that* a design will be faulty to a severe degree. Example: How, without operational test could you determine that a design would be faulty to such an extent that 4400 trucks would have to go directly from production line to repair depot for storage to await a redesigned transmission? Good design can insure that parts known to need regular servicing are quickly and easily accessible. Design cannot be expected to take account of a part breakdown that occurs more frequently in the operational environment than anticipated, and is a part that requires major dissassembly to reach.

The judgment of risk and test adequacy cannot be made until test results are in hand. In the real world a manager or decision maker would not allow a judgment, formed by matching test results to the criteria proposed at the start of the program, to stand without review when test results become available. Circumstances may change between establishing the criteria and getting test results or, the implications of the circumstances may be differently perceived. For example, to a decision maker at program initiation the prospect of a 30 percent cost overrun may appear unacceptable. Cost overrun may prove unavoidable, unpleasant, but nonetheless acceptable if, after the program test results are evaluated another course of action does not offer a better alternative. To a manager faced with getting a complex piece of equipment built and working, a possible durability limitation with a propulsion unit may seem a manageable problem for solution by a production team. After the euphoria of getting the equipment working has passed, a public debate on the durability problem might be something the same program manager would prefer to avoid until after a further serious effort is made to solve the "manageable" problem and to confirm the solution.

ACHIEVEMENT-PACED SCHEDULING

The after-the-fact character of decisions based on test results translates into a well-known, but infrequently implemented policy of: *Achievement-paced scheduling.* The difficulties of this policy include all the reasons described earlier for deferring test, *and* the well-known problems of making an achievement-paced program fit the constraints of a yearly budget cycle. The difficulties are nowhere more evident than when achievement-paced scheduling is applied to system reliability.

A method has not been found to eliminate the difficulties, but there is one technique that can mitigate them. The method is to develop and organize experience into meaningful checklists of essential test actions. Comprehensive check lists were published and distributed widely by the DD(T&E) in conjunction with the report of the Defense Science Board Task Force on Test and Evaluation in April 1974. The series of checklists was presented in nine volumes, in addition to the general checklists and software test proposals contained in the Task Force report. The nine volumes addressed individual generic system types: aircraft, missiles, ships, ground vehicles, antisubmarine warfare systems, airborne electronic countermeasures systems, airborne general surveillance radar systems, command and control systems, and common test gear. These lists remain applicable and can assist in framing and putting a workable scale on the answer to the question: "How much testing is enough"?

It is essential that the fundamentally subjective, indefinite answer be remembered: There is enough testing when it is possible to make a satisfactory estimate of the risks of proceeding into full scale development, or into production, to balance against the effects on US defense of cancelling the program. This concept needs adequate weight in the program development process, with special attention devoted to the identification of essential test actions. Without successful concept accomplishment and identification of essential test actions one should not proceed to the next major phase of the program.

Mr. Howard W. Kreiner is a Staff Member with System Planning Corp. From 1972 until mid-1977 he was a Staff Member in the office then designated as Office of the Deputy Director, Defense Research and Engineering (Test and Evaluation). As a Staff Member, Mr. Kreiner served as Executive Secretary, Defense Science Board Task Force on Test and Evaluation (1973–1974). He again held this position in 1976 when the Task Force was reconstituted for a renewed look at Test and Evaluation policy.

Since 1947, Mr. Kreiner has served as an Operations Research Analyst. He worked with the Navy Operations Evaluation Group and its successor, the Center for Naval Analyses, until December 1966. Assignments included a 1-year tenure with the Air Test and Evaluation Squadron 1 (VX-1) plus assignments at: Headquarters, Office of the Commander, Operational Test and Evaluation Force and, Office of the Commander in Chief, Atlantic Fleet. In addition, Mr. Kreiner has served under the Commanders of the Sixth and Seventh Fleets.

Mr. Kreiner is a founding member of the Operations Research Society of America. He received his BS in Chemical Engineering from the University of Pennsylvania (1944), and his MS in Mathematics from MIT (1954).

TEST AND EVALUATION POLICY

by

Charles W. Karns, Staff Assistant
Office of the Undersecretary of Defense Research & Engineering

The large sums spent each year by the Department of Defense to maintain an adequate military defense of the United States require valid decisions relative to the systems proposed for acquisition. Those decisions in the system acquisition process must be based on valid test and evaluation (T&E). Two primary objectives of current DOD T&E programs are to measure, accurately, the operational characteristics of new weapon systems and to make a realistic evaluation of operational effectiveness and suitability of these systems.

INTRODUCTION

The estimated $2.7 billion that the Department of Defense (DOD) will spend on test and evaluation activities in FY 1978, while only a fraction of total defense expenditures, is a large sum.

There are two principal types of test and evaluation conducted on new military systems:

- **Development Test and Evaluation (DT&E).** Development Test and Evaluation is conducted by, or under the supervision of, a development agency to evaluate the technical performance of prototype equipment. This testing is generally conducted by skilled technicians and engineers under controlled conditions.

- **Operational Test and Evaluation (OT&E).** Operational Test and Evaluation is conducted by military personnel to determine the degree to which new equipment fulfills military operational requirements. This type of testing is conducted under conditions that closely duplicate the environment expected in field operations. Operational testing includes maintenance and logistics support, operation of the equipment, and, where appropriate, the evaluation of tactics and techniques for the most effective system use in combat. Operational tests are conducted on research and development prototypes of new equipment as well as on early production models.

The DOD officials who make decisions regarding the acquisition of new systems depend upon results of test and evaluation. Specifically, *test and evaluation provides information* on

- how well a program is meeting current objectives and,

- what the ultimate outcome of the program is likely to be.

Generally, Development Test and Evaluation most often provides information of the first kind by indicating whether a system under development is meeting the technical specifications. Operational Test and Evaluation usually provides information about how well a system will perform once it becomes operational.

Although decisionmakers require other types of information, the information provided by DT&E and OT&E is essential to most major decisions made during the course of a program. Since testing is expensive, a reasoned judgment must be made as to how much testing is needed, or how much risk can be accepted in making program decisions. The factors that influence the judgment about how much testing is needed vary from program to program. Hence, the amount of test and evaluation to be accomplished may also vary.

Aside from being essential for decision-making, test and evaluation is important in other ways. Development testing, for example, is an integral part of the system design and development process. In this process a system design is converted into hardware. The hardware is tested, deficiencies are noted, and the design is modified as necessary. This process is iterated until the system hardware reaches a final design configuration. Development Test and Evaluation provides the system designer with feedback information that would be difficult or impossible to obtain by analytical processes.

Operational testing has a direct relation to the military capabilities of operating forces. The more effort that is made to identify and correct operational deficiencies before the design of a system is frozen, the greater the military capability of the operating forces when the new weapons system is deployed. Also significant cost benefits may be realized when major system deficiencies are corrected prior to operational deployment. Detection and correction of deficiencies prior to deployment is usually less expensive and less time-consuming than is retrofitting systems after the system has entered the operational inventory. Also, such detection and correction is less disruptive to essential activities engaged in during deployment, such as maintaining a state of operational readiness, conducting training, and developing tactics.

DOD TEST AND EVALUATION ORGANIZATION

In 1970 the Presidential Blue Ribbon Defense Panel reported that operational test and evaluation in the Department of Defense had been infrequent, poorly designed and executed, and generally inadequate. In response, test and evaluation activities were restructured—greater emphasis was placed on the role of test and evaluation in the weapons system acquisition process. The recommendations of this Panel were subsequently strengthened by the Commission on Government Procurement.

In response to the recommendations of these groups, the DOD:

- Created a higher-than-Service organization, specifically the Office, Deputy Director of Defense Research and Engineering (Test and Evaluation) DD(T&E), to oversee Defense test and evaluation activities.

- Established within the Office, DD(T&E) a capability to initiate and conduct operational tests and evaluation in joint operations, that is, test and evaluation involving more than one Service.

- Strengthened the Services' operational testing of newly developed systems and required that the performance of the systems be tested (in an environment approximating the expected operational conditions), and evaluated prior to commitment of the system to full production.

Within each of the Military Departments there has been established an operational test agency, independent of both the developer and user. The operational test agencies are charged with conducting operational evaluations of new equipment.

DSARC PROCESS

The current approach to managing weapons systems acquisition in DOD is based on the milestone concept. At each significant milestone in the life of a major weapon acquisition program its progress is carefully reviewed to determine if the program should be advanced to its next phase. These reviews are conducted at four major program milestones.

AD-A050 752 DEFENSE SYSTEMS MANAGEMENT COLL FORT BELVOIR VA F/G 5/1
DEFENSE SYSTEMS MANAGEMENT REVIEW. VOLUME I, NUMBER 5.(U)
1977 W E LOTZ, I E DONOVAN, P W POWERS

UNCLASSIFIED NL

2 OF 2

AD
A050 752

END
DATE
FILMED
4-78
DDC

Milestone 0 marks the point at which exploration of alternative solutions to a defined mission need is begun and the program is initiated.

Milestone I is the point where one or more system alternatives are selected for demonstration and validation of ability to fulfill the defined need.

Milestone II is where a decision is made on whether to enter full-scale development of the system.

Milestone III is the final milestone. Here a decision is made on whether to initiate full-scale production and deployment of the system.

In addition to these fixed milestones, program reviews may be held at any point in the life of a program when major issues arise or when the program breaches or threatens to breach a prescribed cost, schedule, or performance threshold. Most information used at reviews to assess program progress and likely success is derived from test and evaluation results. Depending on the assessments made at the reviews, a program may be advanced to its next phase, cancelled, reoriented, or held in its present phase.

Major program reviews conducted at the OSD level are the responsibility of the Defense Systems Acquisition Review Council (DSARC). The DSARC reports its recommendations directly to the Secretary of Defense, who makes the final decision on whether the program is to be advanced to its next phase. The principal members of the DSARC have been OSD officials at the Assistant Secretary level—the Director of Defense Research and Engineering, the Comptroller, the Assistant Secretary of Installations and Logistics, and the Director of Planning and Evaluation. Although some changes in the precise membership of the DSARC will follow the recent OSD reorganization, membership will consist of those officials responsible for the functional areas represented in the past. The Deputy Director of Research and Engineering (Test and Evaluation), not a DSARC principal, normally participates in the Council's deliberations and provides to its members an assessment of the testing already completed and the adequacy of the plans for tests to be conducted in the future. Thereafter, the DD(T&E) submits a written assessment directly to the Secretary of Defense for use in reviewing DSARC recommendations.

A new acquisition program may be initiated (at Milestone 0) when a DOD Component Head perceives the existence of a mission need and, having submitted a statement of that mission need to the Secretary of Defense, requests approval to identify and explore alternative solutions.

A Decision Coordinating Paper, prepared for use at the time of Demonstration and Validation Decision (Milestone I), identifies critical issues and areas of risk to be addressed by test and evaluation. The Decision Coordinating Paper provides the test objectives and measures of effectiveness related to satisfaction of mission need, and a summary of schedules and resource requirements applicable to test activity prior to Milestone II. The DSARC (or the Systems Acquisition Review Council (SARC) in one of the Services for those programs delegated to it) determines the adequacy of the critical issues, test objectives, and test schedules.

When full-scale engineering development is proposed (Milestone II), the revised Decision Coordinating Paper gives:

- the results of test and evaluation accomplished to date;
- an updated statement of critical issues, test objectives, and areas of risk needing further assessment;
- a summary of performance criteria goals and thresholds;
- an overview of test plans, milestones, and program interrelation.

Supporting details of test plans and test results are made available as requested by the Deputy

Director (T&E). The DSARC (or designated Service Council) assesses and comments to the Secretary of Defense as to the adequacy of test and evaluation progress, and of the test and evaluation planned to occur prior to the first major production decision.

At Milestone III results of testing and plans for future testing are again assessed, and the DSARC (or designated Service Council) comments to the Secretary of Defense as to the adequacy of these test results to support a decision relative to full scale system production and deployment.

Because test and evaluation is essential to the decisionmaking process, it is important that the Deputy Director (T&E) be independent and objective in making assessments. Current DOD directives provide the means by which such independence and objectivity can be exercised. Example: Even though the Director of Defense Research and Engineering is the direct superior of the Deputy Director (T&E), the latter has direct access to the Secretary of Defense, particularly when weapons system acquisition decisions are to be made. In this respect, the position of the Deputy Director (T&E) is very similar to that of the heads of the independent test agencies within the Services.

NOTE: Involvement of the Deputy Director (T&E) in a program is not confined to the limited period of time required to prepare for the conduct DSARC reviews. If the DD(T&E) were to wait until the meeting of the DSARC to report that the testing of a system was unsatisfactory, unnecessary additional program costs could accrue. A loss of valuable time could result because of the need to develop new test models, or modify existing ones, and the requirement to repeat tests. It is most effective for the DD(T&E) staff to be involved in tests on a daily, weekly and monthly basis—from the test planning through the final evaluation of test results. The action officers in the office of the Deputy Director (T&E) bring to bear a military operational background and experience gained from observing large numbers of tests on a wide variety of equipments, as well as an objectivity nurtured by remoteness from the development process. By reviewing test plans before the tests are initiated and by observing results as the test proceed and even actively participating in certain tests, the Deputy Director (T&E) can advise a Program Manager early in the development cycle as to whether or not the equipment is advancing toward meeting the standards by which the weapon system will be judged. Thereby time is allowed for corrective action to be taken prior to the milestone decision point.

RECENT DEVELOPMENTS

The office of the Deputy Director (T&E) is constantly trying to improve the quality of test and evaluation information available for acquisition decisionmaking. Changes are being made in test and evaluation policies and directives to emphasize the need for operational testing early in the weapon system development cycle. In the early 1970's emphasis was placed almost exclusively on conducting Operational Test and Evaluation just prior to the Milestone III major production decision. This was because of the large procurement funding commitment made at Milestone III. Lately, it has been realized that earlier Operational Test and Evaluation may be needed for four reasons:

- In considering system concepts during the early stages of a program, it is necessary to know if the concepts are technically feasible, and operationally viable, and tactically sound.

- To have impact on the research and development phase of a program, Operational Test and Evaluation must be conducted much earlier than at DSARC III. By DSARC III most research and development resources will have been expended.

- The earlier in the development cycle a deficiency is identified and corrected, the less the cost of corrective action.

- A trend toward initiating limited production before the major production decision results in an earlier commitment of procurement funds.

For these reasons, emphasis now is on conduct of Operational Test and Evaluation as early as possible in the acquisition cycle. Early OT&E insures that relevant test information is available for decisionmaking during the early development stages of a program.

A particular area in which test and evaluation is making a significant contribution is that of equipment reliability. This is not a new area of interest. The Services have long concentrated on reliability and have given close attention to the achievement of improved reliability through proper design and through quality control in the production process. The Office, Deputy Director (T&E) is in full agreement with these efforts, and works cooperatively with the Service development and materiel agencies that carry out design and quality control actions. Test and Evaluation plays a major role in the measurement of reliability and the isolation of failures. The actual correction of defects (for the most part) is accomplished by redesign of the equipment or improved quality control in equipment manufacture.

Laboratory and bench tests of components and assemblies by the contractor are the mainstay of reliability testing and form the basis for proving that equipment complies with contract specifications. Often such tests are required to further pinpoint and analyze field failures so that engineering corrections can be devised.

What laboratory and bench tests do not do is expose the system to:

- the full range of operational stresses;
- the harder handling of less well-trained military operation and maintenance personnel;
- the cumulative effect of harsh field conditions.

If reliability weaknesses are to be discovered before large numbers of equipment are deployed, the equipment must be tested under typical operational conditions. Therefore, the Services now are required to specify reliability requirements in operational terms. These requirements are to be measured under operational conditions by operational personnel.

A limiting factor in operational testing is that of obtaining sufficient hours of system operation under the very conditions that make those hours most costly; that is, full system operation in an actual field environment. To augment these valuable hours it is necessary to insure, to the extent possible, that development testing of new systems is done under realistic conditions. The failure data then can be combined with the results of operational testing, to give a comprehensive, yet the most economical method of evaluating weapon system reliability.

Closely allied with reliability is production quality control and quality assurance. The estimate of reliability that can be achieved, based on prototype testing, will not be realized without adequate production line procedures and tests. Although production engineering and quality control are not functions of the Office of the Deputy Director (T&E), close coordination must be exercised with those who are charged with performing those functions. The early equipment production models will be subjected to follow-on OT&E, that is, operational test of production models in a tactical environment. Both those responsible for test and evaluation and those responsible for production engineering expect the production models to exhibit maturity.

Reliability is just one of the measures of system performance for which concern must be expressed. Regardless of how performance is measured, a continuing goal must be early detection of major system deficiencies if such deficiencies exist. Early detection and correction of deficiencies can result in more operating force capability and less likelihood of costly retrofit programs.

Although test and evaluation is making a significant contribution to the weapons systems acquisition process, it is the job of the Deputy Director (T&E) to insure that unnecessary tests are not conducted. There is a point at which the incremental value of information obtained by additional testing is not worth the expense. To the extent possible, each test conducted must contribute to the knowledge of the essential capability of a system in rough proportion to the resources committed to the test.

In 1976 the Director of Defense Research and Engineering asked the Defense Science Board to form a Task Force, under the Chairmanship of Dr. Eugene Fubini, to look into the questions of emphasis placed on test and evaluation and the amount of time and money devoted to these activities. This Task Force affirmed the fundamental and integral character of test and evaluation in the system acquisition process. The major test and evaluation system elements, that is, the DSARC and its interfaces with the Office of the Deputy Director (T&E), were found to function well.

In the view of the Task Force, under applicable DOD Directives, there appears to be little or no overtesting done. The testing done contributes its full value to the improvement and verification of system performance. It was concluded that the management overview provided by the Office of the Deputy Director (T&E) adds to the completeness and quality of system testing. The Task Force believes that the Deputy Director (T&E) does in fact apply the precepts recommended by an earlier Defense Science Board Task Force (1974).

The main concern of the Task Force is that within DOD a mechanism does not exist to insure that risk assessments are made on programs wherein the test results are not altogether satisfactory and correction of deficiencies is indicated. The Task Force recognized that such risk assessments go beyond the responsibilities of the Deputy Director (T&E).

The Task Force was concerned that rigid application of test and evaluation policy directives sometimes leads to the so-called "T&E Gap."* The report restated the alternatives for avoiding the gap proposed by the earlier Defense Science Board Task Force on Test and Evaluation. These alternatives are

- early provisions for additional research and development hardware, to be research and development funded and built for Initial Operational Test and Evaluation.
- an additional phase of testing to cover the T&E Gap;
- advanced availability of production items through early availability of long lead time production funding and low rate initial or pilot production funding.

Response to the Task Force's recommendations is contained in a forthcoming reissuance of DOD Directive 5000.3, "Test and Evaluation." The DOD Directive 5000.3 is responsive also to revised DOD Directive 5000.1, "Major System Acquisition," and DOD Directive 5000.2, "Major System Acquisition Process." Significant changes in these acquisition policy directives that impact on test and evaluation are the establishment of the Milestone 0 Program Initiation Decision and the creation of a Systems Acquisition Review Council (SARC) in each of the Military Departments. These councils, referred to as Service SARC, advise the respective Secretaries of the Military Departments in a manner equivalent to the DSARC advising the Secretary of Defense.

To keep acquisition programs in close consonance with mission needs,* the Secretary of Defense now requires submission of a Mission Element Need Statement (MENS) when a mission need is perceived to exist and a new capability is required to meet that need. Milestone 0 is marked by the approval of the MENS and the direction to one or more DOD components to systematically and progressively explore and develop alternative system concepts to satisfy the approved need. After this, when appropriate and prior to Milestone I, test and

*A break in testing between the end of the development phase and the beginning of production (caused by the requirement for completion of adequate testing prior to a production decision).

*Required by OMB Circular A-109.

evaluation is accomplished to assist in selecting preferred alternative system concepts and to assess the operational impact of the candidate technical approaches.

Under the procedures set by revised Directives 5000.1 and 5000.2, the DSARC will review those programs classed as strategic, nuclear, joint-Service, multinational, intelligence, or communications, command, and control, at Milestone I. The DSARC will review all major programs at Milestones II and III. In cases where a DSARC review is not held, and in most cases where a DSARC is held, the cognizant Service SARC is to advise the Service Secretary relative to the Secretary's recommendations to the DSARC or to the Defense Acquisition Executive and the Secretary of Defense. Test and evaluation inputs to the Service SARC are to be made by development and operational test and evaluation agencies, in much the same manner as the Deputy Director (T&E) supports the DSARC.

Another significant positive effect on test and evaluation arising from the revised DOD directives on acquisition is the provision for the Secretary of Defense, when he approves the selection of a system for full-scale engineering development at Milestone II, to authorize procurement of long lead production items and limited production of items for operational test and evaluation. This is consistent with the recommendations of the Defense Science Board Task Force.

In summary, recent changes in test and evaluation policy have been designed to provide effective inputs to decisions concerning acquisition of major systems. The purpose of test and evaluation is to evaluate, under conditions as realistic as possible, the operational effectiveness and operational suitability of new systems. Possible serious deficiencies are to be highlighted at early stages of development to avoid costly remedies, otherwise required in later stages. The Deputy Director (T&E) and the entire test and evaluation community are continuing efforts to provide valid inputs to the DOD decisionmakers—decisionmakers who must make valid decisions concerning the acquisition of many expensive systems for the defense of the country.

Mr. Charles W. Karns is a civilian Staff Assistant in the Office of the Undersecretary of Defense Research and Engineering. He has been in his present position for the past 6 years. Mr. Karns is a former staff member, Center for Naval Analyses and its predecessor, the Operations Evaluation Group. In addition to duty in the Washington area, he served on field assignments with Naval Staffs in Norfolk, Virginia, Hawaii, and the Mediterranean. Mr. Karns received a BA from Dickinson College and an MA from Northwestern University.

AIRCRAFT OPERATIONAL TEST AND EVALUATION

by

Lt Col Michael J. Butchko, Dept of the Air Force

In this article the author presents the evolution of operational test and evaluation (OT&E) within the Department of Defense (DOD) and the Air Force. Recommendations are provided for the Program Manager that should ensure that valid operational test and evaluation requirements are in the development programs. Emphasis is placed on initial operational test and evaluation (IOT&E)—that portion of operational test and evaluation accomplished prior to the initial production decision—and the role of the Air Force Test and Evaluation Center (AFTEC).

BACKGROUND

Flight test of powered, heavier-than-air aircraft has been going on in the United States since 17 December 1903, when the Wright Brothers flew what was probably the shortest successful test flight on record. In February, 1977, the Space Shuttle, piggybacked on a Boeing 747 made its first test flight. Both events were unique—both events were exciting. The first event represents the efforts of two extraordinary men. The latter event represents efforts of thousands of persons. Between these two aviation milestones, hundreds of aircraft have gone through extensive flight test programs with varying degrees of success. While test program success does not necessarily result in the production of a new aircraft, an unsuccessful test program just about guarantees that a plane will not be produced.

The test program has accomplished its purpose if it verifies that the system under development is ready for production. Also, the test program has accomplished its purpose if it identifies a major fault that leads to program cancellation. If significant problems surface after an aircraft is in the operational units, the Air Force is accused of fielding an underdeveloped system. However, as the time between Milestone 0 (Program Initiation) and Initial Operational Capability (IOC) grows longer and more expensive, the Air Force is accused of testing to obsolescence. The pendulum swings back and forth. The Program Manager, the man charged with *the development of the system*, has the responsibility to ensure that the program has an efficient and effective test program.

Here the author investigates the evolution of operational test and evaluation (OT&E) policy in the Air Force aircraft acquisition process. The methodology employed traces Department of Defense and Air Force policy on test and evaluation since 1970 with particular attention to the proposed revision of DOD Directive 5000.3.

DEFINITIONS

To preclude confusion, Development Test and Evaluation (DT&E) and Operational Test and Evaluation (OT&E) are defined as per *DRAFT* DOD Directive 5000.3 dtd 12 July 1977.

- **"DT&E is that test and evaluation conducted to assist the engineering design and development process and verify attainment of technical performance specification and objectives. DT&E is normally managed and monitored by the DOD components' developing agency...."**

- **"OT&E is that test and evaluation conducted to estimate a system's operational effectiveness and operational suitability...."**

OT&E POLICY 1970–1976

The policies governing test and evaluation have changed over the years in response to changing needs and in response to a variety of internal and external criticisms. Since 1970, the majority of policy changes and criticisms have been directed toward the operational aspects of testing.

In 1970 the flight test portion of the aircraft acquisition process was separated into three categories. Category I (CAT I), ***Subsystem Development Test and Evaluation***, was generally contractor conducted and was totally developmental in nature. Air Force Regulation 80–14 defined CAT I as "development testing and evaluation of the individual components, subsystems, and in certain cases the complete system.[1] There were Air Force Preliminary Evaluations but these too were development test and evaluation oriented.

In Category II (CAT II), ***System Development Test and Evaluation,*** the participants changed somewhat but the emphasis did not. The Air Force, through Air Force Systems Command (AFSC), had the lead role with the contractor still heavily involved but in a secondary role. Air Force Regulation 80–14 defined CAT II as "test and evaluation spanning the integration of subsystems into a complete system in as near an operational configuration as practicable." The using command became involved during CAT II by assigning the Deputy Test Director and by participating in "hands-on" maintenance with the contractor at the test site. However, CAT II was seldom "operational in nature."[2] The production decision for the system was generally made during CAT II.

Category III, ***System Operational Test and Evaluation,*** was the responsibility of the using command. It included "...all components, support items, personnel skills, technical data and procedures...under as near operational conditions as practicable."[1] Since the tests were conducted on the production item after delivery to the using command, "such testing was to determine how best to operate a system, rather than to provide data for decisionmakers to use in determining whether or not to acquire a system."[3]

Such was the situation when the Blue Ribbon Defense Panel made its "Report to the President and the Secretary of Defense" on 1 July 1970. The Executive Summary statement relative to operational test and evaluation was direct and to the point. "Operational test and evaluation has been too infrequent, poorly designed and executed, and generally inadequate."[4] The development test and evaluation was "well understood and faithfully executed...(and)...not considered to be a major problem area." The report went on to point out that operational test and evaluation was not just physical testing but that "To be effective, OT&E must be a total process, using all appropriate methods of evaluation, which spans the entire cycle of a system from initial requirement until it is phased out of the operational forces."[4]

Referring specifically to the Air Force, the panel stated:

> "There are three principal problems with Air Force OT&E, as currently done. First, operational considerations receive much too little attention in Categories I and II. Second, the operational commands responsible for Category III and Operational Employment Testing lack both the

personnel and facilities to be effective. Finally, all of the categories are too duplicative and time-consuming."[4]

The panel made sweeping recommendations on organization within the Office of the Secretary of Defense (OSD). The most significant recommendation, from the testing standpoint, concerned the independent Operational Test and Evaluation agencies. The wording is somewhat indirect but the meaning is clear. "The Secretary of Defense should communicate to the Military Departments...his conviction (that) the cause of effective OT&E is best served when independent OT&E organizations report directly to the Chiefs of Service, Service Secretaries or both."[2]

The DOD response was immediate and positive. The then Secretary of Defense, The Honorable Melvin Laird asked the Services to forward individual proposals for meeting operational test and evaluation requirements to him not later than 1 September 1970.[5] The results were inconclusive and resulted in a memorandum from the then Deputy Secretary of Defense David Packard on 11 February 1971. He wanted an operational test and evaluation "...agency which is separate and distinct from the developing command and which reports the results of its T&E efforts directly to the Chief of Service."[6]

The Air Force response did not establish the independent organization but it did make significant changes. The concept of categories was eliminated and replaced with the DT&E/OT&E philosophy. Also introduced was initial operational test and evaluation to be conducted prior to the production decision. Active participation by the using and supporting commands was incorporated and the user was required to provide an initial operational test and evaluation report to the production decision process.[7]

Parallel with the memorandum-response cycle discussed, the first DOD Directive 5000.1, "Acquisition of Major Defense Systems," dated July 31, 1971, was prepared and released. The purpose of DODD 5000.1 was to provide formal acquisition policy within DOD, to establish the Defense Systems Acquisition Review Council (DSARC) and, to require the Development Concept Paper (DCP) for major programs. Mention of direct user involvement prior to DSARC II (Full-Scale Development decision) was not made nor was there any requirement for an independent operational test agency. Note that this directive followed by 5 months Secretary Packard's memorandum calling for an independent agency. However, the Directive was explicit on the role and importance of the operational assessment:

> "Test and evaluation shall commence as early as possible. A determination of operational suitability, including logistic support requirements, will be made prior to large-scale production commitments, making 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 the DSARC at the time of the production decision."[8]

The Air Force had already implemented both the spirit and the intent of DODD 5000.1 through a Chief of Staff letter on Operational Test and Evaluation dated 1 Apr 71 which called for "operational test and evaluation...conducted by operational forces to determine operational effectiveness and suitability..."[3] The reporting channel for the test results was independent of the developer and the Air Force felt that this complied with the intent of the DOD guidance and the Blue Ribbon Defense Panel recommendations.

Although the essence of the policy had been directed by the Deputy Secretary of Defense, Mr. Packard, in a memorandum dated 11 February 1971, it was not until 19 January 1973 that specific DOD policy on test and evaluation was formalized into DODD 5000.3, "Test and Evaluation". The Directive confirmed the existing policy that operational test and evaluation would be conducted by "operational and support personnel".[9] Initial Operational Test and Evaluation was to be accomplished during Full-Scale Development *to assist* in evaluating operational effectiveness. The most significant requirement was that "one major field agency (or a limited number of such major field agencies) separate and distinct from the developing/procuring command" be established. This agency was to report directly to the Chief of Staff and "Insure that the OT&E is effectively planned and conducted."[9] The operational test and evaluation was to be separate from development test and evaluation but initial operational test and evaluation could be combined with development test and evaluation if "separation would cause delay involving unacceptable military risk, or would cause an unacceptable increase in the acquisition cost".[9] In either eventuality, the operational test and evaluation was responsible for:

- insuring that the tests were planned and executed so as to provide the necessary operational test data;
- participating actively in the tests and,
- separately evaluating the test results.

There was no mention of the seeming conflict with DODD 5000.1 that stated "The development and production of a major defense system shall be managed by a single program manager..." and the split test and evaluation responsibility specified in DODD 5000.3.

The Air Force established the Air Force Test and Evaluation Center (AFTEC) as a separate operating agency on 1 January 1974. The Air Force was the last of the three services to initiate an independent operational test agency. With approximately 200 personnel, AFTEC was to assess the operational effectiveness, suitability and logistic supportability of new Air Force systems and to report the results directly to the Air Force Chief of Staff. The using command would continue to provide personnel to actually conduct the operational tests under AFTEC management. Thus the independent evaluating and reporting functions were established but did not result in a significant change in the number of persons assigned nor in the tasks performed by these persons. The only visible change at the test sites was that an AFTEC individual was assigned to the test force and given operational control over the using command personnel who were already in place. This action created considerable consternation and misunderstanding of the AFTEC role. Joint Test Force (JTF) Directors perceived the JTF "test teams" being fragmented and the Director's authority diluted. It was difficult if not impossible to draw a Joint Test Force organizational chart that really depicted the organization. A strong adversary relationship existed as a result of the new organization being imposed upon already existing and smoothly running test forces.

For nearly 4 1/2 years, DODD 5000.1 went unchanged, while DODD 5000.3 experienced only minor changes. The minor changes to the latter directive deleted the option of having more than one independent agency and further stipulated that the one remaining field agency would be independent of the using command and the developing command. The 22 Dec 1975 change to DODD 5000.1 left intact the concept of the single Program Manager being responsible for development of the system and the paragraph on test and evaluation was repeated verbatim. Changes were not made to the document that would indicate a shift in test and evaluation policy from the emphasis on early development test and evaluation and initial operational test and evaluation.[10]

CURRENT OT&E POLICY

DOD DIRECTIVE 5000.1

Major changes occurred approximately 1 year after the December 1975 change to DODD 5000.1 when the new DODD 5000.1 titled "Major Systems Acquisition" was issued (18 January 1977). Its purpose was to implement OMB Circular A-109 dated April 5, 1976 also titled "Major Systems Acquisition," and to update DOD policy on acquisition management. Considerable discussion on the role of the Program Manager is presented in DODD 5000.1. The Directive serves to further emphasize the Program Manager's overall responsibility for the development of the system. The test and evaluation paragraph had two changes.

Without defining the terms, the Directive added the requirement for "An estimate of military utility and of operational effectiveness..." to those of operational suitability and logistic support requirements.[11] Significantly, the new test and evaluation paragraph deleted the requirement that the operational test results be presented to the DSARC at the time of production decision. The former change appears to be a direct quote from the January 1973 DODD 5000.3 and is not significant. The latter was clarified in a subsequent interview with Director of Defense Research and Engineering (Test and Evaluation) officials. The change reflects the decentralization trend called for in OMB Circular A-109. The initial operational test and evaluation results are now to be reported to the Service SARC which is to be chaired by the Service Secretary or Undersecretary. The Secretary then will report the combined DT&E/OT&E results to the DSARC as appropriate.

DOD DIRECTIVE 5000.3

While the new DODD 5000.1 has little impact on test and evaluation, the subsequent revision of DODD 5000.3 contains major policy changes. Some of these changes appear to stem from the April 2, 1974 Defense Science Board "Report of Task Force on Test & Evaluation."[12] In testimony before the Research and Development Subcommittee of the House Armed Services Committee on 16 March 1977, General Lotz* stated "...we are making changes in our test and evaluation policies and directives to emphasize the need for conducting operational testing earlier in the development cycle of a weapons system." General Lotz went on to state four reasons, paraphased below, for the need for earlier operational test and evaluation.

a. "In considering system concepts during the early stages of a program, we need to know not only whether they are technically feasible but also whether they are operationally viable or tactically sound.

b. To have any impact on the research and development phase of a program, OT&E should be conducted much earlier than at DSARC III since by then most R&D resources will have been expended.

c. The earlier in the development cycle a deficiency is identified and corrected, the less the cost of corrective action.

d. A trend toward initiating limited production before the major production decision results in a commitment of procurement funds prior to the major production decision."[13]

The author was privileged to review the 12 July 1977 draft revision to DOD Directive 5000.3, that shows the current thrust of OSD thought, although the revision may be further modified during the review and coordination process. An analysis of the draft document revealed the following significant changes envisioned relative to planning, timing and reporting of operational test and evaluation. In general, the directive would require that the "... critical issues, test criteria and measures of effectiveness related to the satisfaction of mission need shall be established prior to the commencement of tests."[14] Further, the directive requires the establishment of a management reserve both in schedule and funds to cover contingency testing if significant test objectives have not been met.[14] The term "management reserve" is not used but the intent is clear.

Specifically, operational test and evaluation will be directly related to the modified milestone concept and the requirements to assess the system's vulnerability and its capability against enemy countermeasures will be added. The 1973 policy did not call for operational test and evaluation prior to the full-scale development phase. The 1977 policy requires that initial operational assessment begin in the conceptual phase following Milestone 0, to assess the operational impact of the proposed technical

*Lt General Walter E. Lotz, Jr., USA (Ret), then Deputy Director (T&E), Office of the Director of Defense Research and Engineering (ODDR&E).

approaches. During the demonstration and validation phase, initial operational test and evaluation would "...provide information relative to projected operational effectiveness and suitability of the candidate system."[14] Significantly, initial operational test and evaluation results *may* (as opposed to *will*) be considered at Milestone II as may any commitment of funds for items having a long lead time or limited production items. The previously specified full-scale engineering development phase involvement remains the same except that the initial operational test and evaluation report to support the production decision will be made to the Service SARC instead of the DSARC.

To determine what portion of development test and evaluation will contribute to the accomplishment of operational test and evaluation objectives, the Air Force Test and Evaluation Center is now directed to participate in development test and evaluation planning. Early coordination should serve to reduce duplication of effort, minimize required test resources and provide maximum data to satisfy common needs of the developer (Air Force Systems Command) and AFTEC.

A subtle change in words relative to combined development test and evaluation and initial operational test and evaluation has put a more realistic emphasis on the issue in these times of tight money. The 1973 directive stated that "...operational testing should be separate from development testing. However,...(they)... may be combined where separation would cause delay,...or would...increase the acquisition cost of the system."[9] The new directive modifies the statement to say "Development testing and operational testing may be combined...."[14] The remainder of the words are essentially the same but the emphasis is shifted to combined testing as combined testing is stated first. Interviews with DDR&E officials confirmed the intent of the change. However, a sentence added at Navy insistence clouds the issue slightly: "As a normal practice the operational tests supporting a production decision will be conducted independently by the OT&E agency."[14] Despite this add-on, the availability of resources, time and cost considerations will remain the final determinents. But the emphasis is definitely shifting toward the combined testing concept being regarded as the most efficient and cost effective method of test and evaluation. The key is to ensure:

- that the combined tests are planned and executed to realistically provide the necessary test information;
- that AFTEC participates actively; and,
- that AFTEC conducts a separate evaluation of the test data.

The 1974 Defense Science Board report stated that IOT&E "tests should not be conducted until the primary objectives of the DT&E have been met."[15] In an apparent partial response, the new DODD 5000.3 calls for AFTEC to "Monitor and review the results of DT&E...to assess the readiness of the system for operational testing."[14] Close, early involvement of AFTEC in the acquisition process should reduce the inevitable debate between the engineer who is never finished and the operational tester who is anxious to start.

There is an increased emphasis placed on the timely preparation of the Test and Evaluation Master Plan (TEMP). Up to now it had to be prepared "prior to initiation of Full-Scale Development" and was to "integrate the effort and schedules." Now the TEMP must be prepared as early as possible in the acquisition process and "...should identify and integrate objectives, responsibilities, resources, and schedules for all test and evaluation to be accomplished..."[14] This requirement alone should result in earlier involvement of AFTEC because without AFTEC participation the TEMP would be incomplete. The same emphasis is put on the test portion of the DCP (now called Decision Coordinating Paper). The Air Force Test and Evaluation Center inputs will be required to prepare a complete Decision Coordinating Paper.

The reader is reminded that the foregoing assessment of the draft DOD Directive 5000.3 may have to be modified if major alterations in the draft occur before final promulgation.

AIR FORCE REGULATIONS

Air Force policy on test and evaluation has not had time to react to the new DOD policies. The Air Force test and evaluation policy regulation, AFR 80–14, and the AFTEC regulation, AFR 23–36, are both mid-1976 documents. The most current element is the Air Force Systems Command Supplement to AFR 80–14, dated 3 January 1977.

Two key elements of the current AFR 80–14 are the relationship between the Program Manager and AFTEC and the conduct of combined testing. "The Program Manager has overall responsibility for a system acquisition program (except the management of OT&E.)"[16] The Program Manager will include operational test and evaluation requirements in the test program and support operational test and evaluation as appropriate. "AFTEC has responsibility for managing the OT&E in a major acquisition program...AFTEC will plan, direct, conduct, control, and independently evaluate and report on OT&E."[16]

Paragraph 17 of AFR 80–14, "Conducting a Combined Test Program" seems to negate any sort of team concept in combined testing. The planning aspects of the paragraph conflict with the new DODD 5000.3 in that these planning aspects call for separate development test and evaluation and operational test and evaluation test plans with the Program Manager responsible for integrating the operational test and evaluation plan. Joint planning is not mentioned. The separation of roles and functions continues down to the Joint Test Force. The test force director, appointed by Air Force Systems Command, is responsible for development test and evaluation, integrating combined test events, insuring availability of resources and insuring the safety of the test program. The operational test and evaluation test director, provided by AFTEC, manages the operational test and evaluation portion of the combined program. Previously this function was performed by the user detachment commander serving as Deputy Joint Test Force Director as opposed to "OT&E test director." Thus there is one Joint Test Force with two directors—a situation that tends to splinter the "test team" and is difficult to depict on an organizational chart. Air Force Systems Command Supplement 1 to AFR 80–14 attempts to improve the situation by stating that "This test force must be able to function efficiently as an ***entity*** to accomplish overall program test objectives; ***management relations must be clearly defined***" (underlining added for emphasis).[17] Unfortunately neither AFR 80–14 nor AFSC Supplement 1 specifies one individual who has complete test program responsibility. Air Force Regulation, AFR 23–36[18] generally follows the guidance set forth in AFR 80–14. Appropriate Air Staff agencies are aware of the difficulties with the combined test program guidance and hopefully will consider again this combined guidance in the next revision of AFR 80–14.

The Air Force Test and Evaluation Center has been in existence for more than 3 years. Initially there were serious growing pains as this new organization was imposed upon ongoing programs. An adversary had been thrown into a predominantly advocate environment. Very few individuals really understood the role of the independent test agency. The situation has now had time to stabilize and AFTEC is fully operational. In fact, AFTEC recently was authorized sixty additional personnel. The AFTEC FY 1978 budget calls for $17.9 million in 3080 (Other Procurement), 3400 Operation and Maintenance (O&M) and 3600 Research Development Test and Evaluation (RDT&E) money. This is more than double the $8.46 million FY 1976 budget and the size of the budget is expected to continue to rise. Figures were not available on the cost of AFTEC managed operational test and evaluation but it can be assumed to be significant since approximately 15 percent of the annual DOD Research Development Test and Evaluation cost for weapons systems is used for test and evaluation.[13]

COMPARATIVE VIEWS ON OT&E

With the relative significance of AFTEC and operational test and evaluation established, a series of structured interviews was conducted to determine the current view of operational test and evaluation in the Air Force. Through these interviews, in which eight questions were asked, the views of various Department of Defense and Air Force officials were compared with those of the AFTEC Commander. Officials from both development test and evaluation and operational test and evaluation organizations were interviewed as well as senior DD(T&E) officials.

The approach used to analyze the interviews was to determine if a consensus of opinion existed on the issues. Next the results were compared with the AFTEC view. Note that the interviewees were expressing personal opinions and not necessarily Air Force nor Department of Defense policy.

Question 1.

> DODD 5000.1 calls for a "strong SPO to achieve program objectives." AFR 80–14 says that the Program Manager has "overall responsibility" except for Operational Test and Evaluation. Does this in effect dilute the Program Manager's authority?

With minor reservations, the consensus is that significant problems with this concept do not exist at this time. The view is that since the Program Manager has control of the resources of the program, the dollars and the test articles, he is in a strong negotiating position when the original test plans are being developed. The negotiating task is a delicate one that requires diplomacy and tact in establishing effective working relationships with AFTEC. The same reasoning also applies at the test force level where the split roles are evident. The Joint Test Force Director has the necessary authority through his "ownership" of the test articles and responsibility for safety of operations. The concept of independence is important but not at overall program expense. In the final analysis the golden rule applies: "He who has the gold, rules." Therein lies the authority that the Program Manager needs and that can be used as appropriate.

The AFTEC Commander, Major General Leaf, views this issue in a similar fashion. Rather than viewing the Program Manager's authority as being diluted, General Leaf sees the situation as merely an affirmation of the differing roles of *development* test and evaluation and *initial operational* test and evaluation, the former being specification oriented and the latter being oriented toward the operational environment. Furthermore, because of the complementary nature of development test and evaluation and initial operational test and evaluation, the Program Manager's ability to produce an operationally effective and suitable system is enhanced.

As AFTEC has matured and the rest of the Air Force has gained a better understanding of its role, significant conflicts in this area have diminished. However, the issue should be put into clearer context in a subsequent change to DODD 5000.1, the basic acquisition management directive, and AFR 80–14. Further definition would preclude problems from reoccurring when the personalities involved are rotated to other assignments.

Question 2.

> How much should initial operational test and evaluation requirements influence total Full-Scale Engineering Development test schedules and the Milestone III date?

There is no clearcut consensus on this question. Since all programs are different, the effect will be different. The common thread throughout the responses was that well planned, achievement oriented programs (vs time-schedule oriented programs) would not have problems with the DSARC process. There is a danger of unnecessary program stretchout if the test program is not well structured since the requirement for an initial operational test and evaluation input to the production decision process is a firm one. Both the Program Manager and AFTEC must realize that initial operational test and evaluation is not an end in itself but rather it is one of the means to the end objective of selecting the correct production decision. ***When the areas of risk have been adequately addressed, go for production, don't continue to test to obsolescence.***

Again AFTEC concurred by emphasizing that adverse influence should not be experienced if the initial operational test and evaluation requirements are integrated into the program early and are properly reflected in the Statement of Work and the contract.

Question 3.

> Have we swung from too little to too much testing (DT&E or IOT&E) prior to Milestone III?

There is wide disagreement on this question. Interestingly, the developer oriented responses indicated that we are doing too much testing or at least leaning that way. There is also a feeling that, regardless of the amount of testing, the areas of emphasis need to be scrutinized. There is a concern that overemphasis on initial operational test and evaluation at the expense of development test and evaluation is dangerous and that developmental "show stoppers" could be missed. The best summation in this area is that the system should be developed first and the concept of combined developmental and operational testing (with independent evaluation) be used to the maximum extent practicable.

The AFTEC response quoted the 1977 Defense Science Board report:

> "There appears to be little or no overtesting done under the (DOD) directives; what testing is done contributes its full value to the improvement and verification of system performance."[19]

Question 4.

> What are we getting now in terms of operational assessment that was not available prior to 1974 when Air Force Systems Command and the users were working together? In other words, what new contribution is AFTEC making?

Question 5.

> The AFTEC budget projection for FY 1978 is $17.9 million and will be higher in the out years. Do you feel that we are getting our money's worth?

The only apparent consensus on these two questions is that the AFTEC main contribution is its independent test report to the Chief of Staff and to the AFSARC.* Otherwise, the responses are mixed. Some feel that AFTEC contributions are very worthwhile but that the cost effectiveness has yet to be proved. Others feel that the Air Force was doing just as well prior to AFTEC involvement. Whether or not the AFTEC contribution is cost effective will not be known until someone can determine a way of measuring costs avoided because of the AFTEC input. Any deficiency found early enough to correct during production instead of through retrofit equates to costs avoided. In the future, it may be possible to develop a means to compare the number of postproduction changes before AFTEC inception (1974) and after AFTEC. Only then can some valid, statistical measure of cost effectiveness be made. Until then, the AFTEC contribution is recognized as significant in the Air Force and at Department of Defense and Congressional levels.

The AFTEC response reinforced the positive aspects of operational test and evaluation in the acquisition process by citing several specific contributions. In addition to the reporting aspect, the contributions include:

a) revision of the deficiency reporting process to provide better visibility, "prioritization" of deficiencies, and identification of potential fixes through coordination with the developer and the user:

b) standardization of operational test and evaluation planning, criteria, execution and support as well as evaluation methodology;

---

*Air Force SARC.

c) improved software evaluation through the use of a software evaluation group;

d) earlier involvement of operational test and evaluation personnel at the program office and at contractor facilities; (Note that user personnel were collocated at both program office and contractor facilities on the B-1 program prior to the inception of AFTEC in 1974.)

e) improved logistics and manpower assessments and operations and support cost estimates.

From a cost effectiveness (cost avoidance) standpoint, the early identification of problems, when problems are relatively cheap to fix, is most significant. The example of the F-15 production radar is a case in point. Major General Leaf pointed out that the APG-63 production radar would not have been fixed without AFTEC having surfaced the problem at high levels in the Air Force.

Question 6.

AFTEC has been criticized as a 'nay sayer'. Is this a valid criticism?

Initially, the AFTEC reports seemed excessively negative and without balance. The interviewees generally agreed that this had been true but that it is no longer a significant factor. The feeling is that AFTEC personnel have matured in the job and have a better understanding of the AFTEC role as does the rest of the Air Force. The consensus is that AFTEC should be objective—neither success nor failure oriented—and should present a balanced report. There remains the danger of the initial operational test and evaluation comments being taken out of context outside of the Air Force. This latter danger makes the need for balanced objectivity even more important.

The AFTEC response agrees with the above. The AFTEC charter calls for evaluation without bias or prejudice. The objective is to provide an operational assessment that highlights both the positive and negative aspects of the system being evaluated.

Question 7.

Does it really matter who gathers the flight test data (developer or operator) if the test conditions are operationally representative?

This has been an emotional issue. In some programs, such as the B-1 and F-16, the "negotiations" reached the general officer level to determine the required mix. In an unlimited resource environment there would not be an issue since initial operational test and evaluation and development test and evaluation could be clearly separated. In combined programs with limited resources, the mix is significant. The responses to this question varied from, "you can't have an operational test if it is all done by developers" to the feeling that "a data point was a data point." Overall, there was agreement that the operational pilot has a different perspective than that of the trained test pilot who has been out of the operational environment for a number of years. Qualitative assessment by both developer and user pilots are required for a balanced evaluation. Another aspect was colorfully tagged the "hamburger" factor. Pilot skills vary. Test pilots and the best operational pilots evaluating a new aircraft, might subconsciously compensate for an aircraft characteristic that a pilot of lesser skill could not find acceptable. Thus there is a need for a mix of pilots. The Program Manager and AFTEC must jointly agree on the proper mix.

From the standpoint of pure performance and flying qualities data taking, AFTEC does not see a requirement to specify who flies the test missions since the missions are quantitative in nature. This is not true for the operational effectiveness evaluation. For initial operational test and evaluation, the user input is mandatory. Thus AFTEC concurs with the overall consensus of opinion.

Question 8.

AFTEC currently is limited to "managing" the IOT&E program. Should it also have the capability to "conduct" IOT&E as does its Navy counterpart?

The unanimous response to this final question was "NO". Such a concept would inevitably lead to another "super organization" or "operational monster" that the Air Force cannot afford. Such concept would be redundant and would absorb MAJCOM efforts that rightly belong to the user for follow-on operational test and evaluation. Further studies may offer a better system than currently exists but the current concept is considered much better for the Air Force than is the Navy approach of a separate operational test and evaluation force.

Not surprisingly, the AFTEC response was in complete agreement with the unanimous response noted above. It was pointed out that the current concept is in concert with the planned new DODD 5000.3. Since the AFTEC test team personnel are "borrowed" from the using commands the team is current in operational matters—more so than if the team members were assigned to an independent agency. There is, as a result, less danger of AFTEC personnel becoming "professional testers" and losing sight of the AFTEC role. Furthermore, the sharing of test resources fosters combined DT&E/IOT&E testing which is inherently more cost effective if properly done.

INCORPORATING OPERATIONAL TEST AND EVALUATION INTO THE PROGRAM

Regardless of the variety of opinion on the utility and effectiveness of AFTEC in the acquisition process, the fact remains that AFTEC is here. The AFTEC contribution is highly visible and time sensitive in the decision process. However, in the final analysis, it is the Program Manager who is held responsible for the program outcome. How effectively the *valid* operational requirements are incorporated into the program will be a major determinent of the program outcome. What can the Program Manager do to ensure that all valid operational requirements are adequately covered in the program?

The answer to this extremely important question can be broken into three major areas:

- Establishing the baseline
- Organizing
- Planning

Establishing the Baseline

Establishing the baseline involves a number of elements. First of all, know the significant DOD Directives and AF Regulations—DODD 5000.1 for overall guidance; DODD 5000.2 for the Milestone checklists and Decision · Coordinating Paper content; DODD 5000.3 for the specific DOD test and evaluation policy; and, AFR 80–14 and AFSC Supp 1 for Air Force test and evaluation policy. Also, "A Guide to Program Management", AFSCP 800–3, is an excellent reference.

Second, know the operational concept for the system. The operational concept is the key baseline. It has to be included in the system design as well as the test plan. This concept will serve as the criterion against which the system effectiveness is measured. A new Air Force policy for initiating, developing, updating and approving operational concepts for new and improved major systems is being developed and the Program Manager has to be involved if the system under development is to meet the need. Figure 1 depicts the flow of the operational concept.[20] The Required Operational Capability (ROC)* has always been part of the system documentation. The ROC leads to the *Mission Element Need Statement (MENS)* which is a new requirement under DODD 5000.1. The Program Manager needs to be intimately familiar with both the ROC and the MENS. The proposed operational concept will be a dynamic document tied to program milestones for update and refinement. Air Force Systems Command supports this concept and the requirement for approval of operational concept changes to preclude inadvertant breaches of program thresholds. The concept format contains operational thresholds, forces, standards for development, organization, basing and support, and significantly, the updated baseline for

*Recently renamed General Operational Requirement (GOR).

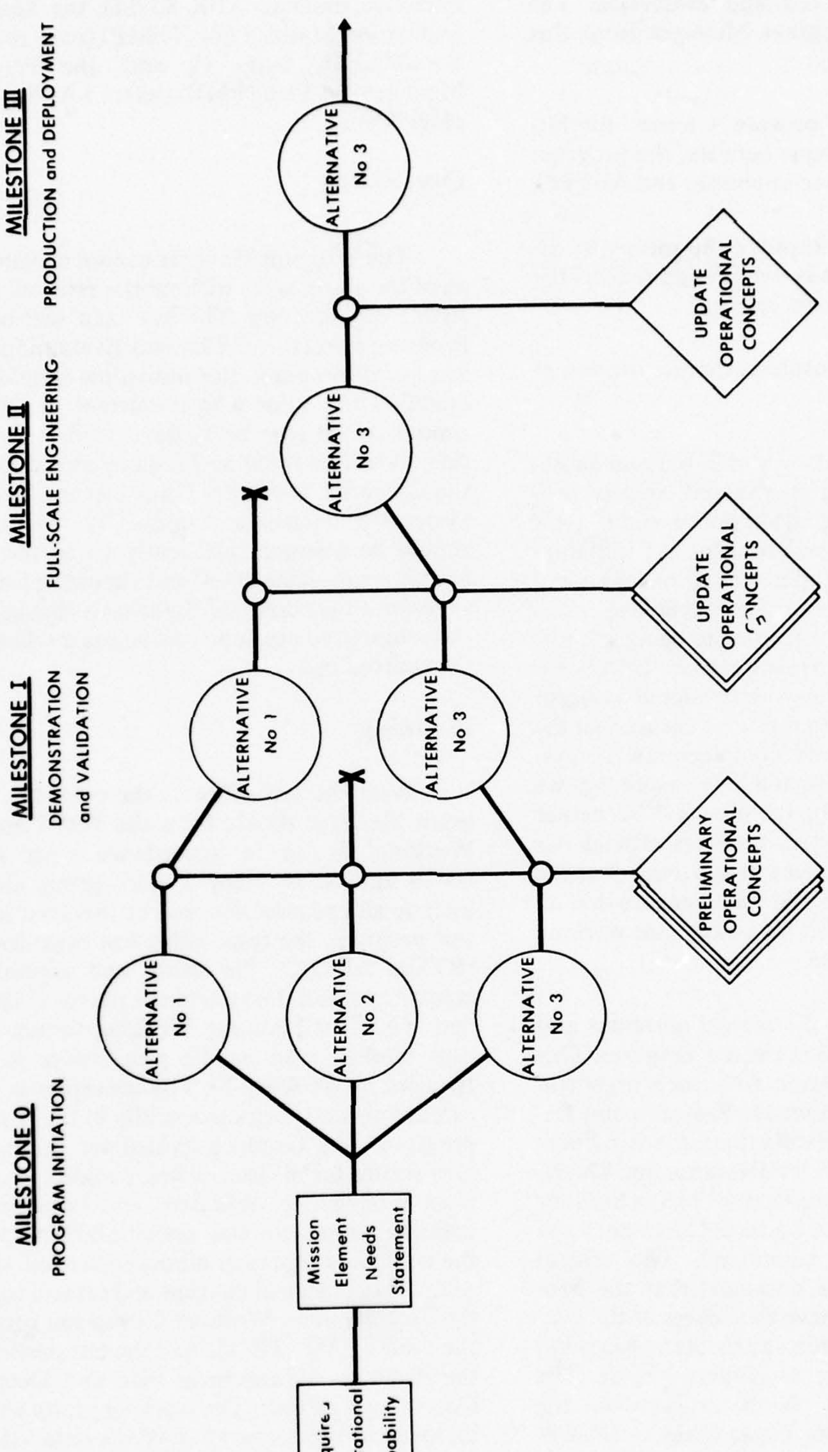


Figure 1. Proposed Operational Concept Role

system design and test and evaluation. The benefits for the Program Manager from this concept are:

- that it will provide a forum for improved dialogue between the program office, the user, supporter and AFTEC;
- that it will improve the means to ensure that the system design reflects the operational concept; and,
- that it will ensure corporate reviews of changes.

"The corporate inputs to the initial and subsequent updates of the operational concept provide the user, developer, supporter and tester a voice in the operational concept. At initiation and each update, the operational concept document is required to go to all applicable agencies for review and comment prior to being submitted by the user to Headquarters USAF for approval. The proposed operational concept directive will provide a means to ensure that the operational community complements the system acquisition process and help ensure that we get the best system for the money."[20] Whether or not the above system becomes official Air Force policy, it is the type of approach that the Program Manager should use to ensure that all parties concerned with his system are working from the same baseline.

Third, establish the critical questions and issues to be addressed in the test program. This will require coordination with such organizations as Deputy Director for Research and Engineering (Test and Evaluation), the Air Force Deputy Chief of Staff for Research and Development, the Air Force Deputy Chief of Staff for Operations, Air Force Systems Command, and the AFTEC/Using Command. The critical questions will be the questions that the Program Manager will have to address at the program Milestones to gain approval to proceed to the next phase of the development cycle. The questions and issues will be reflected in: the Decision Coordinating Paper (refer to DODD 5000.2); the Test and Evaluation Objectives Annex (TEOA) of the Program Management Directive (refer to AFR 80–14); the Test and Evaluation Master Plan (TEMP) (refer to AFR 80–14/AFSC Supp 1); and, the Program Management Plan (PMP) (refer to AFSC Pamphlet 800–3).

Organizing

The Program Manager cannot accomplish all of the above tasks without the assistance of a strong organization. The key man will be the Program Director of Test and Evaluation. For an aircraft program, this individual should be a qualified test pilot who is current in aircraft similar to the type being developed. The position should be rated and require attendance at the Executive Refresher Course at the Defense Systems Management College. The test office should be manned sufficiently to be the program's single voice of test and should be responsible for all ground and flight tests throughout the program to preclude conflicting guidance to the contractor.

Planning

With the test office as the core, the Program Manager should form the Test Planning Working Group in accordance with AFR 80–14 and AFSC Supp 1. This group should include all agencies that will be involved in the test program; the responsible test organization (RTO); AFTEC; the using and supporting commands; and the system contractor if identified. The Test Planning Working Group concept worked well in the past before it was required by AFR 80–14. The concept was used extensively and very successfully in the B-1 test program. The Group provided for the necessary forum for all test related subjects ranging from establishing objectives and baselines to defining organizational responsibilities. With the combined inputs of all organizations along with the operational concept and critical issues, the Test Planning Working Group can prepare the draft TEMP, TEOA, and the test portion of the Program Management Plan and Decision Coordinating Paper. The working group will be involved in the preparation of the detailed test plans whether or not the contractor is assigned primary planning responsibility.

The Program Manager has to be directly involved in this planning process. Based on the program charter, he must provide specific guidance to the Program Test Director. While the Program Manager is responsible for integrating the initial operational test and evaluation requirements into the program, he should not blindly accept them. All requirements and changes to requirements must be questioned to verify validity and need. The requirements should be listed in order of priority as specific objectives rather than as percentages of flying hours or missions. The Program Manager must ensure that the aircraft is adequately developed through the accomplishment of orderly development test and evaluation objectives before jointly agreeing with AFTEC that the system is ready for initial operational test and evaluation. Finally, the Program Manager must ensure that the output of the planning process is a jointly developed strategy that is agreed to by all parties concerned. Only a coordinated, program oriented plan can ensure that there will not be surprises later in the program because of a previously unidentified requirement that is either out of scope or is one that severely impacts the schedule. Finally, the Program Manager will have complied with both the spirit and the intent of DODD 5000.3 through the early incorporation of the initial operational test and evaluation requirements.

SUMMARY

The criticisms of, and changes in, Department of Defense and Air Force policy on operational test and evaluation from 1970 through early 1977 have been reported in this article. The new DODD 5000.1 and the draft DODD 5000.3 have been analyzed. The views of various Department of Defense and Air Force elements on the effectiveness of AFTEC and the AFTEC view of itself were compared. Finally, the policies and views were combined into recommendations for the Program Manager to follow to ensure that the valid operational requirements are in the development program. Each program is different and a universal set of guidelines cannot be developed that will apply to all. But the Program Manager will consider all guidelines and tailor the tactics to his specific situation. In that way the Program Manager can be as assured of success as anyone can in the changing political and military environment of today's world.

REFERENCES

Cited References

1. Dept of Air Force, "Research and Development: Test and Evaluation of Systems, Subsystems, and Equipment," AFR 80–14, Washington, DC, 14 Aug 63, pp 4,5.

2. Presidential Blue Ribbon Defense Panel, "Report to the President and the Secretary of Defense on the Department of Defense—Staff Report on Operational Test and Evaluation," Appendix F, Washington, DC, Jul 70, pp 26,9.

3. Col Fred W. Ermel, Jr., USAF, "Operational Test and Evaluation Within the Department of Defense: A Reexamination Two Years After The Blue Ribbon Defense Panel Report," Washington, DC, 9 Mar 73, pp 85,89.

4. Presidential Blue Ribbon Defense Panel, "Report to the President and the Secretary of Defense on the Department of Defense," Washington, DC, 1 Jul 70, pp 2,88,90.

5. Melvin R. Laird, to Secretaries of the Military Departments and Director, Defense Research and Engineering, Memorandum, subject: "Conduct of Operational Test and Evaluation," 13 Jul 70, p 1.

6. David L. Packard, to Secretaries of Army, Navy and Air Force and DDR&E, Memorandum, subject: "Conduct of Operational Test and Evaluation," 11 Feb 71, p 1.

7. ________, to Secretaries of Military Departments and Others, Memorandum, subject: "Test and Evaluation in System Acquisition Process," 3 Aug 71.

8. Dept of Defense, "Acquisition of Major Defense Systems," Directive 5000.1, Washington, DC, 13 Jul 71, p 5.

9. ________, "Test and Evaluation," Directive 5000.3, Washington, DC, 19 Jan 73 (w/chg dtd 12 Apr 74), pp 3,4.

10. ________, "Acquisition of Major Defense Systems," Directive 5000.1, Washington, DC, 22 Dec 75.

11. ________, "Acquisition of Defense Material," Directive 5000.1, Washington, DC, 18 Jan 77, p 8.

12. Defense Science Board, "Report of Task Force on Test and Evaluation Policies," Eugene G. Fubini, Chairman, Washington, DC, 2 Apr 74.

13. Dept of Defense, "The DOD Program of Test and Evaluation, FY 1978," Overview Statement by DD(T&E), ODDR&E (95th Congress, 1st Session), Washington, DC, 16 Mar 77, pp 9,10.

14. ________, "Test and Evaluation," Directive 5000.3, Washington, DC, 12 Jul 77, (DRAFT) pp 2,3,6,8,9,7,13.

15. Defense Science Board, "T&E Guidelines for Aircraft Systems," Eugene G. Fubini, Chairman, Washington, DC, 2 Apr 74, p 4.

16. Dept of Air Force, "Test and Evaluation," AFR 80–14, Washington, DC, 19 Jul 76, p 3.

17. Air Force Systems Command, Headquarters, AFSC Supp 1 to AFR 80–14, "Research and Development, Test and Evaluation," Andrews AFB, MD, 3 Jan 77, p 3.

18. Dept of Air Force, "Air Force Test and Evaluation Center (AFTEC)," AFR 23–36, Washington, DC, 19 Jul 76.

19. Defense Science Board, "Report of Task Force on Test and Evaluation Policies," Eugene G. Fubini, Chairman, Washington, DC, Jan 77, (DRAFT) p 4.

20. Lt Col Robert S. Delligatti, USAF, "Operational Concept," Briefing to Air Staff Board, 17 Mar 77, p 12B.

Additional References

Boykin, James A., Consultant, ASD/AV, Wright-Patterson AFB, OH, Telephone Interview, 21 Mar 77.

Delligatti, Robert S., Lt Col, and Lt Col Barry J. Howard, Interviews at USAF/XOODE, Pentagon, Washington, DC, 17 Mar 77.

Hale, Alen B., Col, USAF, Interview, USAF/RDX, Pentagon, Washington, DC, 11 Feb 77.

Kollmorgen, Leland S., RADM, USN, Interview, AD(SSST&E) ODDR&E, Pentagon, Washington, DC, 11 Feb 77.

Leaf, Howard W., Major General, USAF, Commander, AFTEC, Kirtland AFB, NM, Written response to interview questions, 28 Mar 77.

Rushworth, Robert A., Major General, USAF, ASD/CV, Wright-Patterson AFB, OH, Telephone interview and written response to interview questions, 21 Mar 77.

Ward, Donald T., Col, USAF, AFSC/TEV, Andrews AFB, MD, Telephone interview, 11 Mar 77.

Lieutenant Colonel Michael J. Butchko, US Air Force is assigned to Headquarters, USAF, Office of the Deputy Chief of Staff for Research and Development (Strategic Requirements). His previous assignment was Deputy Director, Development Test and Evaluation, Deputy for the B-1, collocated with the B-1 Joint Test Force at Edwards AFB, CA.

Earlier assignments of Lieutenant Colonel Butchko include service as a KC and RC-135 pilot, Strategic Air Command (1964–1969); EB-66 pilot, Southeast Asia (1969–1970); student and subsequently an Instructor, USAF Test Pilot School (1970–1973).

Lieutenant Colonel Butchko holds a BS from the USAF Academy and an MS, Systems Management, from the University of Southern California. He is a graduate of Squadron Officer School, the Armed Forces Staff College, the USAF Test Pilot School and the Defense Systems Management College.

USE OF TRAINING EXERCISES FOR TEST AND EVALUATION

by

Lt Col Martell D. Fritz, USA
US Army Operational Test and Evaluation Agency

How many times have you heard a major automaker or other industrial giant announce the recall of an expensive product for correction of an operational or safety defect? This process is costly, time consuming and can severely impact the future growth of the corporation by virtue of consumer dissatisfaction. To reduce the number of recalls, industry devotes significant resources to improving its capability to predict product performance. This philosophy began to pervade the Department of Defense community during the David Packard era and became known as the "fly before buy" approach. In this article the author describes how operational testing combined with training serves to deter the fielding of items that lack operational effectiveness.

INTRODUCTION

In May 1970, the Department of Defense (DOD) formally recognized the need for independent operational test and evaluation (OT&E) to determine system operational effectiveness. The "fly before buy" philosophy was to be fully implemented by all services. Each DOD component was directed to establish a major field agency and give it responsibility for the Operational Test and Evaluation of materiel systems. The agency was directed to be separate and distinct from the development, procurement and user commands. Further the Army determined that the DOD directive could best be implemented by having the Army's independent agency report directly to the Army Chief of Staff. In September 1972, Department of the Army (DA) established the US Army Operational Test and Evaluation Agency (OTEA) and charged this Agency with the Operational Test and Evaluation mission for DA.[1] This Agency would provide the decisionmakers with valid test data and independent evaluations of system suitability in keeping with the threat, organizational and doctrinal needs, tactics, techniques and training. The concept was further keyed to conduct of testing with typical soldiers in a realistic operational environment.

The essence of *Operational Test and Evaluation* is to determine the usefulness of a system in combat. The objective is to measure and evaluate a system's operational effectiveness and suitability based upon realistic operational conditions. The distinction between testing and evaluation should be noted. *Operational testing (OT)* is the gathering of information. *Evaluation* involves the assessment of the information gathered to achieve an objective *evaluation*.

To accomplish its mission, OTEA normally conducts two formal sequential tests which are keyed to DOD decision reviews. Operational Test I (OT I) occurs early in the materiel acquisition process. Here concentration is on the primary system functions, man-machine interfaces at the operator level, and the demonstration of system feasibility. Operational Test II (OT II) occurs following full-scale development and normally is the most rigorous test in terms of obtaining data with statistical confidence, particularly in the areas of functional performance and reliability, availability and maintainability (RAM). The Operational Test II focuses on mission requirements with regular troop units in controlled field exercises and addresses logistics support to the maximum extent. The results of OT II may be used to support a production decision. If OT II reveals significant deficiencies, or if the decision review determines that critical issues remain unanswered, the decision authority may direct that corrections be made and that appropriate check tests be conducted during a contingency operational test, OT IIa. When directed, OTEA will conduct a follow-on evaluation subsequent to the production decision to provide information not gained from previous operational test and evaluation. Additionally, OTEA approves the plan and monitors the conduct of force development testing and experimentation (FDTE). The FDTE is used to:

- develop concepts of employment,
- determine operational feasibility,
- estimate the operational advantages of a proposed system, and
- assist in the development of requirements documentation.

The Operational Test and Evaluation Agency is the Army point of contact with the Office of the Secretary of Defense (OSD) on joint user test matters. The Agency has overall Army management responsibility for OSD-directed joint user test programs. Joint user tests are those in which the Army participates with one or more of the other Services to evaluate systems or concepts having an interface with or requiring a test environment of another Service.

REQUIREMENTS FOR OPERATIONAL TESTING

Test site selection is influenced by many factors. Perhaps the most important factors are test environment, availability of troops, and host installation logistics support. The test environment includes the weather, type and size of terrain, and the type and amount of vegetation. To insure validity of operational testing the test environment must closely approximate the actual combat conditions under which the system is designed to function. Equally important is the availability of player, aggressor and support troops. The troops must possess the required military occupational specialities to operate, maintain and support the system under test. Simulation of a realistic operational environment often necessitates the use of combat, combat support, and combat service support troops. If these troops are not available at the test installation, such troops must be brought in from other installations. A significant increase in cost is a result. A third major factor influencing test site selection is adequacy of logistics support. For example, if mechanized infantry or armor units are involved, a site where such units are assigned is preferred. Due consideration must be given to the adequacy of administration and supply support (billeting, messing, office space, communications, transportation, etc.) required for test directorate personnel.

After the location has been selected the Operational Test and Evaluation Agency normally will send a small task force (three or four persons) to the installation to discuss test requirements with the installation commander and his staff. The primary purpose of this visit is to integrate the test requirements with the training objectives of the units selected to support the test. The coordination visit allows the host installation to participate in the planning phase and is a key step in the detailed test planning process.

Since testing can be costly, the US Army Operational Test and Evaluation Agency is continually searching for ways to reduce costs without degrading the adequacy, quality and credibility of tests. Continual coordination with the Program or Project Manager, the developmental test community, and the host installation, on cost matters provides information necessary to address cost implications.

USE OF TRAINING EXERCISES

The combining of operational testing and training exercises can benefit the attainment of both operational test and training objectives. One way to attain mutual benefit is to satellite an operational test on a planned training exercise. Another valid approach is to define the operational test objectives and requirements and then plan a training exercise that accommodates those requirements. In either approach, the operational test objectives are met while the test units are achieving training objectives.

The Operational Test and Evaluation Agency has gained significant experience in the use of training exercises to accomplish required operational testing and evaluation. This experience has served to highlight numerous advantages and to identify the problem areas associated with the integration of operational test and evaluation with training exercises. The following examples illustrate how training exercises have been used for operational test and evaluation and the relative success achieved.

BART

The Baseline Armor Reliability Test (BART) was conducted to determine the reliability of the new M60A1, the overhauled M60A1, and the overhauled/converted M48A5 tanks. Five tanks of each type were driven 2250 miles. Each tank fired 450 rounds from its main gun. Data concerning repair parts usage, boresight zero retention of the main gun, and subjective information on overall tank performance were obtained. The test was conducted by the TRADOC Combined Arms Test Activity, with OTEA participation, at Fort Hood, Texas (26 April to 12 November 1976). The test support unit was the 2nd Battalion, 8th Armored Cavalry, 1st Cavalry Division. The test involved three phases with each phase consisting of five 10-day increments. Test events were scheduled and controlled by the test unit (consistent with the requirement to travel 150 miles and fire 30 main gun rounds). The battalion commander had virtually a free hand to plan and conduct each exercise. The battalion was able to conduct 6 months of excellent Army Training and Evaluation Program (ARTEP) training without cost to the unit. All expenses were paid from test funds. The three tank companies completed the standard tank gunnery qualification courses and conducted extensive platoon level training. The results of this test illustrate an outstanding example of the increased combat readiness that can be attained through the combination of testing and training exercises.

DRAGON OT IIIa

The DRAGON weapon system is a command linked, line-of-sight guided missile system. This system provides firepower for the soldier against tanks, armored vehicles and fortified battlefield emplacements. The DRAGON OT IIIa was conducted at Fort Bragg, North Carolina (18 September through 10 December 1975). The primary purpose of the test was to determine whether the approved logistical support concept was capable of supporting the DRAGON missile system in addition to the TOW and SHILLELAGH systems (under wartime conditions) and to assess the adequacy of reliability design changes to the DRAGON tracker. In Phase III of the test, three maneuver battalions of the 3d Brigade, 82d Airborne Division, conducted separate 10-day tactical field exercises. Each battalion developed an operations order that incorporated the tactical scenario for the test as well as the unique training objectives of the battalion. Each exercise included day and night live fire exercises and nonfiring offensive, defensive and retrograde operations against a mobile agressor force. Live fire support was provided by Army aviation, division artillery, and Air Force tactical aircraft. Two hundred DRAGON training missiles and nineteen TOW missiles were fired during Phase III. This exercise proved to be highly successful in terms of combining an operational test with a training exercise. The battalions received excellent training that included the benefits of the additional ammunition allocated for the test. Seventy-six DRAGON gunners were rotated through the live fire portion of the exercise, enhancing the antiarmor training readiness within the division. Since the initial planning of the battalion commanders incorporated all requirements for the operational test, the training exercises were conducted without interference. The only significant test limitation was the short duration of the exercises (10 days). This test period did not allow sufficient time for all supply actions (parts identification, requisitioning, processing, delivery and application) to occur.

SOTAS OT I

The Stand-Off Target Acquisition System (SOTAS) consists of an airborne, long-range, high resolution Moving Target Indicator (MTI) radar and a data processing and display ground station. An integrated tracker/data link connects the airborne platform to the ground station. The SOTAS is being developed to provide the commander with an improved capability to monitor the battlefield and provide a target acquisition capability for engagement of targets at long ranges.[2] The SOTAS underwent user testing in September 1976 during the execution of REFORGER 76 conducted in the Federal Republic of Germany. This exercise was a two-sided, force-on-force, free play tactical exercise conducted by units of the North Atlantic Treaty Organization. A UH-1H helicopter was used as the airborne platform for SOTAS. The Test Design Plan for the SOTAS REFORGER test was prepared by the US Army Combat Developments Experimentation Command (CDEC). In the planning phase, OTEA worked extensively with CDEC to develop data requirements for the user test—data requirements that would be comparable to the requirements for an Operational Test I. After the user test, OTEA determined that the results were adequate to waive the requirement for an Operational Test I of SOTAS. Although the test was successful from an OT I standpoint (feasibility proved and military potential demonstrated), some limitations were evident. The nature of the free play exercise made it difficult to control events and gather precise data such as the true location of targets identified by SOTAS. Additionally, since the primary objectives of REFORGER were related to training and readiness with the SOTAS test being secondary, the troops did not recognize the role of test team data collectors and did not allow data collectors access to the operations center. Thus, the flow of information from SOTAS to the Division G2 could not be recorded with accuracy.

MCTNS DT II/OCT

The man-portable, common thermal night sights (MCTNS) are a family of night sights developed to provide a lightweight long-range nighttime capability for the TOW and DRAGON missiles. The MCTNS include the Night Observation Device-Long Range (NOD-LR) and the Ground/Vehicular Laser Locator Designator Thermal Night Sight. The MCTNS are designed to be used for battlefield surveillance, target acquisition and missile firing. The Developmental Test II/Operational Climatic Test of MCTNS in a cold region, winter environment was conducted near Fort Greely, Alaska (January through April 1977). A portion of the test was conducted in conjunction with a battalion level Army Training and Evaluation Program (ARTEP). The ARTEP is a diagnostic tool used to evaluate performance and program training to achieve a specified level of proficiency.[3] Normally a formal battalion Army Training and Evaluation Program is conducted once every 12 to 18 months and often serves as a basis for evaluating the performance of the battalion commander. During the ARTEP for the 4th battalion, 9th Infantry, two TOW missile squads with night sights and one ground surveillance section with two night observation devices-long range, were attached to the battalion. Numerous benefits accrued from combining the MCTNS test with the Army Training and Evaluation Program. The benefits included:

- a reduction in resources required,
- a reduction in cost,
- a cost saving, realistic RAM environment,
- a realistic logistics burden for the test and,
- use of real world tactics and doctrine.

These advantages provide strong rationale for extensive use of ARTEP for accomplishing testing requirements; however, there are some limitations that must be considered. One of these limitations is that the nature of the Army Training and Evaluation Program makes it difficult to replicate events. Although replication is not impossible, such as in a force-on-force free play exercise, any test that requires considerable replication of events (for statistical confidence) would tend to degrade the realism of the ARTEP. A more serious limitation is the added burden that is placed on the commander. Since the ARTEP is instrumental in evaluating his performance, a commander is not usually receptive to additional tasks that may complicate mission performance during the ARTEP. The commander must be assured that the data collected in the test will not be used for any portion of the ARTEP evaluation. Although the MCTNS Operational Climatic Test did not interfere with the ARTEP in this instance, many tests could have an impact on the training, employment and logistics missions of the battalion. Thus, it is apparent that each test has its own unique requirements and not all systems should be tested in conjunction with an Army Training and Evaluation Program.

CORPS AUTOMATION REQUIREMENTS (CAR)—FORCE DEVELOPMENT TESTING AND EXPERIMENTATION (FDTE)

Project Corps Automation Requirements is an evolutionary program designed to identify, validate and test corps management information system automated data processing requirements, less Army Tactical Data Systems, for the midrange (3 to 8 year) time frame.[4] In the period September 1976 to March 1977, the XVIII Airborne Corps conducted the Project CAR Baseline Identification Test to define the current Corps automated data processing baseline for combat service support. Subtest IV of this force development and experimentation was conducted in conjunction with an XVIII Airborne Corps Command Post Exercise, CABER WARRIOR IV, at Fort Bragg, North Carolina (1 through 7 March 1977). In this subtest, the 14th Data Processing Unit, 1st Corps Support Command, operated the mobile S/360–40 computer configuration in a field environment. The purpose of this subtest was to examine the test unit's Modified Table of Organization and Equipment and to observe operating procedures in a tactical environment. Additionally, the subtest was to verify certain tactical information requirements identified in earlier subtests. This test is an excellent example of employing Operational Test and Evaluation as a satellite on a previously scheduled training exercise. Both the exercise and the force development testing and experimentation were successfully completed without degradation to either.

CEWI BATTALION-FDTE

The Combat Electronic Warfare Intelligence (CEWI) Battalion (Division) is a conceptual organization that consolidates all tactical electronic warfare and intelligence resources (personnel and equipment) to be assigned to a US Army division. The battalion organization is being developed in response to a recommendation made in a Department of Army Intelligence Organization and Stationing Study, a study in which tactical intelligence integration was addressed. The study group recommended the placing of tactical support resources in single units under the command and control of tactical Army commanders. Because of this recommendation, the CEWI Battalion was evaluated by the TRADOC Combined Arms Test Activity (TCATA) during GALLANT CREW 77 (a Joint Chiefs of Staff directed US Readiness Command joint readiness exercise). GALLANT CREW 77 was conducted at Fort Hood and other locations in central Texas (25 to 31 March 1977). The purpose of this force development test and experimentation was to evaluate the concept of organization and operation for the CEWI Battalion of the 2nd Armored Division. Although the objectives of the exercise were met in an outstanding manner except for the conduct of operations in an electronic warfare environment,[5] the force development testing and experimentation were substantially less successful. There were three major reasons for the limited success of the force development testing and experimentation. Probably most important was the lack of early involvement by the Operational Test and Evaluation Agency in the planning phase. The late entry of OTEA limited the ability of the test team to interject situations into the exercise that would stress functional areas of the CEWI Battalion. Because of the free play nature of the exercise, the US Readiness Command was reluctant to interject any artificiality into the scenario for fear of jeopardizing attainment of the exercise objectives. Secondly, severe personnel and equipment problems affected the sensitivity of the test data. There was a shortage of personnel having certain skilled MOS and personnel with required security clearances. The equipment shortages occurred because of low item density in special types of electronic warfare equipment. Thirdly, the level of training of the Battalion was less than desired. The training shortfall resulted from limited training time and late key personnel and equipment fills.

FUTURE POSSIBILITIES

The success achieved to date provides a strong incentive to examine future testing for possible integration with training exercises. There are numerous tests that may fall into this category. Some of the more promising possibilities are outlined here.

• ***Division Restructuring Study FDTE.*** The purpose of this force development testing and experimentation is to evaluate the overall effectiveness of a restructured heavy division. Phase II of the test currently is scheduled to be conducted during 1978. Envisioned is the testing of a division (minus) in an operational environment. The test will involve elements of III Corps, 1st Cavalry Division, and the 2nd Armored Division. This phase of the test offers excellent training opportunities and is a candidate for integration with a large exercise.

• ***M198 Follow-on Evaluation.*** The M198 155mm towed medium howitzer was type classified and approved for production on 14 October 1976. The purpose of the follow-on evaluation, currently scheduled for Oct 78-Jan 79, is to verify that system deficiencies have been corrected in the areas of reliability, availability and maintainability, tactical mobility, training, and human factors and in the concept of logistic support. The test appears ideal for integration with pre-ARTEP training of the initial operational capability unit.

• ***IFV/CFV OT II.*** The Infantry Fighting Vehicle/Cavalry Fighting Vehicle (formerly known as the MICV) OT II is scheduled for 1979. The Infantry Fighting Vehicle is the companion vehicle for the XM1 Tank and must be tested in a combined arms operational environment. Although the test does not appear to be appropriate for use as a satellite on a training exercise, it should be extensive. This test has great potential for thorough combined arms training of the test units with a resultant improvement in combat readiness.

• ***EW/CAS Joint Test.*** The Electronic Warfare During Close Air Support (EW/CAS) Joint Test is a US Army/US Air Force/US Marine Corps operational test to be conducted during the period 1978 through 1980. The purpose is to determine the impact of an intense electronic warfare environment on command, control and communications systems. Emphasis will be on close air support and attack helicopter support during combined operations. The scope and nature of this test offer an opportunity to combine testing with division, brigade, or battalion training, or joint exercises.

BENEFITS

The Army's efforts to combine testing with training exercises have yielded tangible benefits. The results, although some are difficult to quantify, definitely contribute to the overall goal of improved combat readiness. Certain significant benefits are summarized below.

• ***Improved command readiness.*** Because of the extra training received in conjunction with an operational test many test units have been able to achieve a much higher level of combat readiness than had been previously obtained. Commanders at all levels are concerned about the cost of training and the attainment of training objectives within budget limitations. Budget constraints usually result in the establishment of definite limits on the quantity of fuel, ammunition and spare parts available to support training objectives. The BART and DRAGON OT IIIa are two excellent examples where commanders improved unit combat readiness by innovative and resourceful planning for tests.

• ***Realistic test environment.*** Training exercises that are planned and executed by combat units usually provide a more realistic environment for a test than do exercises that are planned by "expert testers" using "school solution" doctrine and tactics. Real world problems and limitations are more likely to appear in training exercises than in "canned" scenarios.

• ***Cost savings.*** Cost is an important consideration in every test. Cost often becomes a critical factor in size and scope of test determinations. As systems become more complex and testing becomes more expensive, every possible cost reduction must be explored. Accordingly, placing satellite tests on previously scheduled training exercises may become more of the norm in the future.

• ***Realistic logistics burden (including reliability, availability and maintainability).*** A difficult task in the design of a test is the creation of an environment that will provide a reliable indication of the logistics supportability of a system and the adequacy of reliability, availability and maintainability characteristics. Usually the operational environment is simulated by employing a small slice of the direct support and general support capability and dedicating this small contingent to the support of the system being tested. Thus the support units are rarely placed in a "true" operational environment. One obvious advantage of testing in conjunction with large training exercises is that support units must function within an overall tactical context and provide support for the various types of units and equipment used in the exercise. Any additional logistic burden imposed by the system being tested will be readily identified.

• ***Reduced test support requirements.*** Support of user testing involves extensive planning to provide for administrative and logistic support of test directorate and player personnel, use of aggressor forces, control of personnel, training, availability of equipment and numerous minor tasks. Many support requirements may be significantly reduced when testing is integrated with a training exercise.

• ***Development of doctrine and tactics.*** Training exercises better accommodate the evaluation and refinement of doctrine and tactics for the system being tested. In the context of a training exercise real world problems can be surfaced and possibly resolved prior to the fielding of a system.

PROBLEM AREAS

The Army's experience with operational test and evaluation has revealed problem areas that, in some cases, limit the value of combining testing with training exercises. These limitations underscore the need for caution before a decision is made to combine a test with a training exercise. Significant problem areas include:

- ***Inability to control events.*** It is impracticable to control events in a free play exercise. Although this constraint may not affect some tests, in other tests it is essential that certain system capabilities be tested or stressed. When considering a free play exercise, the system must be reviewed to insure that critical capabilities will be completely tested. If this assurance cannot be obtained, the exercise will be of little value in meeting the test objectives. This fact was amply demonstrated during the CEWI Battalion force development test and experimentation.

- ***Possible burden on the Commander during the ARTEP.*** A formal Army Training and Evaluation Program is a key period of evaluation for a commander. Any test that places an additional burden on the unit taking an ARTEP probably would not be well received by the commander. Not only would the combined test be unfair to the commander but also the system being tested probably would not be employed to maximum capability. This situation does not preclude the possibility of using an Army Training and

Evaluation Program as a test vehicle, but it does severely limit the instances where use of such vehicle would be feasible. One type of system that could be tested is one that provides a complementary increase in capability as opposed to a complete replacement system requiring extensive reorganization and/or training.

- ***Possible interference with control and data collection.*** Every test has a requirement for a certain number of controllers, data collectors and deputy test directors. These personnel must be able to observe and record critical activities during the tests. This requirement usually means that these persons must have access to operations centers, command posts, fire direction centers and other key locations. Accommodation of test personnel can become a problem and can place unnecessary pressure on the unit being evaluated if the activities are not carefully planned and controlled. Herein lies the real challenge to the operational tester. Test design plans must be thorough and efficient. The methodology must be used to optimize test data without hampering any portion of the Army Training and Evaluation Program. Equally important is the perception by members of the tested unit that the data collected from the test will affect the evaluation during the Program. *The test director can play an important role in precluding such apprehensions by ensuring that test planning specifically calls for careful separation of test data from that required for the Army Test and Evaluation Program.* Such test control matters should be highlighted during pretest coordination meetings.

- ***Insufficient replication of events.*** Replication, in a scientific sense, is the systematic repetition of a test event under identical conditions based on the approved test design plan. Replication often is required to reduce the chance of errors entering into the collection of test samples. In a training exercise any significant amount of replication becomes difficult as the tactical realism of the exercise tends to diminish. A scenario driven exercise or the use of side excursions (tactical snapshots) can often provide needed replication of events.

SUMMARY

The Army's own study of the materiel acquisition process supported the need for a separate and independent Operational Test and Evaluation Agency. The Army recognized that because of the increasing expense and complexity of materiel systems the systems must be placed in the hands of typical user troops as early as practicable in the development process and evaluated prior to costly production decisions. The challenge facing the operational tester is to be able to accomplish that mission at minimum cost.

Guided by the important phrase "In the hands of typical user troops," one attractive way of fulfilling the operational tester's mission might be an increase in the use of training exercises to accomplish test and evaluation objectives. The Army's operational tester, the Operational Test and Evaluation Agency, continues to depend upon the resources of Forces Command, TRADOC, and other organizations to accomplish operational tests of major and selected nonmajor (Category I) systems. The cost effectiveness of integrating operational testing with training exercises is evident. However, other operational considerations must be examined to insure that this approach meets approved test objectives. The advantages to be gained from the integration of operational testing with training exercises, when feasible, far outweigh problem areas encountered to date. The bonus effects are improved training opportunities and increased combat readiness. Although training exercises cannot be viewed as a universal solution, these exercises do offer an attractive option for conduct of operational test and evaluation.

REFERENCES

Cited References

1. Dept of Army, "Operational Test and Evaluation Methodology Guide," US Army Operational Test and Evaluation Agency, May 76.

2. ________, "SOTAS (REFORGER) Test Design Plan," US Army Combat Developments Experimentation Command, Jul 76.

3. ________, "Army Training and Evaluation Program for General Support Cannon Units, Department of the Army," ARTEP No. 6–165, Sep 76.

4. ________, "Corps Automation Requirements Baseline Identification Test, Test Design Plan," FDTE Project No. FO-049, Project CAR Test HQ, XVIII Airborne Corps; HQ, US Army Training and Doctrine Command; and HQ, US Army Logistics Center; 10 Sep 76.

5. ________, HQ, III Corps, AFZF-DPT-PLX-E to CINC US Readiness Command, Ltr, subject: "Executive Summary, GALLANT CREW 77 After Action Report," 6 May 77.

Additional References

Dept of Army, OTEA, CSTE-EDS, Memorandum for Record, subject: "OTEA *Position on* SOTAS OT I," 30 Nov 76.

________, "Army Training and Evaluation Program for Light Infantry Battalions," ARTEP 7–14, 30 Apr 76.

"Project CAR Aids Commanders," *The Paraglide*, 37: 8 (1977) Ft Bragg, NC, 10 Mar 77.

Dept of Defense, "Major System Acquisitions," Directive, Number 5000.1, 18 Jan 77.

________, "Test and Evaluation," Directive, Number 5000.3, Ch 2, 20 May 77.

Lt Col Martell D. Fritz is Assistant to the Scientific Advisor, US Army Operational Test and Evaluation Agency, Falls Church, VA. Lt Col Fritz has served in command of Battery B, 2d Howitzer Battalion, 5th Field Artillery, USAREUR, 1964–65; Battery B, 2d Battalion, 19th Field Artillery, 1st Cavalry Division, Viet Nam, 1967; and, as Commander, USA NCO Academy/Drill Sergeant School, Fort Knox, KY, 1974. He has held staff assignments at the Gunnery Dept, Fort Knox, KY, 1967–69; in the TACFIRE Project Managers Office, Fort Belvoir, VA, 1969–71; and prior to his present assignment, in the office of J1, United Nations Command, Korea, 1972–73.

Lt Col Fritz, a contributor of articles to "The Field Artilleryman," has completed both the Field Artillery Officers Basic Course, 1962, and the Artillery Officers Advance Course, 1966, at Fort Sill, OK. He later completed the US Army Command and General Staff Course at Fort Leavenworth, KS, 1972.

Lt Col Fritz received his BS, Military Science from the US Military Academy in 1961 and an MS, Systems Management, University of Southern California, 1975.

CHALLENGES IN AIRCRAFT ENGINE TEST AND EVALUATION PROGRAMS

by

David T. Love, Manager
Engine Evaluation Engineering, Aircraft Engine Group
General Electric Company

Many important factors affect an aircraft engine test and evaluation program, whether for military or commercial application. These factors in turn affect the engine development cycle requirements and the overall test and evaluation process. Often new challenges are presented and sometimes surprises as each new system is developed. In this article the author looks at a few of the factors and notes how they relate to the development cycle. Some areas of improvement are suggested and the basic requirements of the test and development cycle are given.

THE ENGINE DEVELOPMENT CYCLE

A typical cycle in the development of major engine components into an advanced state-of-the-art engine is shown in Figure 1. This development cycle is typical of what is required to assure that a new engine meets design requirements. The cycle is based on 30 years of engine development experienced at one company. As shown, a period of basic applied research and component testing precedes full-scale engine development to bring forth a state-of-the-art application of advanced technology. Then, as engine testing is initiated, both component and full-scale engine testing progress toward qualification (or certification), flight test and eventually into production. What is not shown are the advanced technology Advanced Turbine Engine Gas Generator (ATEGG), Aircraft Propulsion System Integration (APSI) type demonstrator programs that precede these events. The demonstrator programs provide the technology base for the new systems and are vital to maintaining technological leadership. In production, service life verification and continued improvement activities logically follow. Few shortcuts have been found that can compress this development cycle. Attempts to leave out steps in the evaluation process usually mean a costly return to do an omitted task.

Comparison of the test and evaluation cycles of an engine for a commercial vs a military application shows the requirements to be parallel. In almost every case, the commercial engines of today were derived from engines originally developed for a military application. Many of the major components—such as the complete core engine system of the TF39/CF6 derivative engines—were first developed and qualified under a military program. As a result, the derivative engines can be certified in a shorter time period when a relatively modest advancement in technology is accomplished. An example of this is the CF6–6 and –50 commercial engines, which were derived from the TF39 engine. The TF39 engine was an outgrowth itself of a military sponsored research and development program (the Advanced Turbine Engine Gas Generator).

As shown in Figure 2, the engines were developed and certified in a relatively short time. Greater emphasis was placed on developing the improved durability and maintainability characteristics. These enhanced characteristics were required by the airlines to maintain a competitive position in the industry relative to cost-of-ownership and utilization rate requirements.

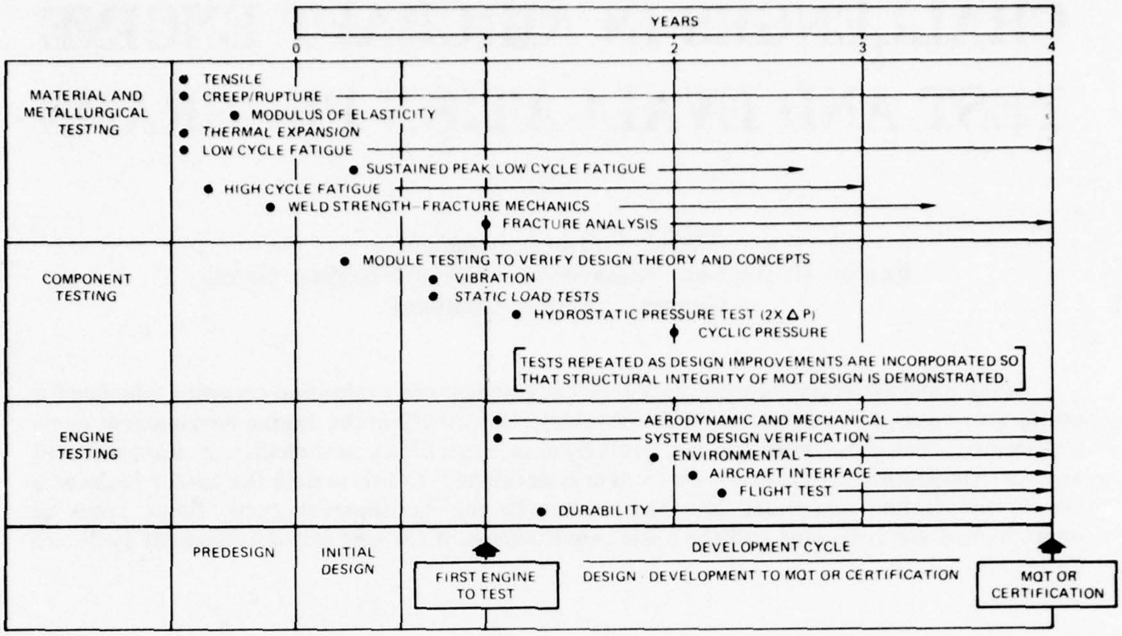


Figure 1. Typical Aircraft Engine Component Development Cycle

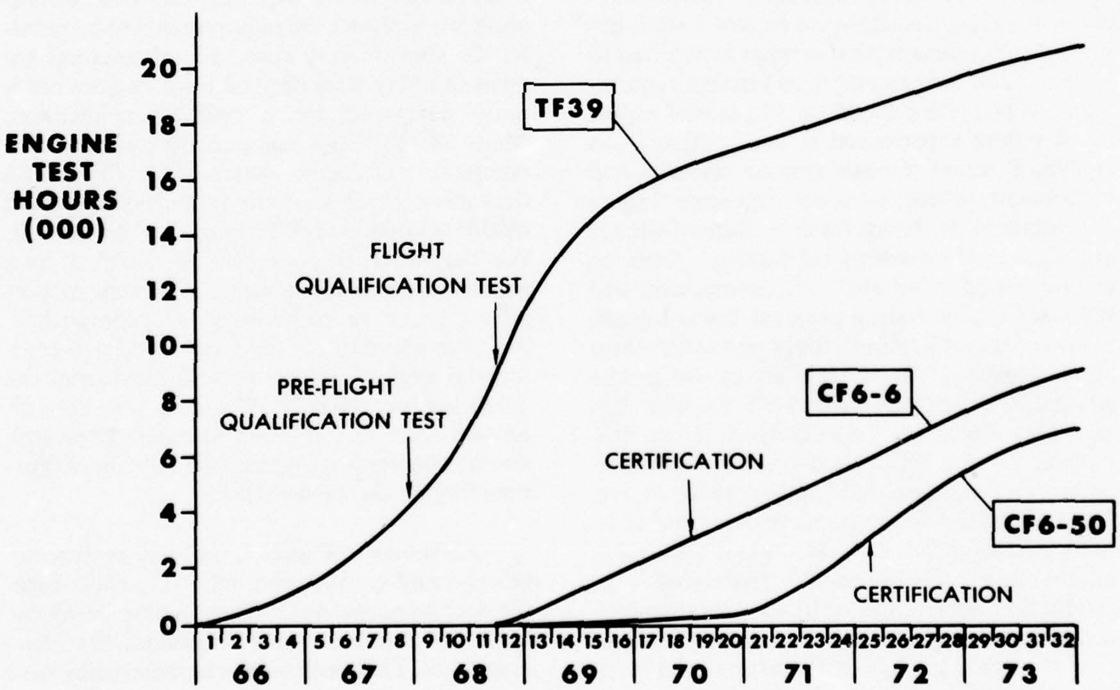


Figure 2. Engine Test Requirements (for military and derivative commercial engines)

Note that because of the derivative engine relationship most of the durability and maintainability improvements were incorporated into the TF39 engine.

Preproduction certification test requirements are—and have been for some time—getting tougher. Emphasis has been shifting toward a better balance among performance, initial development costs, and cost-of-ownership objectives. The producers and users (companies and customers), both military and commercial, are getting more sophisticated. The demands for better simulation and more accurate representation of operating conditions and capabilities have resulted in a complex series of test demonstration requirements. The reliability and serviceability of engines are continually improving and each of the new development programs, in one specific company, has been structured to bring the engines to a development stage of greater maturity. These actions are taken to minimize the risks and potential costs involved in a premature production commitment. This translates to a more complete development program and also to increased test requirements and increased test development program costs.

As an example, General Electric recently completed a very extensive and complicated series of engine and component qualification tests on the F101 engine, which is being developed for the B-1 strategic bomber. The test requirements included two complete back-to-back 157 hour extra severity endurance cycle tests on the same engine, with power usage cycles and operating conditions tailored to the B-1 mission profiles. Three hundred and seventy hours were accumulated on this engine representing more than 1000 hours of anticipated service operations. Simultaneously, another engine—installed in the engine altitude test facility at Arnold Engineering Development Center—completed a thorough performance calibration test series, inlet distortion tolerance demonstrations, and stall margin capability demonstrations at several key points in the flight envelope. These two particular tests—which represent only part of the overall series of qualification test demonstration requirements—easily surpass anything that has been accomplished on prior test programs.

If this test program has succeeded in providing an accurate simulation of the environmental operating conditions and power usage requirements that the engine will be exposed to in service, then initial service life and reliability objectives will have been achieved. Achievement of these objectives will in turn result in a smoother and more trouble-free flight operation, and a significant long term cost savings to the program.

There are a few areas where some meaningful improvements could be made in the test and evaluation process.

NEW APPROACHES TO DEVELOPMENT TESTING

The basic challenge is clear. Costs are increasing. Test requirements are increasing and test costs are increasing. We must, on a continuing basis, push to improve our state-of-the-art test simulation and test measurement capabilities to obtain an increased quantity of meaningful information from each test—more information per test hour and more information per dollar. New techniques both in data acquisition and reduction are available and must be implemented to speed and improve the quality and quantity of data to be obtained in the total evaluation process. This means investment in equipment, research, software, and new approaches to development testing.

An entirely new computer data system has been designed to improve engineering evaluation effectiveness. This data system will have a major impact on the way engine tests are conducted and on the management of test programs. The key to the new system is the availability of an interactive graphic display that the engineer can use to view the data being taken "on line" and plotted in engineering units. These plots will be compared with the predicted results, immediately, in a graphic display so that decisions can be made on the course of the test program and the success or failure of the test being conducted.

This new system will impact the management of the test and evaluation process by providing a number of benefits and challenges. Some of the benefits:

- **More data will be available on line and plotted for faster learning and decision making.**
- **The data as presented graphically can be edited so that only data of interest is analyzed.**
- **A hard copy plot of the data will be available eliminating the time and "dog work" of plotting data.**
- **Test time can be reduced by a faster assessment of data (elimination of unnecessary test points) so that testing can proceed.**
- **An increased percentage of data can be analyzed.**
- **Engineering time can be used more effectively for analysis and decisionmaking.**

Some challenges will be:

- **A greater demand on engineering personnel to predict test results for on line comparison and analysis.**
- **A need to have alternate courses of action as test results uncover problem areas.**
- **Test and evaluation programs will demand more forethought.**
- **As the evaluation process is speeded, a demand for faster problem fixes and test turn around will follow.**

Figure 3 shows a typical type of graphic plot available on-line where the actual test data is compared with the engineering prediction. In addition, an on line calculated projection gives the test director an indication of what may lie ahead as the test proceeds.

Many improvements in engine and component test facilities are being made to keep abreast of the advancements in engine design. High energy X-ray of a running engine, with visual representation of mechanical phenomena occurring inside the engine, is being used to augment analytical techniques. Some recent work with laser measurement techniques to obtain jet velocity profiles or, in one case, icing cloud droplet size information, appear very promising. Certainly, one of the key elements in better test and evaluation lies in the improvement of the testing facilities both in government and industry.

ESTABLISHING AND INTERPRETING REQUIREMENTS

A better job needs to be done of establishing and interpreting basic "design-to" and "test-to" certification requirements. Improvement is needed in the ability to feedback and interpret information gained from operational experience. Improvement is needed in the ability to translate basic cost and usage objectives into meaningful test requirements for use in providing valid technical measures of essential capabilities. It is to be recognized that accurate simulation of *all* inservice operating conditions is impossible. Many of our current requirements are based on subjective, and sometimes inaccurate, interpretation of "lessons learned in the field." Also, the pass-fail connotations associated with certification test programs too easily can be interpreted as a guarantee that unforeseen problems will not occur in the field after an engine has been certified.

In normal service, engines ingest tires, birds, runway debris, etc., and are expected to operate with minimum damage and achieve a high standard of reliability. Normal factory testing requires limited ingestion testing of small, medium, and large typical objects in

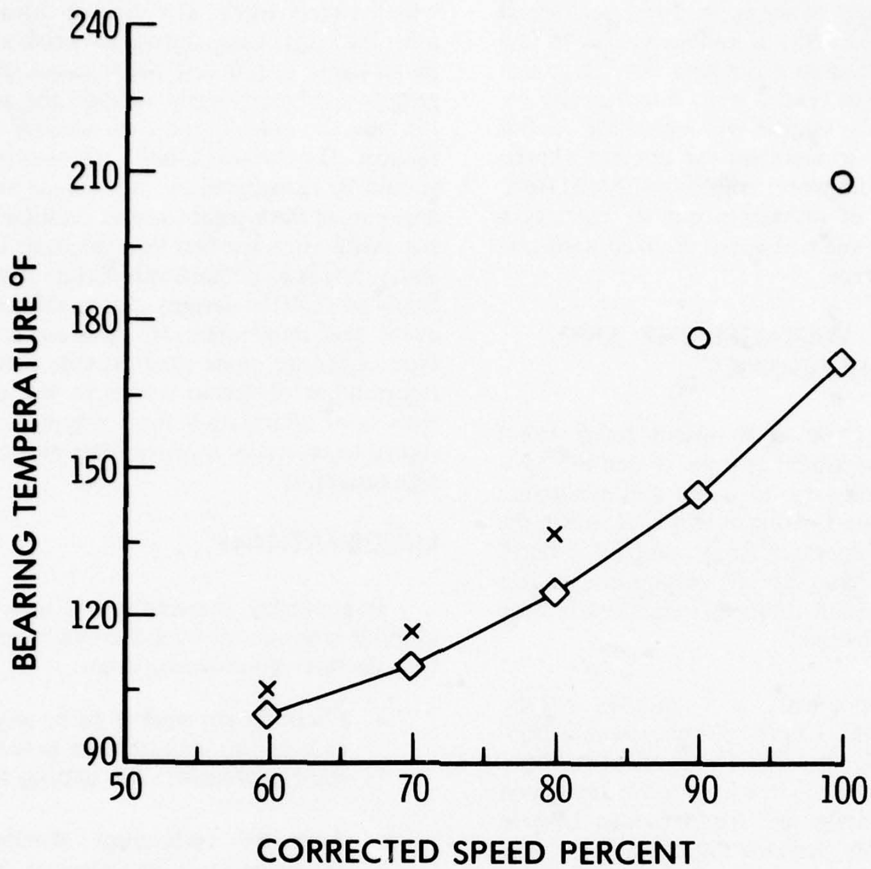


Figure 3. Demonstration of Predicted Values for Bearing Temperature Together with Measured Data (X's) and On Line Calculated Projections (O's)

specified quantities. If these certification tests are passed, the engine is determined to be satisfactory for service.

Experience has shown that ingestion damage, particularly that caused by bird strikes, is quite unpredictable. One ingestion incident may not produce damage while another ingestion incident of the same type may produce significant damage. Statistically, a single ingestion test run in the development program may not produce the results seen in operational service and thus conclusions as to the satisfactory operation of the engine during ingestion experiences may not be correct.

An approach in trying to make engine development testing more compatible with field experience is to produce the effects of major foreign object ingestion, that may include high imbalance owing to blade breakage, and to operate the engine with predicted levels of imbalance to determine the effects on engine operation. This approach removes the variability of certain ingestion tests and substitutes the damage predicted for a worst instance of field ingestion. It is extremely beneficial to determine problems of this kind during development testing in the factory rather than after the many hours of service operation anticipated to produce such ingestion caused damage.

An example of the type of test performed to evaluate the effects of imbalance due to foreign object damage is given here. Two large sections of two fan blades were intentionally severed while the engine was operating at full takeoff power to simulate the damage experienced under the worst conditions in the field. This amount of imbalance can be used as a "baseline" for the evaluation of all components within the system.

PROGRAM MANAGEMENT AND THE "RIGHT-TO-FAIL"

Finally, a few words about development program management aspects. A well thought-out and properly executed test and evaluation program plays a significant and vital role in the development process of many complex systems. The selection and use of certain major test events as program progress indicators is both natural and effective.

The development of a complex system, such as an advanced propulsion system, is typically an iterative process. Feedback of information gained from the test program is translated into refined "design-to" requirements. Also, in some cases, this feedback information is translated into refined "test-to"/"certify-to" requirements as in the object ingestion problem just discussed.

In a typical development program trade-offs are necessary. Unforeseen problems do occur, and management system flexibility is needed to keep the program on track and to maintain a proper balance among cost, schedule, and performance objectives.

The selection and interpretation of individual test program events as major contract milestones must be done carefully to avoid placing too much emphasis on the "successful completion" of any one particular test and to avoid the potential conflict between test program technical objectives and milestone schedule objectives. A related problem is that of the almost universal interpretation of test hardware failures as being "all bad." This interpretation is due in part to the high visibility environment in which testers work. The assessment and ultimate trade off of test objectives vs risk factors—particularly during the early stages of a program—must necessarily include the potential for, and the consequences of, various types of failure. The "right-to-fail" (in the test cell) should be recognized as an inherent and valuable part of the typical test and evaluation process. Aside from the fact that the basic intent of some tests is to demonstrate failure modes and limits so that the designs can be refined, a test event that encounters an unforeseen, unique type of failure mode may provide meaningful information of greater value to the program than does information from a typical test designed to ascertain or prove some particular design capability.

OBSERVATIONS

In summary, there appear to be a number of guide posts in meeting the challenges in future aircraft engine development.

- **There do not appear to be any short-cuts in the development process, and the requirements are getting tougher.**

- **Advanced technology demonstrator programs are a key element in maintaining technological leadership and are the building blocks for new applications.**

- **Continued investment in testing and test measurement capabilities is absolutely necessary to meet future technical challenges.**

- **A better job must be done in establishing and interpreting the "design-to" and "test-to" requirements to insure an adequate evaluation program.**

- **Improvement in the ability to translate in-service usage objectives into meaningful and accurate test requirements is needed.**

Finally, recognition must be given the importance of "successful failures." There is no substitute for good judgment, team work, and management wisdom—and a little bit of luck—if we hope to continue to improve the overall test and evaluation process.

Mr. David T. Love is Manager of the CF6/TF39 Engine Evaluation Engineering, Aircraft Engine Group, General Electric Company, Evendale, Ohio. Mr. Love has had more than 15 years experience in the development of gas turbine engines for aircraft applications.

Mr. Love received his BSME degree from the Carnegie Institute of Technology, Carnegie Mellon University.

RELIABILITY AND MAINTAINABILITY IN THE ACQUISITION PROCESS DOD Directive 5000.x

by

Colonel Ben H. Swett, USAF

An effort to establish overall Department of Defense (DOD) policy on how reliability and maintainability should be addressed during each phase of the acquisition process is underway. The "for comment" draft of a new 5000-series directive, DODD 5000.x, entitled "Reliability and Maintainability (R&M) of Systems and Equipment," was released to the DOD Components and defense industry associations in April 1977. In this article the author discusses the provisions of DODD 5000.x, the underlying rationale, and some of the comments received. Formal coordination and implementation of this directive are expected early in 1978.

PURPOSE

The purpose of DODD 5000.x is to improve the life cycle cost effectiveness of systems and equipment acquired by the Department of Defense. Major goals are to increase operational readiness and effectiveness, reduce maintenance and logistic support cost, and reduce acquisition cost and schedule. Phase-by-phase policy provisions are designed to implement the policies of the Office of Management and Budget OMB Circular A-109, and DOD Directives 5000.1, 5000.2 and 5000.3. Provisions regarding Military Standards for reliability and maintainability are designed to implement the findings of the Defense Science Board Task Force on Specifications and Standards. Five specific objectives are outlined here.

OBJECTIVE 1

Establish the reliability and maintainability characteristics to be managed as major performance parameters of all defense systems and equipment.

Reliability and maintainability can be characterized, defined and measured many different ways. Therefore, *it is necessary to select* the general reliability and maintainability characteristics that link all other performance requirements to operational readiness, effectiveness and ownership costs. The selection will provide a basis for performance/life cycle cost/effectiveness trade-offs. The general reliability and maintainability characteristics selected must be managed as quantitative goals and thresholds during system acquisition to ensure sufficient management attention.

OBJECTIVE 2

Ensure that realistic reliability and maintainability requirements are established for all acquisition programs and achieved in all systems fielded by the DOD.

Emphasis is on realistic, as opposed to idealistic, requirements. Included are both quantitative reliability and maintainability requirements and the series of tasks necessary to achieve these requirements. It was necessary to state that reliability and maintainability requirements *must* be established for all acquisition programs (a practice not now followed in all cases). The objective is to achieve reliability and maintainability requirements in *field service*, rather than in theoretical predictions or laboratory tests.

OBJECTIVE 3

Integrate reliability and maintainability tasks into the acquisition process by identifying the basis for evaluation of reliability and maintainability achievements at each milestone decision.

Assigning specific reliability and maintainability tasks and insuring task accomplishments are Program Manager responsibilities. Evaluating accomplishments of the preceding phase and plans for the next phase are responsibilities of the program review and decision authorities. Both the Program Manager and the decision authorities need a general outline of what should and should not be expected at each milestone. That "report card" is the point at which reliability and maintainability considerations can become an integral part of program management, review and decisionmaking. However, it will not happen until the overriding acquisition cost and schedule problems of the Program Manager and higher authorities are taken into account along with reliability and maintainability considerations.

OBJECTIVE 4

Minimize acquisition cost and schedule, consistent with the above objectives.

Several comments recommended changing this objective to read, "Optimize acquisition cost and schedule...." Such comments could not be accepted. The intent is to accomplish the other objectives through *more efficient* investment in reliability and maintainability programs, not by driving acquisition cost higher and pressing schedule beyond present bounds. There are some specific places where more needs to be invested in reliability and maintainability than has been usual in the past, but the net result must be lower per unit cost and more rapid delivery. Reduction of total cost and schedule has not been a fundamental objective of reliability and maintainability programs, but it must become so in the present cost and schedule environment if reliability and maintainability programs and projects are to survive.

OBJECTIVE 5

Promote continuing improvement in reliability and maintainability engineering and assurance technologies to meet current and future system program reliability and maintainability requirements.

Some comments were that the provisions of DODD 5000.x do not support this objective. The comments may have been based on a misinterpretation of what the objective implies. The objective does not mean continuing refinement of classical practice, but response to a new and broader range of responsibilities. To meet the provisions of DODD 5000.x, Program Managers and higher authorities will need different kinds of assistance from reliability and maintainability staff than has been received in the past. Calling for more than one *reliability* characteristic and more than one *maintainability* characteristic will create challenges for allocation, prediction, design, test and evaluation of those requirements. This, in turn, will create the need for innovation in the reliability and maintainability specialties and underlying technologies.

APPLICABILITY AND SCOPE

The 5000-series Directives, including DODD 5000.x, apply to the Office of the Secretary of Defense, the Military Departments and the Defense Agencies. These directives apply to contracts only insofar as they direct certain activities of the Program Manager or procuring activity. Nevertheless, DODD 5000.x has been developed with contractual implications in mind, and with a conscious effort to incorporate accepted principles of sound business practice. Comments from the defense industry associations were considered on an equal basis with those of the Military Departments and Defense Agencies.

Department of Defense Directive 5000.x differs from other 5000-series Directives in one major respect; that is, it applies to less-than-major systems, subsystems and equipment as well as "Major defense systems." The expansion was necessary because reliability and maintainability problems are not confined to "Major defense systems," and do not originate at the system level of assembly. As in the story of the war that was lost "for the want of a horse-shoe nail," it is at the lower levels of assembly that the correction of most reliability and maintainability problems begins.

DEFINITIONS

The "for comment" draft of DODD 5000.x included eleven terms and definitions. Several comments recommended all terms and definitions be deleted in favor of the "accepted standard terminology" presented in MIL-STD-721. That was not done, because standard reliability and maintainability terminology does not support the performance/cost/schedule or life cycle cost trade-offs required by current DOD acquisition policy, and because Military Standards are contractual documents that must reflect, but do not establish, DOD policy.

The definitions in DODD 5000.x are intended to:

- direct the DOD Components in the selection of existing reliability and maintainability terms that link performance requirements to readiness, effectiveness and ownership cost,
- distinguish between government and contractor responsibilities,
- outline what is meant by an "Integrated" test,
- bring Qualification and Acceptance Tests more nearly in line with established auditing principles, and
- establish a basis for projecting field service reliability and maintainability values from demonstrated test results.

None of these problems has been solved by current policy or practice. Current DOD policy requires clarification of terminology, and is beyond the scope of Military Specifications and Standards. Therefore, although wording changes have been incorporated in response to specific comments on the definitions in DODD 5000.x, it remains necessary to establish a few basic distinctions in reliability and maintainability terminology as a matter of DOD policy.

Definitions of Major Reliability and Maintainability Characteristics with Commentary

Department of Defense Directive 5000.x does not prescribe a set of detailed terms and definitions for all kinds of systems and equipment. The directive defines those general reliability and maintainability characteristics that must be addressed as major performance requirements, and then directs the DOD Components to select (or define) appropriate terms for each of those characteristics, as the characteristics apply to specific systems and equipment. In essence, DODD 5000.x distinguishes between reliability and maintainability as measures of effectiveness and as factors of ownership cost.

- ***Mission Reliability*** will include only missions or mission time, and only those failures that are mission-critical, at a stated level of assembly.

Mission reliability is a factor of mission effectiveness and must be specified as a requirement for all systems and equipment. Mission reliability has little relation to availability or operational readiness, because the system is assumed to be ready at the start of each mission. Mission reliability has little relation to ownership cost, because mission reliability counts only mission-critical failures that occur during the course of a mission. Redundant design and alternate modes of operation increase mission reliability but also increase system complexity, total parts count, and acquisition cost. Complex systems with a high parts count require more maintenance and more spare parts in field service, because there are more parts to fail. Efforts to increase mission reliability are directly related to higher—*not lower*—life cycle cost.

- ***Maintenance Reliability*** will include all potentially useful operation and all failure indications, for a stated level of assembly and all of its component parts.

Maintenance reliability is a new term that refers to reliability as a factor of system demand for maintenance and logistic support. Since every part can create a demand for maintenance, maintenance reliability counts every failure *indication*, whether or not it is mission-critical, whether or not it is subsequently confirmed as a "material failure," and regardless of redundancy in system design. Since failure indications initiate maintenance regardless of time of occurence, maintenance reliability counts all potentially useful operation and not merely missions or mission time. Maintenance reliability, no matter how defined and measured, must answer the question: "How often will this system require maintenance"?

When "failure indication" is defined by detailed performance requirements, maintenance reliability becomes the link between those requirements and the effects of those requirements on availability, operational readiness, maintenance manpower requirements and logistic support cost. Redundant design and alternate modes of operation do not increase, but rather decrease maintenance reliability. One way to improve maintenance reliability is to build simpler systems having a lower parts count. Thus, both acquisition cost and cost of ownership are reduced. Maintenance reliability, *not* mission reliability, is directly related to lower system life cycle cost.

The introduction of maintenance reliability as a separate and distinct reliability characteristic generated much comment and considerable resistance. Apparently a number of people do not realize that investment in reliability programs cannot be justified on the basis of improved readiness or reduced ownership cost when only mission reliability is specified as a system performance requirement.

"Material reliability" was proposed as a substitute for "maintenance reliability" on the basis that it is a well-recognized term. This proposal was not acceptable, because material reliability only counts "confirmed material failures (faults)." In some current equipment, 85 percent of the Built-In-Test failure indications are not confirmed as material failures. Hence mean-time-between-failure (Confirmed) exceeds mean-time-between-failure (Indicated) by a factor between 6 and 7. That is why projections of availability, readiness, manpower and total logistics cost tend to be optimistic when based on demonstrated material reliability. However, material reliability values are needed for the planning of spare parts inventories. Future versions of DODD 5000.x will note that material reliability may be specified as a major reliability characteristic, in addition to, but not instead of, mission and maintenance reliability.

- ***System Maintainability*** will include all system (clock) downtime from initiation to completion of maintenance, excluding delays not dependent upon system design characteristics.

Maintainability, as measured in clock time, is a factor of system availability and operational readiness. Since maintainability is a system design characteristic, it cannot include maintenance delays for such things as towing, getting personnel to and from the job, waiting for spare parts or technical data, etc. And since maintainability is a factor of *system* availability, it cannot include clock time for off-system maintenance of detached components when the system itself is operationally ready. Maintenance reliability asks, "How often will this system require maintenance"? System maintainability asks, "How long will the system be down when it does require maintenance"?

The "for comment" draft of DODD 5000.x called out "mean-time-to-repair (MTTR)" for this system maintainability characteristic, when the standard definition of maintainability—which is the probability that maintenance will be completed within a specified amount of time—would have been equally acceptable. The point is that maintainability is a factor of availability and thus of operational effectiveness; it is not an accurate measure of cost, because manpower and material costs of off-system maintenance and repair are not taken into account. That is why the use of quick-remove-and-replace components improves maintainability and availability, but does not significantly reduce total maintenance or logistic support costs. Also that is why investment in maintainability programs cannot be justified on the basis of ownership cost reduction when only the classic definition of maintainability is specified as a system design requirement.

- ***Maintenance Manpower*** will include all maintenance personnel or all maintenance manhours required to support a given system at all prescribed levels of maintenance and repair.

Maintenance manpower is one of the largest factors of ownership cost for any system that requires maintenance. This factor is sensitive to two reliability and maintainability characteristics—maintenance reliability and maintenance manhours per maintenance action—and it does not go away when on-system maintenance has been minimized in the interests of greater system availability. Maintenance manpower is now being addressed, for some systems, by such terms as "maintenance manhours per flying hour (MMH/FH)"—which is not an accurate measure because it combines the effects of variation in flying hours per month, maintenance actions per flying hour and manhours per maintenance action, without distinguishing among them. The draft DODD 5000.x called out "maintenance manning ratio" for this characteristic—only to have numerous comments point out that it is better to allocate manhours rather than fractions of people as requirements for subsystems and equipment. However it is measured, until this characteristic is addressed as an integral part of the Program Manager's responsibilities, there will be little control over maintenance manpower cost in field service during the course of system acquisition.

- ***Maintenance Material Cost*** will include all replacement parts and components required to support a given system at all prescribed levels of maintenance and repair.

Maintenance material cost is sensitive to two reliability and maintainability characteristics—*material reliability* and *average parts cost.* Maintenance material cost is sensitive to certain design policies and strategies—but in the wrong direction. For example, high-reliability parts typically cost from twice to ten times as much as the commercial-grade equivalents, but these parts do not always last twice to ten times as long. Since less than ten percent of the parts typically cause more than 90 percent of the failures in field service, mandatory use of 100 percent high-reliability parts becomes questionable. In the same way, the use of sealed components as a way to avoid maintenance-induced failures means that the entire component must be removed, shipped, repaired, returned and replaced every time the weakest part in the component experiences a failure. The aggregate impact of such policies and design strategies does not show up unless something like average material cost per repair is managed as a system design characteristic. The Draft DODD 5000.x did not mention the maintenance material cost characteristic, but several comments pointed out the need for it or a similar characteristic.

- ***Availability*** is the probability that a system will be in specified operational condition if called upon to perform its mission at any random point in time.

Availability was not selected as a major reliability and maintainability characteristic, even though it is in wide usage and a number of comments recommended that it be included. Availability can be derived from maintenance reliability and maintainability (downtime). Further, availability is not directly measurable during system acquisition. Future versions of DODD 5000.x will note that the term *availability* can be used instead of *system maintainability,* but not instead of *maintenance reliability,* because the latter is necessary as a factor of ownership cost.

RELEVANT VS CHARGEABLE

Use of the term "nonrelevant failure" has caused a great deal of confusion. In some cases, the Program Manager has deleted a significant percentage of the failures actually observed during a reliability test because the failures were caused by factors beyond the control of the contractor. Thereafter other agencies found that operational readiness, effectiveness and ownership cost calculations based on demonstrated test results were optimistic. In other cases, the Program Manager has deleted only those failures that were obviously test-peculiar. In these cases the contractor was penalized for failures over which he did not have control. None of these choices is an acceptable solution to the problem of analyzing and using reliability test data.

The DODD 5000.x attempts to resolve the dilemma by recognizing that reliability test results must be used for two different purposes: projection of field service reliability values, and determination of contract compliance. The directive establishes a clear distinction between *relevant* failures and *chargeable* failures and expands on that distinction to include clearly defined responsibilities for maintenance downtime, manpower and materiel costs.

- ***Relevant Failures. Relevant failures include all failures, downtime hours, maintenance manhours and materiel costs incurred during test that can be expected to occur in field service.***

- ***Nonrelevant Failures. Nonrelevant failures include only those failures, downtime hours, maintenance manhours and materiel costs incurred during test that are caused by a condition, external to the equipment under test, that is not encountered in field service.***

"Relevant" and "Nonrelevant" refer to field service and to the responsibility of the government Program Manager or procuring activity. The terms therefore are equally applicable to Government Furnished Equipment and Contractor Furnished Equipment components of the item under test. If a significant portion of field service reliability and maintainability problems are found to originate in Government Furnished Equipment components of the system, it is the responsibility of the Program Manager to have the components repaired or replaced.

Note that all test data are considered relevant unless and until such data are proved to be test-peculiar. Only then may this test data be deleted from projections of field service reliability and maintainability values.

- ***Chargeable Failures. Chargeable failures are those relevant failures, downtime hours, maintenance manhours and materiel costs incurred during test which are caused by any of the goods or services provided by a given contractor.***

- ***Nonchargeable Failures. Nonchargeable failures incurred during test include only those relevant failures, downtime hours, maintenance manhours and materiel costs which are caused by, and are dependent upon, a condition previously stipulated as not within the responsibility of a given contractor.***

"Chargeable" and "Nonchargeable" are subsets of "relevant." If test data are not relevant to field service, the failures are not chargeable to the contractor. The limits of contractor responsibility must be defined in advance and stipulated as scoring rules for any test used to determine compliance. "Chargeable" does include the effects of contractor-furnished software, support equipment, training, operation, maintenance or repair procedures, etc., as well as the hardware components of the product. Where goods and services supplied by more than one contractor are involved in the same test, "chargeable" may be more specifically defined to distinguish between the responsibilities of the various contractors. The term "nonchargeable" cannot be used to delete test data from projections of field service reliability and maintainability values.

INTEGRATED TEST

One member of a recently established Program Office surveyed all applicable Military Regulations, Specifications and Standards to see how many tests were required for an avionics subsystem. Thirty-nine separate and distinct kinds of test were found, some of which had to be run several times during the course of the acquisition program. This number did not include tests required for higher or lower levels of assembly. From similar experience throughout DOD, it appears that a larger number of tests or test programs are not needed. What is needed is a rational and cost-effective approach to testing in general.

Such an approach for major weapon systems, at the system level of assembly, is specified in DODD 5000.1, 5000.2 and 5000.3. However, the Military Standard tests for subsystems and equipment are fragmented. Performance is tested under ambient conditions, without consideration of reliability or maintainability. Environmental qualification tests apply high—and in some cases, extreme—stress levels, without equipment operation or without recording equipment reliability under stress. Reliability tests demand extensive equipment operating time, under conditions that do not adequately simulate the operational environment, and under a definition of "failure" that does not begin to count all the failures defined by specified performance tolerance. Maintainability tests are not required for subsystems or equipment, and none of the other tests require a record of maintainability data. Result: Equipment can go through all the standard tests without revealing many—let alone most—of the problems the equipment will display in field service.

Reliability tests are notorious for taking a long period of time or a large number of samples. Department of Defense acquisition policy calls for test results as inputs to the production decision. But the Program Manager does not have many hours of test time between the end of Full-Scale Engineering Development and the start of production. Further, the Program Manager cannot afford a large number of samples for reliability testing. Present cost and schedule constraints dictate a reduction in the amount of time required for reliability testing. Many detailed provisions scattered throughout DODD 5000.x are designed to do just that.

Although specific tests must be tailored to the mission and operational environment of the equipment under test, a few principles of test cost-effectiveness apply in all cases. The DODD 5000.x specifies one integrated test wherein appropriate environmental stresses are combined insofar as practical, and where specified performance, reliability and maintainability characteristics are tested simultaneously.

PROVISIONS OF INTEGRATED TEST

- Test conditions shall reproduce measured or predicted environmental stresses in field service, based on a trade off between maximum test realism and affordable test facilities.

- Test procedures shall simulate operational use and performance monitoring of the equipment under test.

- Reliability and maintenance data shall be recorded during the test, and testing shall be of sufficient duration to measure demonstrated reliability and maintainability values.

- Separate tests that duplicate portions of the more realistic integrated test shall not be required.

Integrated testing serves two purposes: Correlation between test results and field service data is increased by increasing test realism, and the time required for reliability tests is reduced by increasing the rate at which any given number of failures will be discovered. The test time reduction is roughly proportional to the degree that present reliability test conditions are less failure-producing than integrated test conditions. The DODD 5000.x does not state how to structure an integrated test for any system, subsystem or equipment. The directive states principles of test cost-effectiveness as policy for the direction in which the DOD components must move if reliability and maintainability test programs—and more specifically, laboratory test programs—are to become truly cost-effective.

INDEPENDENT TEST AND EVALUATION

Even though the requirement for independent test and evaluation was specifically limited to preproduction qualification tests and production acceptance tests, this provision of DODD 5000.x generated a storm of protest. The DOD components said compliance would be too expensive. The defense industries do not like others testing their equipment. The comments are correct in one respect: the issue of independent test and evaluation overshadows DODD 5000.x—and may overshadow DODD 5000.1, 5000.2 and 5000.3 combined. The issue was surfaced in response to the management principles incorporated in DODD 5000.3, "Test and Evaluation," (including the standard auditing principle that says no one should be asked to evaluate his own product), and a question from the House Appropriations Committee that asked, "Is it true that defense contractors test their own products on behalf of the government"?

It is true. That is the almost-universal practice in current reliability and maintainability programs. Government inspectors are responsible for monitoring qualification and acceptance tests at the larger plants. Smaller plants and vendor facilities are visited only occasionally. Nevertheless, once a product passes these tests and is placed on a Qualified Parts List all higher tier contractors have to use the product unless a waiver can be justified (and obtained). If the higher-tier contractors receive faulty parts, neither these contractors nor the Program Managers have much recourse when the parts already have been accepted as compliant with government specifications.

Department of Defense Directive 5000.x could not resolve the inherent conflict of interest involved in having every producer test his own products on behalf of the government. The directive merely defined independent test and evaluation as "a test conducted by a government or commercial agency having no vested interest in the outcome of the test," and noted that independent test and evaluation is called for by DODD 5000.3. Then DODD 5000.x provided three options—in priority—for the testing of subsystems and equipment.

1. Independent T&E is always preferred for Preproduction Qualification and Production Acceptance Tests, but it is not always feasible nor cost-effective.

2. The second choice is to have prime contractors test the products of their subcontractors, and subcontractors test the products of their vendors, under government surveillance.

3. The third choice is to have the producer conduct Preproduction Qualification or Production Acceptance Tests of his own product under strict government surveillance and independent evaluation of the test results.

The options were considered to be the most logical compromise between further support of the current practice—which essentially reverses these priorities—and the unknown cost impact of requiring truly independent test and evaluation across the board. The directive leaves selection from among these priorities to the Program Manager and the degree of enforcement to higher program review and decision authorities. The fundamental question of independent test and evaluation vs inherent conflict of interest is a matter for the executive levels of the DOD and the United States Congress.

TEST REALISM

Classical definitions of reliability and maintainability all contain the caveat: "under specified conditions." The definitions of major reliability and maintainability characteristics in DODD 5000.x give recognition to the fact that reliability and maintainability data are applicable only to a specified set of conditions. There are two sets of conditions—laboratory test and field service. The relation between these two conditions must be known or estimated if quantitative reliability and maintainability requirements and achievements are to be useful as decision information. Why measure reliability and maintainability values in a laboratory, or use laboratory tests to determine contractual compliance, if one has no idea of the relation between laboratory tests and field service conditions?

Test Realism: The degree to which a specified set of test conditions and procedures simulates a specified operational environment.

This is the reciprocal of the "Environmental K Factor" long used to adjust parts failure ratings for different applications. As shown here, the degree of test realism would be less than 1.0 if the test conditions were less failure-producing than the operational environment; and, greater than 1.0 for accelerated or overstress testing. Realism would have to be obtained by comparing demonstrated values for the same reliability and maintainability characteristics, using the same definitions and the same or similar kinds of equipment, under the two sets of conditions and procedures: Laboratory test and field service.

An estimated degree of test realism is necessary for the translation of field service reliability and maintainability requirements into specified values for a test that does not perfectly simulate field service. Also an estimate of test realism is necessary as an adjustment to the values demonstrated during test when projecting the values expected in field service. Until laboratory test results can be related to field service reliability and maintainability values, the results are useless as inputs to operational readiness, effectiveness or ownership cost models. Even though establishing rough estimates of test realism will require effort on the part of the DOD Components, until it is done, reliability and maintainability or integrated tests cannot be justified on the basis of cost-effectiveness because there is no way to ascertain the cost-effectiveness.

POLICY

The draft DODD 5000.x has been criticized as having gone into too much detail for a DOD Directive. The criticism may have merit. However, there is a pattern to the comments that makes it impossible just to shorten the Directive. First, every provision in the draft was designed to correct a specific problem that has a large and widespread impact on readiness, effectiveness, ownership cost, and acquisition cost or acquisition schedule. Second, those who have encountered one of these problems almost invariably approve of the provision designed to correct it. Third, those who have not encountered a particular problem—or have not recognized it—almost invariably consider the provisions to be unnecessary detail. Except for the provision concerning independent test and evaluation, where the uproar was well-nigh unanimous, the comments to other provisions tend to cancel each other. Future versions of DODD 5000.x will be shorter and more pointed—but will be accompanied by DOD Instruction 5000.x that will include more explanation than is appropriate in a Directive.

GENERAL

It was necessary to state the policy that reliability and maintainability characteristics *will be* considered major performance parameters of all defense systems, subsystems and items of equipment. Presently these characteristics are not so considered. It was necessary to state the minimum set of reliability and maintainability characteristics because those characteristics now selected as performance parameters often fail to support required trade offs or life cycle cost reduction. It was necessary to require that each reliability and maintainability characteristic be addressed at each program milestone decision and contractual review. Too often these characteristics are ignored. Other general policy provisions and the applicable underlying rationale are as follows:

Specific Reliability and Maintainability tasks are to be defined and documented as an integral part of both government and contractual program plans, rather than as an additional, sideline "R&M Program." This action is to ensure that these tasks get the necessary degree of management attention throughout the acquisition process.

Quantitative Reliability and Maintainability Requirements and Achievements in DOD program documents—such as the initial statement of an operational requirement, the Decision Coordinating Paper, Program Memorandum, and equivalent management documents within the DOD components—are to use the terms and definitions by which reliability and maintainability achievements will be measured in field service. There has been unnecessary confusion between field service and contractual definitions of the same reliability and maintainability terms. The confusion results in numbers of wide variation. Explanation should not have to be made to the Congress about why the numbers measured in early field service do not mean the same thing as the numbers stated in the Decision Coordinating Paper.

Quantitative Reliability and Maintainability Values in Contracts differ from field service values. It is not feasible to hold the contractor responsible for failures, downtime hours, maintenance manhours and materiel costs over which he has no control. Laboratory test requirements differ from field service requirements because a set of test conditions does not exist that can accurately simulate the field service environment. Design requirements differ from test requirements by the margins used to establish a specified degree of statistical confidence or decision risk. It is not a matter of using one and only one number for each reliability and maintainability characteristic. The numerical traceability of reliability and maintainability requirements and achievements must be established through a form of audit trail. Department of Defense Directive 5000.x does not tell the DOD components how to establish numerical traceability, but it does say it must be done. Also the Directive does not indicate the major factors that must be included in the audit trail. As stated in several comments, some of the correlation factors do not exist. It will take time and effort to develop these factors. Nevertheless, until reliability and maintainability requirements and achievements are numerically traceable, program review and decision authorities will not know what any reliability or maintainability number really means, and neither will the Program Manager.

Theoretical Reliability and Maintainability Predictions based on paper studies of equipment designs were prohibited as a basis for source selection. The prohibition made a number of people unhappy, even though it complied with the spirit and the letter of all 5000-series directives—directives that make it clear that major decisions are to be based on demonstrated test results rather than paper studies. The prohibition was directed against two common problems: First problem: the cases where one contractor's "predicted mean-time-between-failure" is competed against another's in the process of source selection. The practice establishes a market—or an auction—in which the contract goes to the best predictor. The second problem is the practice of evaluating predicted reliability and maintainability values against field service or test requirements, instead of against design requirements derived via an audit trail such as is discussed here. The procedure allows the design to be accepted without compensation for a series of adjustment factors—which means trouble later when the design is put to test. In those cases where it is absolutely necessary to base source selection* on paper predictions rather than on demonstrated test results the procedure is allowed (if predictions are evaluated against design requirements and are not used as a basis for direct competition among the bidding contractors).

Operational Readiness, Effectiveness and Ownership/Life Cycle Cost Models have been entered with contractor's predictions. The results of these models have been used as a basis for source selection. In addition to creating a paper auction, this procedure leads to highly

* *The* major program decision for the contractor.

optimistic field service estimates. Thus, it was necessary to state only that those *field service* reliability and maintainability values established as requirements to be demonstrated prior to the production decision are to be used to enter such models at any point in the program before field service reliability and maintainability values, based on demonstrated test results, are available. Thereafter, the models must be updated by demonstrated test results, and the outputs of the models must be available as inputs to the production decision.

Integrated Performance, Reliability and Maintainability and Environmental Testing was stated as a policy for all program phases. This policy is *required* only for Preproduction Qualification and Production Acceptance Tests. The policy is to ensure Program Manager authorization to cut down on duplicative or overlapping tests. The policy was not intended to mandate full-blown "Mission profile, combined stress, integrated laboratory tests" where such tests are neither appropriate nor cost-effective. Future versions of DODD 5000.x will clarify this stipulation.

Statistical Estimation Within Specified Confidence Limits was called for, in addition to hypothesis test plans in documents such as MIL-STD-781. Reason: hypothesis tests cannot be used to project the reliability and maintainability values expected in field service. Hypothesis tests yield one of two answers, "Accept" or "Reject." The true values demonstrated by the test are ignored. Statistical estimation techniques support the idea that statistics should be the servant, rather than the master, of acquisition cost and schedule. Statistical estimation techniques provide the Program Manager some confidence in the true values demonstrated as the basis for making a nonpredetermined management decision. The Program Manager then does not have to spend the time necessary to fulfill predetermined accept/reject criteria. Department of Defense Directive 5000.x did not say that is the way it should be done, but there are cases where the Program Manager has to make such a decision. In these cases the Program Manager needs all the help he can get. The Program Manager does not need the restrictions imposed by enforcement of hypothesis test plan assumptions.

Fixed-Length Test Plans were given preference over the "standard sequential test plans," especially for preproduction qualification tests, for a number of reasons. First, all sequential test plans contain built-in cost and schedule uncertainties that the Program Manager cannot control. Second, if program cost and schedule is planned on the basis of the "expected decision point" in standard sequential test plans, the inherent test time uncertainty becomes a built-in cost and schedule overrun that the Program Manager cannot control. (For example, in MIL-STD-781 Test Plan III, there is a 10 percent probability of a 100 percent overrun even though true mean-time-between-failure is equal to the specified mean-time-between-failure. Third, fixed-length test plans can be used in conjuction with statistical estimation techniques. Sequential test plans cannot. Fourth, while some cautionary notes with regard to the inherent problems are included in the recent revision of MIL-STD-781, it is not the only document that contains sequential test plans. Such test plans are often applied without reference to cautionary notes. It was absolutely necessary to state that fixed length test plans *are to be preferred for Preproduction Qualification* Tests where the Program Manager cannot afford schedule uncertainty—much less a built-in overrun—and that when sequential test plans are used, program cost and schedule are to be planned on the basis of "maximum allowable test time," rather than "expected decision point." Thus inherent test time uncertainty results in an underrun rather than an overrun.

Preliminary Studies. The reliability and maintainability characteristics of current systems and forces must be considered as factors of their operational readiness, effectiveness and ownership cost when Mission Area Analyses are performed and when the Mission Element Need Statements (MENS) are prepared. Circular A-109 and DODD 5000.1 prohibit the establishment of specific requirements that presuppose a system solution to the mission need at the onset of the program. The MENS can say about reliability and maintainability only something like, "We need a system with twice the readiness and half the maintenance personnel of the XXX and YYY systems now being used in this mission area."

PROGRAM PHASES

This section of the Directive is essentially a checklist for program reviews prior to each milestone decision. Stated are the reliability and maintainability activities that should be evaluated as accomplishments of the preceeding phase. The activities at each of these reviews should, after evaluation, be addressed as requirements for the following phase. This portion of the directive is structured to implement the policies of OMB Circular A-109 and the other 5000-series DOD Directives.

CONCEPTUAL PHASE

In the Conceptual Phase appropriate reliability and maintainability characteristics must be selected or defined for each type of system being considered as a candidate. Firm quantitative requirements are prohibited. The primary concern must be to ensure that the right reliability and maintainability characteristics are selected and defined in measurable terms. The second concern must be to ensure that a realistic baseline is established for the quantitative requirements that are to be developed later in the program. The third concern must be to prepare a baseline for estimation of technical risk. Many of the problems that plague current acquisition programs can be traced to an idealistic, over-optimistic or ill-defined baseline for the reliability and maintainability requirements.

> ***Two values are required for each major reliability and maintainability characteristic. The first value shall be measured from a similar system or equipment in field service. The second value shall be a tentative field service requirement for the new system or equipment. Both values shall use the terms and definitions selected for the new system or equipment.***

This provision is specifically intended to avoid the practice of generating reliability and maintainability requirements by theoretical analysis of hypothetical missions and with little or no consideration of past and present field experience. Also the provision is intended to ensure that development planners obtain and use field reliability and maintainability data. Too, the provision is to ensure that tentative reliability and maintainability requirements are field service values, rather than attempts to compensate for:

- degree of laboratory test realism,
- the difference between field service and contractual definitions, or
- any margins for statistical confidence or decision risk.

All of these latter adjustments are to be made later in the program, by the Program Office or procuring activity.

> ***Tentative reliability and maintainability requirements shall be based on the sensitivity of operational readiness, effectiveness and ownership cost of similar systems and equipment to changes in the numerical value of each major reliability and maintainability characteristic.***

Tentative reliability and maintainability requirements must not be arbitrary, and must not be set several times higher than measured field service reliability and maintainability values from similar systems without making clear why such improvement is needed. For example, the difference between 90 and 99 percent probability of mission success increases required mean-time-between-failure (Critical) by an order of magnitude. That factor drives the entire acquisition program. Although a 9 percent increase in mission success may be essential, the cost and schedule impact of attaining the increase must be understood and justified from the beginning. Since the new system is still in the earliest stages of concept formulation at this point, the potential payoff for any tentative reliability and maintainability requirement can be estimated only by analysis of field experience with similar systems.

> ***Tentative reliability and maintainability requirements shall be annotated as being within current state-of-the-art or as requiring a significant advance in reliability and maintainability technology.***

The difference between measured field service reliability and maintainability values for a similar system and tentative reliability and maintainability requirements for a new system must be used to provide initial visibility on the degree of technological advance being sought for the new system. We must answer the question, "Is this requirement technically feasible"? The answer must be an estimate of technological risk, based on analysis of experience with similar acquisition programs and a review of the technology base. One or more reliability and maintainability characteristics may be improved by a factor of 2 or 4 or 10 over the previous generation of systems and equipment—but we need to be able to explain why this is believed to be possible, rather than saying, "Do it," thus handing the Program Manager an impossible dream.

DEMONSTRATION AND VALIDATION PHASE

The Demonstration and Validation Phase is where the Program Office has to translate *tentative* field service reliability and maintainability requirements into *tentative* contractual requirements and include these requirements in design solicitations. Requests for alternative design proposals must explain the reliability and maintainability characteristics sought, and must provide industry with sufficient information so that industry may furnish a meaningful response. Conversely, each offeror must be free to propose his own technical approach, main design features, and alternatives to schedule, cost and capability goals. Detailed government specifications and standards are to be avoided during all program phases short of Full-Scale Engineering Development (OMB Circular A-109). Therefore, there is a set of reliability and maintainability activities which are—and a larger set which are not—appropriate during the Demonstration and Validation Phase.

Design Solicitations to Industry shall include:

1. the criteria to be used in measuring reliability and maintainability achievements after the system is deployed,
2. measured reliability and maintainability values for a similar system and tentative reliability and maintainability requirements for the new system,
3. a defined operational life profile to include one or more missile profiles, and
4. either measured environmental stress data from similar systems or guaranteed access to such data as input to the initial design process.

The Department of Defense wants responding contractors to look at reliable, maintainable performance in field service, rather than at the minimum values needed to pass a series of stylized laboratory tests. The contractors must be told to do this and the government must provide the information to the contractors to enable the contractors to achieve the desired result. The contractors must be assisted as well as motivated to design reliability and maintainability into their systems.

> ***Mathematical Modeling*** of various design approaches, to include parts derating and system packaging, shall be used to ensure that all field service requirements, including reliability and maintainability characteristics, are being met. Iteration of these models shall be emphasized as a means of establishing a sound design approach; however, the quantitative values they produce shall not be used as measures of merit for source selection, either directly or as input data for life cycle cost estimates.

At this point government expertise is needed to evaluate how responding contractors propose to go about ensuring that field service reliability and maintainability requirements are met. Several comments to the Draft DODD 5000.x said that the draft directive did not place sufficient emphasis on designing reliability and maintainability into the system. This is where such emphasis belongs. What the contractor proposes to do about stress analysis, parts derating, worst case and sneak circuit analysis, failure modes and effects analysis, and "test-fix-test" engineering development all are more important than is organization of reliability and maintainability staff functions or how high a "predicted mean-time-between-failure" can be produced. Special attention is needed to ascertain how well the responding contractors have integrated performance, reliability and maintainability and environmental considerations in contractor system design and program plans. Although legal staff in both government and industry want and need some kind of hard criteria for source selection, efforts to meet that need cannot substitute for sound engineering judgment. How engineering judgment can be brought to bear in source selection decisions that are made before development, before test data are available, and under prohibitions on the use of Military Specifications and Standards is a problem larger than DODD 5000.x—and it is a major reason for the emphasis on continuing competition into engineering development.

> ***Prototype Demonstrations*** shall be used to refine theoretical reliability and maintainability predictions and validate the proposed design approach. These demonstrations shall test performance, reliability and maintainability characteristics and environmental stresses simultaneously. Full qualification tests shall not be required during the demonstration and validation phase unless the program bypasses engineering development and goes directly into production.

Prototype demonstrations are spot tests intended to explore and resolve particular risks. It is important that the spot tests not be fragmented to the extent that a design is optimized for performance on a laboratory bench. However, sophisticated "mission profile" qualification tests are not appropriate when early prototypes do not represent the design approach and equipment configuration intended for production. For example, it makes little sense to spend much time and money on full-blown qualification tests of "chip and wire" prototypes when large scale integrated circuits will be employed in the production unit. Qualification tests *are* necessary if a decision is made to produce the validation phase prototype.

> ***Trade offs***. Tentative reliability and maintainability requirements shall be subjected to continuing review for technical feasibility and trade offs to balance their future effects on operational readiness, effectiveness and ownership cost with their more immediate effects on acquisition cost and schedule.

Such trade offs are the Program Manager's responsibility. This provision gives notice that reliability and maintainability requirements are not sacred. The provision also indicates that requirements shall not be eliminated or reduced to ease acquisition cost and schedule difficulties *without* explaining the detrimental effects of such reductions on the field service cost-effectiveness of the system. Program review and decision authorities above the Program Manager are alerted to the potential danger of "robbing Peter to pay Paul" when reliability and maintainability requirements are reduced. It is the responsibility of these authorities to make sure the proposed trade offs are in fact life cycle cost-effective. That is about as far as a DOD Directive can go in attempting to resolve what has long been a major problem in the way reliability and maintainability considerations have been addressed during system acquisition.

> ***Enforceable Reliability and Maintainability Requirements*** shall be presented for approval at Milestone II and included as quality assurance provisions in the specification for engineering development. Each reliability and maintainability characteristic shall be represented by: a mandatory growth

function or series of intermediate management thresholds for development testing, and the minimum acceptable value and confidence level to be demonstrated prior to the production decision. Reliability and maintainability growth shall not be presented as continuing into production unless concurrent development and production has been specifically authorized.

Those requirements shown in Section 4, *Quality Assurance Provisions,* of an equipment Specification are enforceable; those requirements shown elsewhere are not. The same is true of environmental stresses and test conditions/procedures. The contractor can be told to design for "world-wide" use, but in fact he has to design for the criteria in Section 4 because that is what the government is paying him to prove.

The purpose of engineering development is to mature the approved system design. Engineering development is the place for "reliability growth"—and maintainability improvement. Effective management requires that reliability and maintainability growth be measured against intermediate points in a series of intermediate milestones so decisions can be made early to modify the equipment. Required growth must reach a minimum value before the system is evaluated, and the system must be evaluated prior to production. Growth does not happen automatically nor without investment of time and money. If growth is predicted to continue through production and into field service, that predicted "growth curve" implies concurrent development and production and maturation in field service—whether or not the implication is made clear. Rather than allow the implication of concurrence to go unrecognized—which may well make it a "primrose path" for the Program Manager and higher decision authorities—DODD 5000.x requires that growth requirements be made explicit. The Program Manager and higher decision authorities then are free to decide each case on its own merits.

Reliability Test Duration. The maximum test time for demonstration of compliance with reliability requirements prior to the production decision shall be established prior to Milestone II and included in quality assurance provisions of the equipment specification for engineering development. Test time and number of test articles shall be selected to balance: (1) total test cost and impact cost of any addition to length of program schedule owing to reliability testing, (2) the cost increase of requiring the contractor to design more reliable equipment or to fabricate a larger number of samples simply to pass a shorter reliability test, and (3) the desired degree of statistical confidence or decision risk.

This is a major performance/cost/ schedule trade off that is now obscured in the selection of a Military Standard reliability test plan. The trade off involves a three-way choice: (1) increase design reliability to shorten test time and number of samples, (2) add test time and samples to reduce pressure on system design, or (3) reduce statistical confidence (increase "decision risk") to reduce pressure on the design, total test time and number of samples. The Program Manager and the contractor need to know the cost of statistical confidence or decision risk before entering into a contract for Full-Scale Engineering Development.

FULL SCALE ENGINEERING DEVELOPMENT PHASE

In this phase both DOD and industry need to place more emphasis on reliability and maintainability considerations than has been done in the past. Although reliability and maintainability have to be "designed into the system," a team of human beings cannot anticipate and eliminate all design deficiencies. The choice is to find the remaining deficiencies and eliminate them during engineering development, or wait for the deficiencies to surface in early field service—where every design change takes longer and costs more. Research and Development (Test and Evaluation) funds are not plentiful.

Hence DODD 5000.x makes a sharp distinction between development testing that is intended to improve the equipment and qualification tests that are intended to determine the compliance of a particular configuration. The directive emphasizes the former and limits the latter. Also the directive is written to ensure that improvements actually are incorporated—because if the improvements are not made, the business of engineering development is a waste of time and money.

> ***Development Testing***. Equipment used for full-scale engineering development shall be operated under progressively more realistic test conditions and procedures to mature the design as rapidly as possible by disclosing and correcting latent design deficiencies.

There is nothing unusual in this definition of development testing. What is new is the emphasis on finding and correcting *reliability* and *maintainability* deficiencies as well as performance limitations. The DODD 5000.x states that Development Test and Evaluation at the system level of assembly is the same thing as Test-Analyze-Find-Fix (TAFF) at the subsystem and equipment levels of assembly. The contractor should conduct Development Test and Evaluation in his own facilities, using his most qualified people (especially failure analysts and redesign engineers), under minimum constraints and with minimum administrative delay. Since this is the only kind of testing specifically intended to get design deficiencies corrected, the government needs to ensure this type testing is not squeezed out of the program. Stipulation of a minimum as well as a maximum time for development testing is one way to inform the contractor the government is serious about maturing the design before producing the item.

Two problems are typical of reliability and maintainability development testing. One problem is that a kind of "reliability growth" can be obtained just by replacing "failed" parts. The replacement amounts to screening faulty parts out of the full-scale development models, and the apparent reliability growth does not transfer to production units. The DODD 5000.x states that such growth indicates the need for 100 percent preconditioning ("Burn-In") of production units. The other problem is the handling of failures, downtime, maintenance manhours and materiel costs that a proposed design change is expected to eliminate. As stated in DODD 5000.x all test data are relevant to field service unless and until a design change is actually incorporated and is verified as being effective in reducing reliability and maintainability problems. The old reliability and maintainability values then become test-peculiar on the basis that they pertain to a former design configuration. Future versions of the directive will have to place emphasis on design change to improve maintainability characteristics. Several comments pointed out that maintainability is "locked in" once the design is frozen.

> ***Incorporation of Design Changes***. Cost-effective reliability and maintainability design changes should be incorporated and verified as early as possible during Development Test and Evaluation/Test-Analyze-Find-Fix.
>
> Design changes shall be incorporated and verified to provide representative equipment for test and evaluation *prior* to the production decision. Acquisition strategy shall ensure that sufficient time and resources are available to incorporate cost-effective reliability and maintainability improvements at this point in the program.

This provision is self-explanatory. All too often when deficiencies are found, there is an inability to correct them because cost and schedule planning are too success-oriented. Granted that engineering development is funded with a different color of money than production, and that less funds are available from Research and Development (Test and Evaluation) than from procurement accounts, the high return-on-investment for funding little "fixes" with big payoffs in terms of field service reliability and maintainability make this provision necessary.

Test and Evaluation Prior to the Production Decision. Final preproduction prototypes shall be tested under the most realistic simulation of field service available to provide demonstrated reliability and maintainability values for evaluation prior to Milestone III.

Test and evaluation prior to the production decision has been a major aspect of DOD acquisition policy for the past 7 years—since former Deputy Secretary of Defense David Packard signed his memorandum on system acquisition in May, 1970—but it is only now beginning to be reflected in Military Standards on reliability and maintainability. The DODD 5000.x requires that performance, reliability and maintainability and environmental qualification tests be integrated insofar as practical and conducted prior to production. The qualification of "first production units" is not an acceptable substitute for the testing of final preproduction units. Preproduction qualification tests of subsystems and equipment are the same as Initial Operational Test and Evaluation (IOT&E) of systems. Both of these kinds of test should be:

- short to avoid the impact cost of program delay,
- realistic to project field service reliability and maintainability values based on test results as inputs to the production decision, and
- independent of the developer to avoid the inherent conflict of interest involved in having the contractor test and evaluate his own product.

Several comments pointed out the need for clarification regarding maintainability demonstrations at the system level of assembly. The demonstrations are different from the collection of maintainability data from an integrated test of subsystems or equipment because demonstrations can and should be accelerated by the introduction of simulated failures (as opposed to waiting for failures to occur). Maintainability demonstrations must be conducted by personnel having the skill level, tools, test equipment and technical publications specified for field service. The use of highly qualified scientists and engineers simply invalidates the demonstration and makes the results optimistic. Some comments said it is impossible for the contractor to have tools, test equipment and technical publications ready prior to the production decision. If so, it is questionable whether the system—being a combination of hardware, software and people—is ready for production and deployment. Again, if concurrent development, production and deployment are the fact, that fact should be made explicit and judged on its merits, rather than obscured in a reliability and maintainability program.

The DODD 5000.x stated that the reliability aspects of a qualification test need not be prolonged beyond the point at which minimum acceptable values have been demonstrated at minimum acceptable confidence levels. Accepting equipment on the basis of demonstrated confidence levels increases the "consumer's decision risk," that is, the probability of accepting less than the minimum acceptable value. However, DODD 5000.x did not say that the Program Manager has to stop the test and accept the equipment on the basis of demonstrated confidence levels. The directive said that such a decision is within the authority of the Program Manager if he chooses to exercise this prerogative. Such a decision is appropriate when the Program Manager has selected a fixed-length reliability test plan and the equipment demonstrates significantly more than minimum acceptable reliability. While this appears to be a highly technical detail of statistical theory, it is in fact a performance/cost/ schedule trade off that Program Managers need to be aware they can make. In some cases, sacrificing the last 10 percent of "consumer's decision risk," without sacrificing "demonstrated confidence level" can cut the time required for a fixed-length test by 85 percent.

To ensure that decision authorities know what the reliability and maintainability values shown at Milestone III really mean, DODD 5000.x states that demonstrated test results will be:

- measured within a specified confidence interval,

- adjusted to remove any difference between field service and contractual definitions,

- adjust for degree of test realism, and

- compiled to project field service reliability and maintainability values at the sytem level of assembly.

This is the "achievements" side of the "requirements/achievements audit trail." The entire audit trail does not have to be presented at Milestone III reviews, but decision authorities have to know it has been accomplished and the audit trail must be available on request to substantiate the reliability and maintainability values presented as achievements.

Enforceable reliability and maintainability requirements for the production and deployment phase must be presented at Milestone III. These too, are field service reliability and maintainability values. If these values are higher than those demonstrated by Initial Operational Test and Evaluation and preproduction qualification tests, the Program Manager should be prepared to explain exactly how that improvement will be obtained and who will pay for it. A theoretical "growth curve" is not acceptable at this point unless concurrent development, production and maturation in field service is part of the production decision. Even then, the growth has to be planned, funded and managed, or it cannot be expected to occur.

Once approved at Milestone III, field service reliability and maintainability requirements in the Decision Coordinating Paper or equivalent DOD documentation must be translated into contractual reliability and maintainability requirements and included in Section 4 of the equipment Specification for the production phase. That makes the requirements enforceable criteria of contractual compliance, to be evaluated by production acceptance tests.

PRODUCTION AND DEPLOYMENT PHASE

Again, DODD 5000.x emphasizes those tests intended to improve the equipment and de-emphasizes those intended to prove or disprove compliance, in terms of the amount of *time and money spent on the tests*. The directive calls for 100 percent inspection and strongly recommends 100 percent preconditioning ("Burn-In") to reduce infant mortality in field service. Further, the directive calls for Integrated (Mission Profile) Production Acceptance Tests, with real "teeth" in terms of contractual compliance, but the directive limits the tests to a sampling in the interests of cost and schedule reduction.

> ***Preconditioning ("Burn-In").*** All deliverable end items shall be submitted to a short test, at the highest practical level of assembly, to disclose weak parts and manufacturing defects for correction prior to delivery.

There were many potentially useful comments in regard to this provision. The difference between "Inspection," "Parts Screening," "Burn-In," and "Preconditioning" was discussed. Several comments highlighted the need for management flexibility with regard to how preconditioning is done and how it can be changed during the course of the production run. None of the comments said preconditioning is not needed nor that it does not more than pay for itself—for both the contractor and the government. The statement in DODD 5000.x to the effect that environmental stresses for preconditioning should be selected to disclose defects, rather than to simulate the operational environment, was well received. So was the provision that these tests must be done by the contractor. However, there was some disagreement as to whether preconditioning tests should be managed against "accept/reject criteria." The position taken in the draft Directive was that it is unwise to penalize the contractor for finding and fixing defects, since that is exactly what the government wants the contractor to do.

Production Acceptance Tests. After items have completed the preconditioning test and have been certified ready for delivery, a few samples shall be selected at random from each production lot and resubmitted to the preproduction qualification test.

Production acceptance test conditions and procedures must be identical, within specified tolerances, to those used for preproduction qualification tests of the same equipment. A more stringent test is not fair to the contractor. A less stringent test allows the contractor to shave corners after the government is committed to produce the design.

Contractual requirements for production acceptance tests must be enforceable. That is, "accept" must mean "compliant," and "reject" must mean "noncompliant," not "take corrective action and retest." (The automatic "reject and retest" feature of MIL-STD-781B has been removed from MIL-STD-781C because this feature meant that, in practice, the only way to stop a reliability test was to accept the equipment—one way or another. Any similar features in as-yet-unrevised Military Standards must be waivered as a matter of policy.)

All-equipment production acceptance tests are discouraged, but not prohibited. These tests are far more expensive than preconditioning tests, because of the need for realistic "Mission Profile" conditions and procedures. All-equipment production acceptance tests were not prohibited because there are a few cases—such as in man-rated spacecraft—where such testing can be justified.

Like preproduction qualification tests and initial operational test and evaluation, production acceptance tests should be independent of the producing contractor. Once a product is on the shipping dock and certified ready for delivery, the contractor should be prepared to have the product delivered to field service or diverted to an independent laboratory that will test the product under specified conditions and procedures designed to simulate field service.

Field Evaluation of early deployed systems shall be the primary method of determining compliance with field service reliability and maintainability thresholds approved at Milestone III. Program plans for production and deployment shall contain provisions for: (a) feedback of measured reliability and maintainability values to the Program Office and the contractor(s), (b) determination of compliance, and (c) rapid correction of residual deficiencies and defects.

The field evaluation provision is neither new nor exotic. The Congress evaluates achievement of reliability and maintainability requirements in the Decision Coordinating Paper against reliability and maintainability values collected by the maintenance data reporting system used in field service. A problem is generated when the definitions differ. Contractors keep asking for feedback of field service reliability and maintainability data. There is no apparent reason why the contractors should not have this information as a matter of course from systems they have produced, or on request from other systems. Several systems have been fielded recently with what amounted to a guarantee, by the developing command to the operating command, that the system would meet Milestone III reliability and maintainability requirements or be corrected by the developing command. That kind of policy is worthy of encouragement and support as a matter of DOD policy.

Field Service Warranties. The production and deployment phase may contain provisions for either developing agency or contractor support of deployed systems. These provisions shall distinguish between maintenance support and further reliability and maintainability improvement.

This distinction is necessary to prevent support and improvement provisions from working against each other. Maintenance support contracts must be defined with care and with regard to who is responsible for various kinds of failures, downtime hours, maintenance manhours and materiel costs. Improvement contracts need broad definition to allow and encourage the contractor to identify needed improvements in hardware, software, training, manning, organization, etc. Where these two concepts have been mixed in the same contract, the results are expectable: the need for improvement will be found in an area not covered by the maintenance contract and the provisions of the maintenance contract will not be enforceable.

RESPONSIBILITIES

Effective management demands clear definition of areas of responsibility, with commensurate authority. Implementation of this principle demands an answer to the question, "Who is involved in this activity"? Where the subject is the reliability and maintainability of defense systems and equipment, the answer must include mention of the operating commands who write initial requirements, the developing commands who meet those requirements, and the supporting commands who inherit the final product. Effective management must include the Program Managers, procuring activities, contractors and subcontractors. Management *must include the program review and decision* authorities in the chain of command above any given Program Manager and the independent review agencies that support those authorities. Effective management must include, but can no longer be limited to, the professional reliability and maintainability staff functions in all of these agencies. Given that a DOD Directive sets policy but does not tell the DOD Components or the contractors how to implement those policies, or how to organize, DODD 5000.x has addressed reliability and maintainability as a command responsibility.

PROGRAM MANAGER

The designated Program Manager or procuring activity is responsible for the field service reliability and maintainability characteristics of both Contractor Furnished Equipment and Government Furnished Equipment. The Program Manager is responsible for selectively tailoring reliability and maintainability tasks, integrating these tasks into each phase of the acquisition program, and ensuring that the tasks are satisfactorily accomplished. Too, the Program Manager is responsible for the translation of field service reliability and maintainability requirements into contractual requirements, and for the translation of contractual reliability and maintainability achievements into projected field service reliability and maintainability values. Further, the Program Manager is responsible for the inclusion of quantitative reliability and maintainability values as an integral part of performance/cost/schedule and life cycle cost trade offs within established thresholds. Portions of specified authority may be delegated to the contractor or to outside agencies, but the Program Manager retains overall responsibility and authority to manage the reliability and maintainability aspects of the program.

INDEPENDENT REVIEW AGENCIES

The independent test and evaluation agencies in each DOD Component are responsible for reviewing reliability and maintainability tests and test results in accordance with the provisions of DODD 5000.3, "Test and Evaluation." The Cost Analysis Improvement Group is responsible for reviewing the acquisition and ownership cost impact of reliability and maintainability requirements and achievements in accordance with the provisions of DODD 5000.4, "OSD Cost Analysis Improvement Group." Organizations having equivalent functions will perform these reviews for less-than-major system acquisition programs managed within the DOD Components.

Several comments to the draft DODD 5000.x recommended that an Independent Reliability and Maintainability Review Authority, similar to the Independent Test and Evaluation Agencies and Cost Analysis Improvement Groups, be established in each of the DOD Components. Such recommendations were found to conflict with established OMB and DOD policies with regard to management layering and restriction of the Program Manager's authority. It was felt that another review would add administrative delays to the acquisition schedule, and that establishing reliability and maintainability as a separate organization would not serve to integrate reliability and maintainability considerations into the mainstream of program management. Future versions of DODD 5000.x will continue to treat reliability and maintainability as system performance and design characteristics. This approach allows the DOD Components to staff reliability and maintainability functions, at all levels of organization, as necessary and appropriate to comply with the substantive provisions of the Directive.

LINE AUTHORITIES

The line authorities, in the chain of command above a given Program Manager or procuring activity, are responsible for the reliability and maintainability thresholds they approve. The line authorities shall ensure that:

- reliability and maintainability thresholds are realistic,
- reliability and maintainability values used as decision information are relevant to field service and are based on demonstrated test results,
- projected field service reliability and maintainability values satisfy operational readiness, effectiveness and ownership cost requirements at affordable acquisition cost and schedule, and
- the Program Manager or procuring activity is provided with sufficient resources to enable compliance with the provisions of this DODD 5000.x.

The degree to which reliability and maintainability considerations become an integral part of defense system acquisition depends upon the degree of management attention received from program review and decision authorities. Reliability and maintainability staff who report to these authorities must ensure that all reliability and maintainability considerations are clear and presented in a form amenable to high level decisionmaking. Once reliability and maintainability characteristics are properly defined, and line authorities know that these measurable characteristics are key, controlling factors of operational readiness, effectiveness and life cycle cost, a greater degree of management attention can be expected.

MILITARY DEPARTMENTS

The Military Departments are responsible for:

- establishing realistic reliability and maintainability requirements for all acquisition programs by ensuring adequate coordination among operating, developing and supporting commands,
- providing measured field service reliability and maintainability data and environmental stress data to the Program Offices and contractors,
- maintaining an adequate reliability and maintainability technology base and corporate memory in support of present and future acquisition programs, and
- revising Military Regulations, Specifications and Standards as necessary to comply with the provisions of DODD 5000.x.

Realistic reliability and maintainability requirements must be based on experience rather than on theoretical analysis of the reliability and maintainability values desired. The operating commands now write initial requirements for new systems and equipment. The supporting commands will have to become more involved in writing the maintenance reliability and maintainability portions of those requirements. Effort will be required to make use of existing reliability and maintainability and environmental stress data, and to fill the present gaps where data is lacking. The alternative is continued reliance on stylized Military Standards that do not deliver satisfactory reliability and maintainability in field service. New technology will be needed to actually improve reliability, and that means greater attention to reliability on the part of both government and industry laboratories. Major efforts will be required to upgrade present Military Regulations, Specifications and Standards, many of which were written to support total package procurement and have not been subjected to needed revision. All of these activities must be supported, funded and managed by the command line in each DOD Component if improved readiness and effectiveness, with reduced life cycle cost, is to be attained.

OFFICE OF THE SECRETARY OF DEFENSE

The Acquisition Executive has prime responsibility for DODD 5000.x and for its implementation. The Assistant Secretary of Defense (Manpower, Reserve Affairs and Logistics) and the Director (Test and Evaluation) will participate in review of implementing documents.

CONCLUSION

Like the weather, everybody talks about reliability and maintainability problems but little seems to get done about them. Like the weather, no possible solution is going to please everybody. The farmers pray for rain while the golfers pray for sunshine. Nevertheless, the magnitude of reliability and maintainabiiity problems in deployed systems and equipment, and the ever-increasing squeeze on both acquisition and ownership cost, indicate the time has come to suspend debate and move out. While it is a long way from perfect and a long way from comprehensive, DODD 5000.x is a major step in the direction of reliable, maintainable systems and equipment at lower life cycle cost. On that basis, and on that basis only, the present and future revisions of DODD 5000.x should be reviewed and evaluated.

REFERENCES

1. Executive Office of the President, Office of Management and Budget, "Major System Acquisition," Cir A-109, 5 Apr 76.
2. Dept of Defense, "Major Systems Acquisitions," Directive, Number 5000.1, 18 Jan 77.
3. ________, "Major System Acquisition Process," Directive, Number 5000.2, 18 Jan 77.
4. ________, "Test and Evaluation," Directive, Number 5000.3, 19 Jan 73.
5. ________, "OSD Cost Analysis Improvement Group," Directive, Number 5000.4, 13 Jun 73.
6. ________, "Report of the Task Force on Specifications and Standards," Defense Science Board, Washington, DC, 7 Mar 77.
7. ________, MIL-STD-721B, "Definitions of Effectiveness Terms for Reliability, Maintainability, Human Factors and Safety," Washington, DC, 25 Aug 66.
8. ________, MIL-STD-781B, "Reliability Tests, Exponential Distribution," Washington, DC, 15 Nov 67.
9. ________, MIL-STD-781C, "Reliability Qualification and Production Acceptance Tests, Exponential Distribution," Washington, DC, 21 Oct 77.

Colonel Ben H. Swett, United States Air Force, is the Assistant for Reliability and Maintainability, Office of the Undersecretary of Defense Research and Engineering. The immediate prior assignment of Colonel Swett was that of Military Assistant to the Assistant Director for Planning, Office of the Deputy Director Research and Engineering. Earlier assignments include service as: navigator, avionics maintenance supervisor, scientist-specialist, development engineer and Research and Development Director.

Colonel Swett attended the Air Force Squadron Officer's School, the Army Special Forces School, and the Armed Forces Staff College. Colonel Swett holds an MA in Political Science. He has performed postgraduate work in the Philosophy of Science, R&D Management and Industrial Management.

B-1 STRUCTURAL TEST PROGRAM

by

John W. Rustenburg, Department of Air Force
Mark A. Owen, Department of Air Force
William L. Geese, Department of Air Force

The B-1 Bomber Program began in 1970 during a time when "fly-before-buy" was driving aircraft system procurement into long periods of development prior to production commitment.

Theoretically, subsystems may be replaced many times during an aircraft's useful life, **but basic structure is expected to continue to function as designed. Careful attention was therefore given in the B-1 Program to disciplined detail structural design and to extensive ground and flight testing.**

The B-1 full scale structural test program contained many separate elements that may be considered under the broad headings of static, fatigue and flight testing. This separation into testing regimes was not intended to imply independence. The results from one regime may be ***critical inputs to another. The thrust of the testing was toward finding and eliminating potential*** **structural problems in one phase of testing before moving on to the next. The intent was to arrive at an airframe whose structural integrity is fully understood and one that can be built in production quantities with high confidence.**

STRUCTURAL INTEGRITY

The results from testing are used to optimize system design as well as to aid in the production decision making process.

One aspect of B-1 development receiving major attention was structural integrity. The F-111 and C-5A recovery programs had demonstrated the need for a unified approach to structural integrity during the development phase. Aircraft structural considerations are of basic importance since the structure forms a foundation for the remainder of the weapon system.

Figure 1 is a schematic representative of the various elements of the B-1 structural test program.

STATIC TEST PROGRAM

The purpose of static testing on the B-1 was to verify analytically predicted margins of safety to clear the way for later flight tests. (As used here, "static" refers to discrete properties of as-designed structural assemblies, including strength, stiffness and natural frequency.) The elements of the B-1 static test program included design verification and ground vibration testing prior to the first flight of Aircraft 1 and a structural ground test program prior to flight of Aircraft 2.

In the past an airplane's static capabilities could only be evaluated by loading the full airframe to design load levels plus a factor of safety, and very often to destruction. Over the years airframe static design practice has advanced to the use of sophisticated finite element computer programs. As finite element analysis techniques developed, math modeling improved. As the accuracy of these methods was demonstrated, static testing innovations naturally occurred. If complex pieces of structure

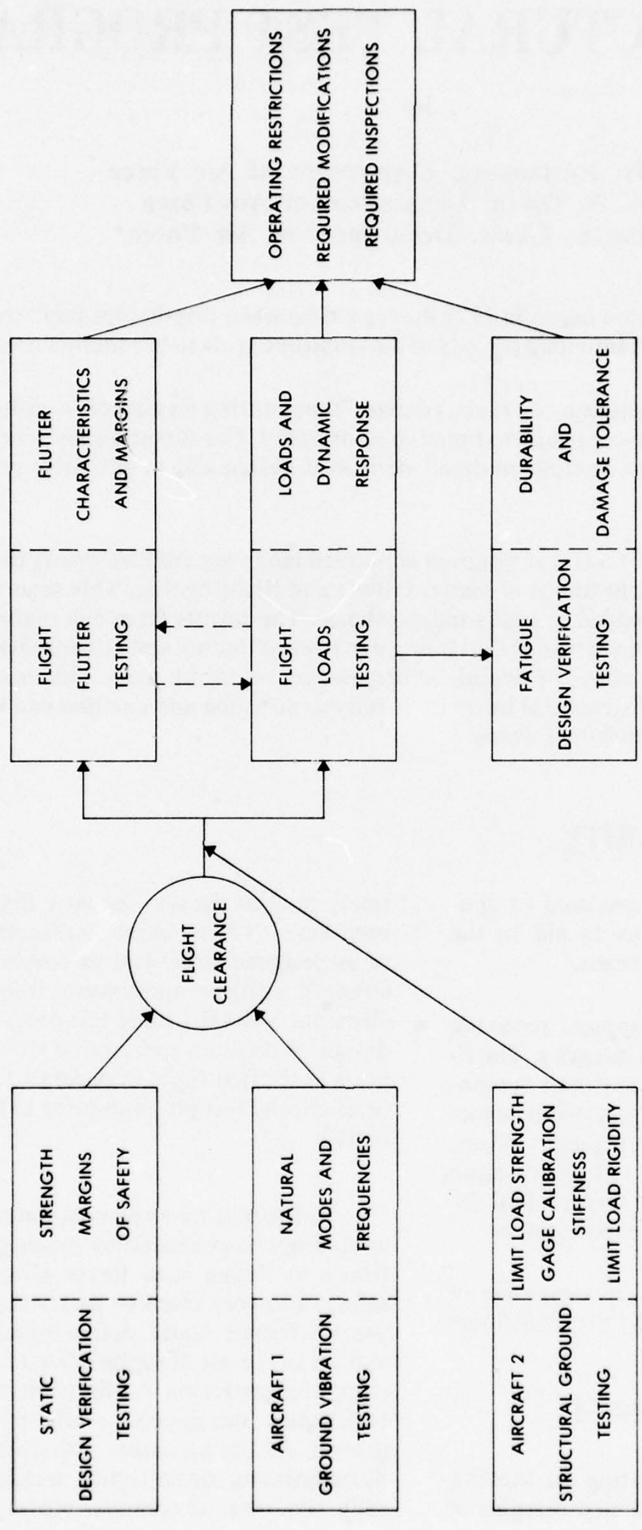


Figure 1. B-1 Full Scale Structural Testing

can be designed and analyzed using these finite element techniques, and then be verified by tests, a more simple structure designed and analyzed using the same methods will generally be adequate. The expense of test verification of the simpler structure therefore need not be borne.

In the case of the B-1, original plans called for a full airplane test to destruction in addition to extensive structural configuration development tests, design verification tests and preflight operational and proof loads tests. Because of the expense of this program, an innovative look was taken at the testing requirements to determine the design, analysis and testing absolutely necessary and realistic. As a result the full airplane test to destruction was eliminated. Significant program dollars were saved with almost no additional risk to the B-1 structural integrity. This decision allowed the earliest possible testing of the most critical structural assemblies in the design verification test program.

Design Verification Testing

The purpose of static design verification testing was to validate the predicted structural strength margins of safety prior to flight of any B-1 aircraft. The test articles consisted of components and assemblies representing the most critical areas of the airframe. The articles included all splices, joints and fittings appropriate to the specific test item with realistic boundary condition restraints. Selection of the test components was based on:

- Importance of the component's structural stiffness, integrity, and function in the vehicle,

- Degree of structural redundancy and complexity, and

- Method of construction and materiels used in the component as compared to current technology.

Figure 2 shows the location of these design verification test/articles within the B-1 airframe.

The static design verification test program on the major assemblies required 24 months to complete. The program met the requirement of verifying strength margins prior to flight of Aircraft 1. In the course of testing, design loads were altered and updated, substructures that had failed were redesigned and replaced, test conditions were rerun and various other changes were made to the program. Finally, the major wing carry through article was subject to twelve different design load conditions (six to limit load, the maximum load that any B-1 would be expected to experience in service, and six to ultimate, 150 percent of limit) in order to thoroughly test all portions of the structures. Catastrophic failure was finally induced at 111 percent of the wing design ultimate load.

A number of structural failures occurred during testing that contributed substantially to the time required to complete the program. These were failures for which remedies could be made on the flight articles. Almost all failures occurred in various parts of the fuselage structure as a result of one or both of the following:

- Poor detail design in corner attachment locations and in similar areas of abrupt load changes resulting from discontinuous structure; such design deficiencies compounded an already severe load redistribution problem.

- The use of undersize (shear critical) attachments for joining heavily loaded members; such fasteners were unable to take advantage of the substantial strength of the basic structure being joined.

The airplane fix involved simply increasing the size and material strength of the fasteners and adding a few large attachments at the critical corner locations. This modification often resulted in doubling the allowable joint strength.

An exception to this "easy fix" solution was a bulkhead failure in the main landing gear wheel well region of the aft intermediate fuselage where an unexpected load redistribution occurred. Design measures involving various

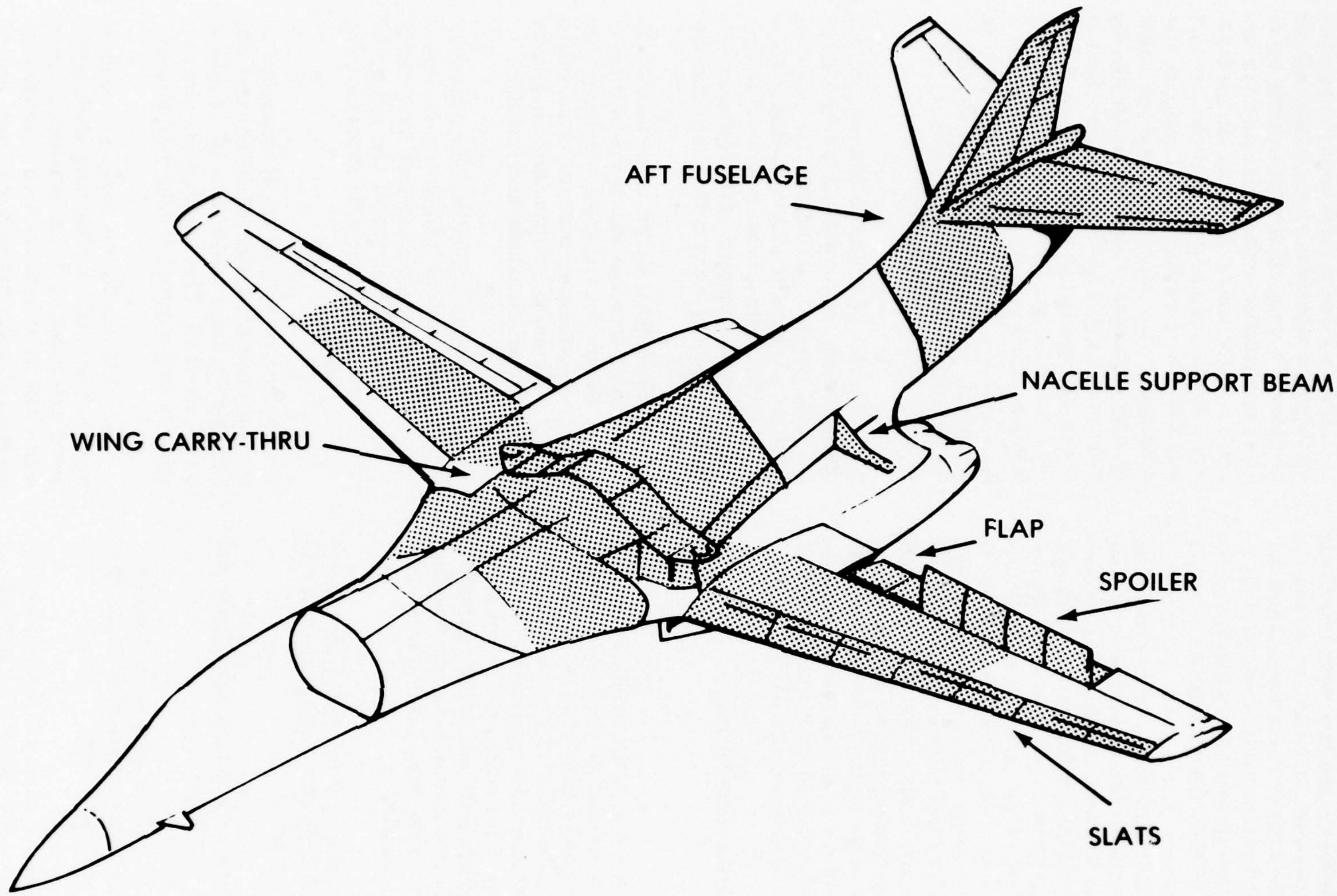


Figure 2. Design Verification Test

doublers, stiffeners and increased attachment strengths were tried with little success. The aluminum web of the bulkhead was finally replaced with a heavy steel member of exceptional structural integrity. A similar repair was used on the flight test aircraft and demonstrated on the design verification test article.

The static design verification test articles were a tool through which design strength could be checked, inadequacies uncovered and redesigns made. The tests further served as a vehicle for testing repairs before part installation in a flight article. When a generic problem was found with implications to structure not being tested (such as the undersize fasteners mentioned earlier), similar changes were made throughout the airframe. The result was high confidence in the structural strength of the flight articles as they entered a very demanding flight test program.

Ground Vibration Testing

The purpose of the ground vibration test is to determine the natural vibration modes and frequencies of the complete airplane. Ground vibration test constitutes a proof test of primary importance in determining structural flight limits with respect to flutter as well as other areas where vibration modes are important, such as the flight control surfaces, nose boom, overwing fairings and the landing gear.

The ground vibration test was conducted on Aircraft 1 prior to first flight over a nonstop, 12 and 1/2 day period. The entire airplane was mounted on a soft suspension system with attached vibration shakers and tested in six configurations including four different wing sweep angles and simulated single and double hydraulic failure modes of the horizontal tail. Symmetric and antisymmetric frequency sweeps were made for each of the six configurations resulting in a total of twelve separate test sequences.

The vibration modes obtained in the ground vibration test were compared with analytically predicted modes to assess the stiffness of the actual flight article. Also the vibration modes were used in calculations to determine "corrected" flight limits and enable further comparison with similar measurements on wind tunnel flutter models to establish the airplane flutter limits. This information provided the necessary confidence that the airplane would exhibit adequate flutter speed margins at the design flight limits and that flight flutter testing on Aircraft 1 would be initiated safely.

Aircraft 2 Structural Ground Testing

The structural ground test program on Aircraft 2 included the limit load proof test intended to demonstrate, in conjunction with the design verification tests, the static strength structural integrity of the aggregate airframe. Although this was the primary purpose, three additional objectives were satisfied, a structural deflection test, a limit load rigidity test and a strain gage calibration test. Total testing extended over a period of approximately 6 months.

Limit Load Proof Tests

For all the structural ground tests, Aircraft 2 was placed within a load reaction frame and mounted on main and nose landing gear fixtures. Nacelle attach point beam fixtures were used to transmit simulated engine and nacelle inertial loads. Weapon fitting loads were applied through simulated weapons launcher spools, that also reacted fuselage x-axis inertial loads. Wing, empennage and fuselage loadings were induced through tension, shear and point loading fixtures.

Fourteen limit load proof test conditions were chosen that represented selected aircraft ground handling, landing, braking, flight and weapon launcher conditions. Some test load conditions were truncated and in some cases two truncated conditions were combined and run as one condition.

A truncated load condition is one that loads only the vehicle section critical for a given condition. Balance loads are applied elsewhere on the vehicle for reaction purposes. In this way the total truncated condition is less complex than if a total airplane load condition were applied. Yet assurance of structural integrity for the critical test area is provided. Two truncated conditions could be balanced together in this same manner.

A large number of strain measuring devices had been installed during manufacture of the aircraft. Many of these (called "safety-of-flight" gages) were mounted on structural elements whose failure could result in serious damage to the aircraft. The location of these devices was identical to that of gages installed on Aircraft 1 and the design verification test components. Response from the gages during the application of test loads provided verification of internal structural load distributions. Too, a data interface between the design verification test and Aircraft 1 and Aircraft 2 flight tests was provided. During flight testing, the stress readings could be compared with established "never-exceed" limits.

As a demonstration of structural integrity at 100 percent of design limit load, the test reached a successful conclusion. Primary structural failures did not occur. Typical aircraft skin panel buckling did occur in a wide range of components, but after unloading, the buckles disappeared without leaving evidence of load induced permanent set. Further, the strain gage data generally followed the analytical predictions within reasonable bounds. Thus, the overall soundness of the structural analysis and design approach were successfully demonstrated.

Structural Deflection Tests

The B-1 development program originally included the requirement for structural influence coefficient testing on the complete airplane. The purpose of structural influence coefficient testing was to substantiate the stiffness characteristics that are used in the flutter analyses, the design of flutter models and the external loads analyses. A structural influence coefficient describes the deflection of the structure at a point on the airplane caused by loading applied at that and at other points. In a structural influence coefficient test the load is applied consecutively at a number of control points, while simultaneously measuring the resulting deflections at all those points. For the B-1 this would have resulted in 451 separate load applications and up to 143 deflection, twist angle or slope measurements at each load application.

Review of the aircraft industry's experience in performing structural deflection coefficient testing revealed that such testing is often performed on components such as pylons, actuators or control surfaces where detailed knowledge of the stiffness characteristics is important. Structural deflection coefficient testing over a complete aircraft is never performed and doubt was expressed that the accuracy of the results would be better than results obtainable by simpler means.

As a consequence, the requirement for structural influence coefficient tests was deleted and replaced with structural deflection tests. Structural deflections were measured at all the structural deflection coefficient control points during the application of two-limit load proof test conditions. The measured deflections could be compared with predicted deflections to obtain indirect evaluation of the influence coefficients. Results from ground vibration testing provided additional information used in this evaluation.

Limit Load Rigidity Tests

Limit load rigidity tests were required to determine if changes in stiffness occur from the application of limit load. These tests consist of deflection measurements similar to the structural deflection test, but at a considerably reduced number of control points. The deflection measurements help to determine the degree and character of any nonlinearity the structure might exhibit at limit load magnitudes. Deflection measurements for the limit load rigidity test were performed in conjunction with the limit load proof tests at all the applied test conditions. The results showed that a reduction in stiffness does not occur when the airframe is subjected to limit load.

Strain Gage Calibration Tests

The primary task assigned to Aircraft 2 included the structural airloads/dynamic response testing. The purpose of such testing is to measure in-flight and ground operational loads for substantiation of the structural design loads analyses. Too, structural integrity of the airplane for the critical loading conditions is flight demonstrated.

There are two basic methods for measuring structural loads—the strain gage method and the pressure survey method. Both methods provide the load parameters in terms of bending moment, shear and torsion and both were used on the B-1. Figure 3 shows the distribution of the structural airloads instrumentation.

The use of strain gages for measurement of bending moment, shear and torsion requires a calibration. This calibration is to provide equations in terms of mathematical combinations of strain gage outputs that will yield values of bending moment, shear and torsion. The most commonly used calibration technique requires the application of incremental loads at a number of individual loading points, one at a time. This point load method of calibration is not completely satisfactory since it calibrates the system for only a portion of the loads ultimately attained in actual flight test operations. A more ideal way of performing a calibration is through the application of a series of distributed loads representative of expected loading conditions. This method is not normally used because it requires a static test fixture and is more time consuming, complicated and costly.

The decision to submit a B-1 aircraft to a limit load proof test afforded an ideal method of calibration using distributed loads representative not only of actual flight conditions, but of predicted critical conditions as well. Calibration tests were run in conjunction with the limit load proof tests where load magnitudes and center of pressure locations were compatible. Additional special calibration conditions, not compatible with proof test conditions, were run separately to complete the calibration test. The loads equations derived were then available for use in the flight test program of Aircraft 2.

FLIGHT TEST PROGRAM

Perhaps the most visible portion of aircraft development is the flight test phase. The true test of an aircraft comes when it is called upon to fly and perform as designed. From a structural integrity standpoint, design performance has less visibility during flight test. The structural flight test program seeks to measure load distributions and approach, but not reach, aeroelastic instability points. Success is measured by a failure to find higher than predicted loadings, and the demonstration of adequate structural damping within the flight envelope.

Flight Flutter Testing

Flight flutter tests are performed to substantiate that the airplane is free from aeroelastic instabilities and has satisfactory damping up to limit speeds of the flight envelope. This is accomplished through the definition of modal parameters, principally frequency and damping, of the more important aeroelastic vibration modes of the aircraft in flight.

To accomplish the flight flutter tests, flutter critical substructures on the aircraft were forced into vibration and the resulting aeroelastic response was measured, recorded and analyzed. The mechanism for forcing the vibration used oscillating inertia shakers that were installed at the tips of the wings and at the tips of the horizontal and vertical stabilizers. Excitation of flight vibration modes was accomplished by sweeping the frequency range from 1 to 60 hertz or by selection of a discrete frequency. The magnitude of the shaker inertia forcing function could be varied.

The Aircraft 1 flight flutter test program covered practicable variations in parameters such as speed, altitude and gross weight. Of particular interest was the effect of operation of the B-1 automatic flight control (Stability and Control Augmentation) and terrain following ride quality (Structural Mode Control) systems. Flight flutter testing for flutter clearance of the flight envelope is performed in a "build-up" manner and must precede the structural flight loads testing. By the same token, Aircraft 1 is

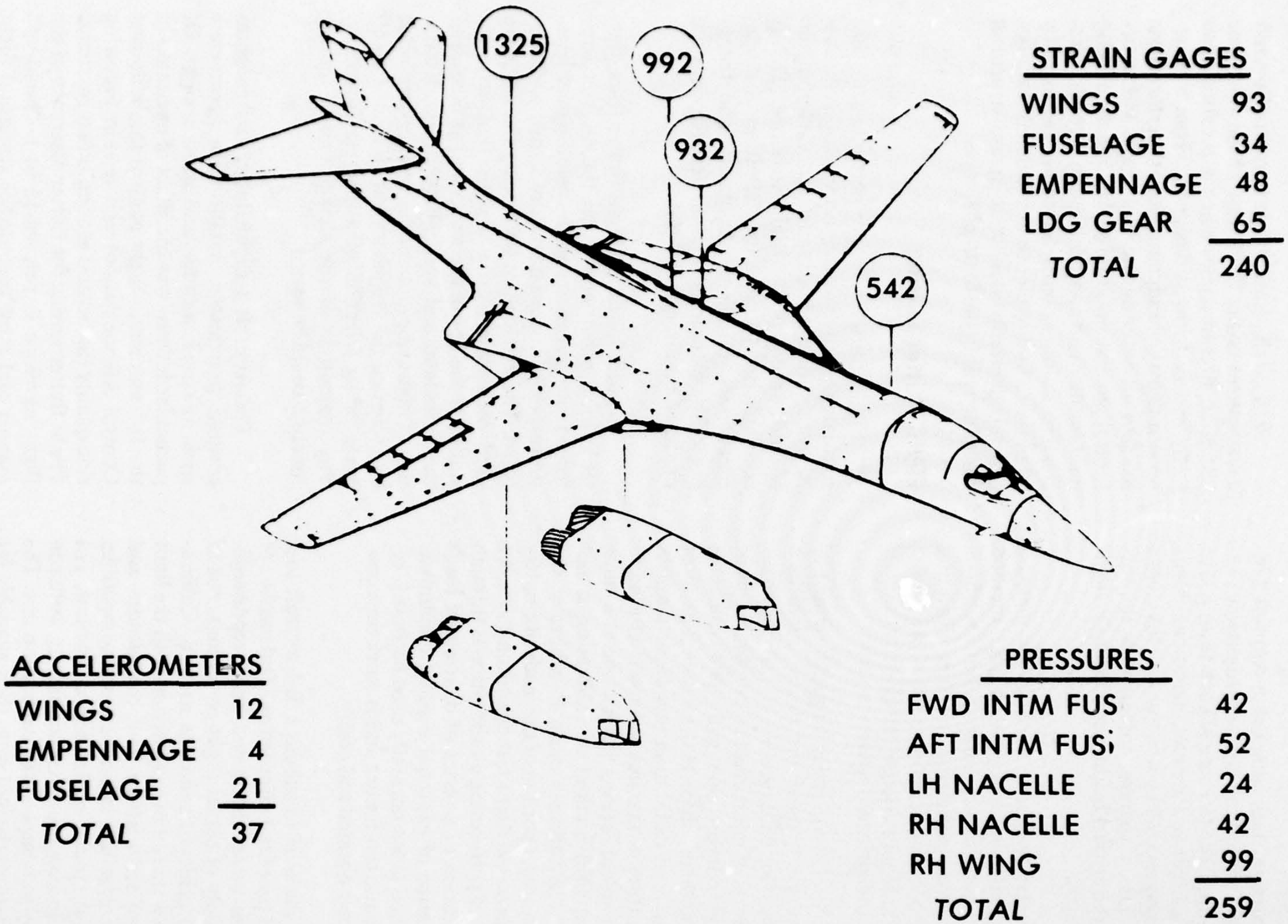


Figure 3. Structural Airload Instrumentation

limited to 80 percent of design limit load until completion of the flight loads testing to 100 percent of design limit and the incorporation of any necessary modifications resulting from either the flight or static test programs.

Flight flutter data were compared with theoretically predicted damping trends for the more important structural vibration modes. When practical these trends were used to demonstrate that, even with an increase of 15 percent in equivalent airspeed at all points on the limit speed envelope, flutter *would not* be expected.

Flight Loads Testing

Structural flight loads tests consist of flight and ground operation load surveys and dynamic response tests. The purpose is:

- to substantiate the structural loads analysis,
- to establish the variation of component loads with speed, altitude and type of maneuver for use in determining critical loading conditions and
- to demonstrate aircraft structural integrity for these critical conditions.

The flight and ground operations load survey is conducted in two distinct phases. The initial phase consists of a series of specified load survey maneuvers to 80 percent of design limit load. The test conditions cover the various operating regimes (gross weight, load factor, speed, altitude, etc.) so that the measured loads can be used to correlate with and substantiate the structural loads analysis. Upon completion of the initial phase, the measured loads data is extrapolated to 100 percent conditions for comparison with predicted design values and determination of critical maneuver conditions. The final phase then consists of build-up maneuvers to 100 percent of design limit load to demonstrate the conditions which were indicated as critical by analytical calculations, static tests and initial flight test data.

While the flight and ground operations test aircraft, Aircraft 2, proceeds to load levels equal to design limit, other test aircraft remain restricted to 80 percent load levels. The release of structural operating limitations for test and service inventory aircraft is contingent upon successful completion of the structural certification program, including static tests, flight tests and structural modifications resulting therefrom.

Dynamic response tests are performed concurrently with the loads survey and demonstration tests. These tests will include flight, landing and taxi conditions simulating actual operations to verify the dynamic loads analysis and response characteristics. Flights will be made through atmospheric turbulence for the conditions where significant dynamic response is anticipated. Landing tests will cover a range of sink speeds, and taxi tests will include runs over specific runway roughnesses at various speeds.

The load measurements obtained from Aircraft 2 are used to evaluate the adequacy of the loading conditions which were applied in the strength tests and, if necessary, to provide corrections to the test results or to any further ground tests.

In addition to strength, the measured loads data is particularly relevant to the estimation of structural fatigue life. Flight measurements are used to check the load distributions and the frequency of occurrence of load magnitudes for application to structural fatigue tests. Lessons learned during flight testing must be taken into account in the fatigue test program if structural integrity is to be assured throughout the aircraft's service life.

At this writing, the initial phase of the flight load survey is approximately 80 percent complete. Results of the testing accomplished have shown the measured loads to be equal to or less than predicted for the primary structural components.

FATIGUE TEST PROGRAM

The structural life discipline has made more rapid changes in recent years than has any area of structures technology. In 1970 when the B-1 contract was awarded, Air Force structural requirements included a "safe-life" fatigue design using a scatter factor of four.[1] This factor was assumed to account for the effects of initial structural quality, environment and variation in material properties. Fracture mechanics requirements had not been defined and fatigue testing could be done on a production aircraft using a "block" loading spectrum. In 1974–75 the Air Force requirements were changed to the current version.[2] Fracture mechanics design and testing were made requirements along with fatigue testing prior to production commitment using a representative flight-by-flight loading spectrum.

The B-1 program occurred in the middle of this evolution in structural philosophy and helped to pave the way for the present day requirements. The B-1 was the first Air Force program in which fracture mechanics was instituted as a design requirement and fracture control was made part of the development process.[3] Also, the B-1 program was the first program to employ major fatigue design verification test articles, the successful testing of which was a prerequisite to DSARC III.

The requirement to achieve early fatigue test results carried with it some consequent penalties. The B-1 DSARC III was scheduled for November 1976. The schedule required fatigue testing to commence by February 1976 on the major fatigue design verification test articles, or completion of manufacturing by the end of 1975. Design release of hardware drawings was needed approximately 2 years ahead of this date to meet material procurement and manufacturing schedules. Thus, the design of the fatigue design verification test articles had to be completed in the 1973 time period, *before any B-1 had flown and when there were still many unknowns about the fatigue loading spectrum.*

The B-1 structural design loads were released in late 1972 and early 1973. They were derived for a B-1 which was still very much on paper. Preliminary weights and inertias had been iterated to an extent, structural influence coefficients were based on general internal sizings and some wind tunnel data was available. However, structural dynamics had not been included and only a preliminary description of the B-1 stability and control augmentation system had been made.

As the time approached to begin design verification testing, more information had been obtained. Revised mass distributions and inertias had been calculated, along with updated influence coefficients. Structural dynamic effects were available for the vertical and lateral axes. Aerodynamic definition had been refined such that vertical and horizontal tail interference effects were predicted. As a result of information gained from the flight control simulator, the importance of the stability and control augmentation system to the B-1 fatigue loading spectrum was becoming clear. Finally, weight growth over specification goals had taken place that needed to be offset by changes in operational gross weights.

The program was faced with the choice of updating the design of the test articles to reflect these latest inputs or testing the original design to higher than design loads. In view of the disparities known to exist in fatigue analysis and design techniques, the decision was made to proceed with testing of the as-designed articles using a fatigue loading spectrum derived with the latest loads information. In this way the test itself would provide information about areas needing redesign rather than relying solely on analytical predictions for final structural sizings.

The decision to update the structural loads for test purposes was proved sound when the B-1 entered flight test. Measured load levels and frequency of occurrence of load magnitudes showed excellent correlation with the revised loading used during design verification testing. That revised loading spectrum was nevertheless more severe than the one to which the structure had been designed. This increased the risk of structural cracking and failures in the test articles, but a fully representative test spectrum was deemed important enough to offset risks to the test schedule, especially in view of the criticality of the structure being tested.

The advantage of design verification testing is that large scale component assemblies can be built quicker and cheaper than can an entire airframe thus allowing test results to be obtained earlier. However, some rationale must be available to select which sections of the aircraft will be tested in design verification test and which will wait for a later time. In the case of the B-1 determination of critical structure for the fatigue design verification test was made on the basis of the following:

- Stress levels (life assessment)
- Flight impact if failure occurs (loss of aircraft, aborted mission, complete mission)
- Complexity (load paths, advanced state of the art)
- Cost impact to repair (patch vs reskin)
- Historical problem areas (F-111, C-5, etc.)

The design verification tests were scheduled so results obtained could be incorporated into the production design and avoid expensive structural retrofit.

In view of the above, none of the cracking that finally occurred during the course of testing came as any surprise; that was the reason for the tests. In every case where cracking occurred, design improvements were incorporated into Aircraft 4, the test article was repaired and testing continued.

Fatigue testing on the B-1 was made up of two phases—durability testing and damage tolerance testing. In the first phase, which ran for two fatigue lifetimes, cracks developed naturally during the course of testing, identifying structural hot spots. These cracks were for the most part of the nuisance variety for which minor detail changes were generally sufficient. In the damage tolerance phase that followed, small artificial flaws were installed at various locations in critical primary structural areas. The rate of growth of these cracks under fatigue loading could then be used to assess the remaining structural integrity in the presence of undetected cracks of the type that might occur in actual service. This remaining structural integrity is called damage tolerance.

The most ironic observation to be made about the results of the fatigue design verification tests deals with the disparities in fatigue analysis methods. The structural areas which were of most concern from a fatigue life standpoint (e.g., wing lower surface, wing pivot lugs and fuselage longeron joints) exhibited almost no cracking problems. The areas that cracked were largely not safety critical parts and included shear webs, fuselage skins, flanges and fittings. Little attention is devoted to these types of parts in fatigue literature and so testing, rather than analysis, must be used to identify them as potential problem areas. This type of cracking, while not a safety problem, might have posed a maintenance and inspection burden in service had testing not identified this cracking for redesign. Therefore, in addition to the primary purposes of eliminating safety of flight problems and structural recovery programs, the fatigue test program may bring about a reduction in cumulative structural inspection and repair costs.

CONCLUDING REMARKS

This brief discussion of the B-1 full scale structural test program has shown how the static, fatigue and flight test elements are interrelated and focus on a common goal of assuring the structural integrity throughout the useful life of the aircraft.

As part of the B-1 structural development many other tests were performed that have not been discussed. More than 3000 material allowables coupons, 1700 fracture mechanics allowables elements and approximately 700 design development test components were run to support the design development process. More than 3000 hours of flutter model testing to substantiate flutter analyses and flutter margins, and approximately 5600 hours of wind tunnel testing to define aerodynamic force characteristics and pressure distributions were conducted. These small scale development tests contributed in large measure to the eventual successful progress of the full-scale structural testing program.

The primary lesson learned during the B-1 structural development was the need for continual interaction between the various engineering disciplines involved. Structural integrity is the result of interdisciplinary trade off and optimization. The structural testing elements as performed on the B-1 reflect this interrelationship and thus support the attainment of a quality engineering product to a stage where quality is left less to chance and more to the integration of disciplined design and testing practices.

REFERENCES

1. **Harold M. Wells, and Troy T. King, "Air Force Aircraft Structural Integrity Program: Airplane Requirements," ASD-TR-66-57, May 70.**

2. **Dept of Air Force, MIL-STD-1530A (11), "Aircraft Structural Integrity Program, Airplane Requirements," 11 Dec 75.**

3. **Mark A. Owen, "Evolution of Fracture Control in the B-1 Program," American Society for Metals—American Society of Nondestructive Testing (ASM-ASNT) Fifth Annual Forum: Prevention of Failure in the Transportation Industry, Tarpon Springs, FL, Jun 1977.**

Mr. John W. Rustenburg is an Aerospace Engineer in the Systems Project Office of the Aeronautical Systems Division, Wright Patterson Air Force Base, Ohio. Mr. Rustenburg has been with the B-1 Program throughout its development phase and has authored a number of technical reports and papers. Prior experience includes the XB-70 and the C-5A Programs.

Mr. Rustenburg received a BS (Aerospace Engineering) from West Virginia University in 1959.

Mr. Mark A. Owen is an Aerospace Engineer in the Directorate of Flight Systems Engineering, Aeronautical Systems Division, Wright Patterson Air Force Base, Ohio. Mr. Owen has been with the B-1 Program throughout its development phase and has authored a number of technical reports and papers. Prior experience includes the XB-70 and the C-5A Programs.

Mr. Owen received a BS (Aerospace Engineering) from Purdue in 1971. He received his MS (Aerospace Engineering) from the Ohio State University in 1977.

Mr. William E. Geese is an Aerospace Engineer in the Strategic Systems Project Office of the Aeronautical Systems Division, Wright Patterson Air Force Base, Ohio. Mr. Geese has been with the B-1 Program throughout its development phase and has authored a number of technical reports and papers. Prior experience includes the XB-70 and C-5A Programs.

Mr. Geese received a BS (Aerospace Engineering) from the University of Cincinnati in 1969.

PRODUCT LIABILITY

by

Lt Col Robert B. Machen, United States Army

The basis for this thought provoking article by Lieutenant Colonel Robert B. Machen was a short article written by Mr. Timothy Kellerher.* The material has been amplified and adapted by Lieutenant Colonel Machen to stimulate the thoughts of *Review* readers.

Beyond the introduction Lieutenant Colonel Machen provides a discussion that penetrates the present confusion surrounding reliability, responsibility and redress—three R's that have a direct relation to the Government's own test and evaluation of military equipment.

Sharp insights are shared with the reader who will gain an enhanced awareness of the implication of products defects. The policy maker, the systems acquisition expert and the Program Manager alike will find this discussion of product liability informative, stimulating and most importantly, beneficial.

"The Court of the Town of Williamsburg, County of York, is now in session." Let's imagine the town crier of 1782 giving notice of the case of Jonathan Thatcher vs Thomas Wyth, gunsmith. Jonathan's militiaman son, Isaac, lost a hand in an accident with a musket and Jonathan is suing for the loss of the son's hand, the pain and suffering endured, disfigurement, and loss of future employment in the militia. Thomas admits that he made and sold the musket in 1759 and argues that the gun had changed hands many times before it was acquired by Isaac as an issue item of the Commonwealth. The Commonwealth admits that the stock had been replaced and the sights adjusted several times by unknown users while the musket was out of the control of the gunsmith for more than 20 years from the time of manufacture. Thomas also argued that he had made all his muskets for squirrel hunting and had not made instruments of war, but the lawyer for Isaac insisted that the basic design was wrong and that the hammer should not have been sharp on one side. If the gunsmith had simply rounded both sides of the hammer, the hammer could not have cut the thumb of Jonathan's left-handed son and caused an infection, which led to loss of the hand. The lawyer was very persuasive in his argument that the gunsmith should have known that the equipment would be used in an environment where infection was likely, and arguably, following a certain tea party and in retrospect, it was also foreseeable that the colonists could rebel against the Crown before the musket would wear out. The lawyer showed that the chain of events were crystal clear to those who would only look.

In vain did Thomas protest that Thatcher had already won his negligence suit against Dr. O'Neal, who failed to save the hand. The results of that suit had driven the physician to bankruptcy. The law was straightforward. The manufacturer was responsible for his product. The passage of time did not make a difference. The gunsmith had to pay for his poor design.

Consider the more recent history of product liability in the United States starting with the landmark decision in MacPherson v. Buick Motor Company. Justice Cardoza's decision handed down in New York in 1916 was one of

*Timothy Kellerher, "Hear Ye, Hear Ye," Alexandrian, Vol 3 (11): 23 (Nov 77).

the first major steps away from what is called privity of contract; that is, the governing relationship between two people in contracts excluding any interest of a third person regardless of impact. Cardoza's turn away from the protection afforded by privity toward strict liability in the interest of third parties opened an entirely new field of law in cases against manufacturers. Simply stated, MacPherson bought a new car from the Buick Motor Company who manufactured automobiles. The car was sold to a retailer who sold it to MacPherson. While MacPherson was driving the car, a wheel collapsed and he was thrown out and injured. One of the spokes was made of defective wood. Although Buick had bought the wheel from another manufacturer, it was shown that the defects could have been discovered by simple inspection. The holding that the manufacturer was responsible for the finished product, even though the product was processed through a retailer to a customer, is known to almost everyone who studies law. What is not commonly known, however, is that there has been a virtual explosion of court decisions during the seventies favoring the user. It stands to reason that the surface has just been scratched, for the National Safety Council estimated that in 1975 over thirty million injuries were related to products defects. The explosion of litigation can, in part, be attributed to the nonapplication of old standby defenses, such as contributory negligence and assumption of risk. Both, along with the defense of privity, have been cast aside by recent court decisions.

The absurdity of the situation can be shown by further paraphrasing of the article referenced. The facts reveal the following in a recent suit against "Old Saw Manufacturing Co." In 1942, some 35 years ago, the company made a bench saw with the customary safety guards. The saw was delivered under wartime contract and met all of the specifications applied by the government to such purchases. Sometime after World War II, when it was decided that the saw was worn and surplus, it was sold. The saw then changed hands a number of times. It was worn out, rebuilt and, after more than 30 years, all of the saw that was recognizable was the original nameplate "Old Saw Manufacturing Company." In its last resting place, the saw was set up without the safety guards of a bench saw and used by personnel who apparently knew little of the tradeoffs between convenience of work and the danger of an exposed blade.

A man was hurt and went for relief. The attorney knew that the small operation which was using a 30 year-old saw had no money, so he looked up the manufacturer. That's what a good lawyer is paid for in our present culture. Sure enough the Dunn & Bradstreet rating was good, so "Old Saw Manufacturing Company" was sued! Any lawyer that smart had to be successful. He was able to obtain a jury trial and the "Old Saw Manufacturing Company" paid through the nose for a verdict on the design of the old saw.

The liability of manufacturers is approaching that of the physician as product liability insurance costs increase drastically. The costs of product liability claims rose more than seven times in the last 5 years. One manufacturer who lost a suit had his liability insurance premiums increase from below $5,000 in 1971 to almost $500,000 in 1977.

The full impact of the drastic changes of the past few years has not become apparent to the manufacturers of defense products. Only time will tell whether or not the first hypothetical story becomes a reality. One only has to look at a weekly summary of military aircraft accident reports or other similar summaries to realize that products fail daily. The next question to be asked is, When will the parents of the young soldiers, sailors, and airmen realize the significance of product liability? Will they soon discover that because it's painted Army Green, Battleship Gray, or Air Force Blue, it makes no difference?

How does the foregoing affect the policy makers of the Armed Forces, the Systems Acquisition Managers and the Program Managers of defense systems? Perhaps the most telling impact is recognition of the significance of manufacturer responsibility and how legal determination of product liability is established.

The program Manager who nurtures a major program from inception to completion is indeed a fortunate individual if the task is accomplished without the disruptive influences of at least one near (or actual) catastrophic occurrence—an occurrence that is attributable to defective equipment such as a component or part.

When a product malfunction leads to adverse results (injury or death to product users) and a program goes awry, think on these things (all elements of a cause of action):

- Was there an equipment (product) defect?

- Was the defect present when the product first was delivered?

- Was the defect the proximate cause of the injury?

- Was the plaintiff someone that the manufacturer could reasonably *foresee* using the product for which it was manufactured and sold?

In the event that either or all of the above questions are positive, the next step will be to determine:

1. Is there a cause of action based on contract? Such a cause of action requires that there be a contractual relationship between the plaintiff and defendant.

2. Is there an action in tort which is grounded upon negligence? Such a cause of action does not require the allegation of a contractual relationship between the plaintiff and defendant.

3. Is there an action in tort which is based on a breach of duty assumed by the manufacturer or seller of a product when the manufacturer expressed or implied a fitness for use when the product was delivered?

Again if either one or a combination of these questions are affirmative you can rest assured that you are in for a real education—the same type of education that has haunted the medical profession for the past few years. As you work toward satisfying test and evaluation requirements consider the emphasis you have (or have not) placed on the quality of the products that comprise the system of your program.

A reasonable conclusion is that manufacturers will pay for injuries caused by defects in their products. The day has arrived when one can safely say "caveat vendor" instead of "caveat emptor."

Lieutenant Colonel Robert B. Machen, is the Acting Director of the Department of Research and Publications, Defense Systems Management College (DSMC). He has served in positions of responsibility at platoon, company, battalion, group, brigade, division, and major command levels and recently completed a tour at Headquarters, Department of the Army before being assigned to DSMC.

Lieutenant Colonel Machen was graduated from Northeast Louisiana University with a Bachelor's Degree (Geology), from the College of William and Mary in Virginia with a Master's Degree (Counseling), and from American University with a Doctorate (Educational Psychology and Administration). Lieutenant Colonel Machen is a Master Army Aviator and Examiner, with more than 4500 hours of flight experience. A graduate of the US Army Command and General Staff College, Fort Leavenworth, KA, (1971), Lieutenant Colonel Machen is now enrolled as an evening student at the International School of Law, Washington, DC.

DEFENSE SYSTEMS MANAGEMENT REVIEW

The Defense Systems Management Review is published quarterly by the Defense Systems Management College, Fort Belvoir, VA 22060. Publication of the Review was approved by OASD(PA) May 18, 1976.

The Review is available in microfiche or paper copy from the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22151 and the Defense Documentation Center (DDC), Cameron Station, Alexandria, VA 22314. When ordering from DDC specify the volume and issue number of the DSM Review desired and the date of the issue.

DEFENSE SYSTEMS MANAGEMENT REVIEW

INDEX OF AUTHORS
(VOL I, Issues 1,2,3,4, and 5)

VOL I, NO 1

Mr. James A. Francis
Dept of Navy

THE CONGRESSIONAL BUDGET AND IMPOUNDMENT CONTROL ACT OF 1974: IMPLICATIONS TO DOD

PICTURE NOT AVAILABLE

Capt Robert F. Ewart,
Coauthor, US Air Force

WHAT HAPPENED TO PERT?

Mr. Carroll Garrison
Dept of Army

TECHNOLOGICAL PROGRESS AND LIFE CYCLE SUPPORT

Capt Donald M. Nanney,
Coauthor, US Air Force

WHAT HAPPENED TO PERT?

Lt Col Robert H. Ammerman, Jr.
US Army

TRAINING WITH INDUSTRY

Mr. Wayne L. Hinthorn
United Technologies Corp.

INDUSTRY MANAGEMENT OF DEFENSE VS COMMERCIAL PROGRAMS

Maj Edward J. Dunne, Jr.,
Coauthor, US Air Force

WHAT HAPPENED TO PERT?

VOL I, NO 2

Mr. R. M. Lockerd
Texas Instruments, Inc.

ELECTRONIC TECHNOLOGY PROGRESS AND LIFE CYCLE SUPPORT

INDEX OF AUTHORS, Continued

Lt Col A. J. Driscoll
US Air Force

SOFTWARE VISIBILITY FOR THE PROGRAM MANAGER

Honorable Walter B. LaBerge
Undersecretary of the Army
(formerly Assistant Secretary General for Defense Support, NATO)

A CONCEPT OF A TWO-WAY STREET

Mr. D. D. Acker
Defense Systems Management College

CONTEMPORARY MANAGEMENT-PROFESSIONAL ASPECTS

Brig Gen Frank P. Ragano
US Army Program Manager of the Year 1977

US ROLAND-A GIANT STEP TOWARD WEAPON COMMONALITY

Maj J. W. Allsbrook
US Air Force

ROLE OF CONGRESSIONAL STAFF IN WEAPON SYSTEMS ACQUISITION

Mr. John H. Richardson
Executive Vice President
Hughes Aircraft Company

ROLAND-A TECHNOLOGY *TRANSFER PROGRAM*

Mr. W. W. Thybony
Office of Federal Procurement Policy

WHAT'S HAPPENED TO BASICS?

Brig Gen James A. Abrahamson
US Air Force

THE F-16, NATO'S MILITARY AND ECONOMIC CORNERSTONE

VOL I, NO 3

Mr. William E. Stoney
Scientific Advisor to the Supreme Allied Commander, Europe
(formerly with OSD)

THE PROCESS OF STANDARDIZATION-AN OVERVIEW

Col D. W. Waddell
US Air Force

COMMONALITY-OR, WHAT'S IN A WORD?

INDEX OF AUTHORS, Continued

MAJ John D. Elliott
US Army, Designated a
NATO Fellow, 1977

INTERDEPENDENCE-THE
IMPACT ON US SECURITY

Mr. Eugene Lurcott
RCA

F_2D_2, A SYSTEM
MANAGEMENT TOOL

Lt Col A. Martin Lidy
US Army

NATO STANDARDIZATION-
AN ALTERNATIVE
APPROACH

Dr. Joseph F. Shea
Raytheon Company

OBSERVATIONS ON DEFENSE
ACQUISITION

VOL I, NO 4

Dr. Albert J. Kelley
President, Arthur D. Little
Program Systems
Management Co.

CAN WEAPON SYSTEM
PROCUREMENT BE
MANAGED?

Lt Col Paul B. Demitriades
US Air Force Reserve

ZERO-BASE BUDGETING
AND SUNSET LEGISLATION

Mr. Jacques S. Gansler
VP, Analytic Sciences Corp.
(formerly Dep Asst SECDEF
(Materiel Acquisition))

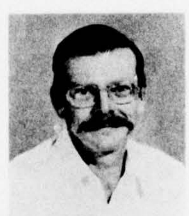
A NEW DIMENSION IN THE
ACQUISITION PROCESS

Mr. Leonard S. Freeman
Dept of Army

THE ARMY BUDGET AND
COMBAT CAPABILITY

Mr. William F. Brown
Program Coordinator

TAILORING PROGRAM
REQUIREMENTS



Mr. James A. Bowen,
Coauthor, Dept of Navy

ESTABLISHING THE FAE II

INDEX OF AUTHORS, Continued

Capt Ronald S. Fry,
Coauthor, US Air Force

ESTABLISHING THE FAE II

CDR Ian E. M. Donovan,
Coauthor, US Navy

OPERATIONAL TEST AND
EVALUATION OF SHIPS:
POLICY AND PRACTICE

Major Robert S. Feingold
US Air Force

COMPUTER SYSTEM
SIMULATION: A
DESIGN EVALUATION TOOL

LCDR Thomas A. Fitzgibbons,
Coauthor, US Navy

OPERATIONAL TEST AND
EVALUATION OF SHIPS:
POLICY AND PRACTICE

VOL I, NO 5

Lt Gen Walter E. Lotz, Jr.,
USA (Ret)

TEST AND EVALUATION IN
THE DEPARTMENT OF
DEFENSE

CAPT T. H. Sherman
US Navy

TEST AND EVALUATION IN
DEFENSE PROCUREMENT

Maj Gen Patrick W. Powers
US Army

STREAMLINING ARMY
TESTING

Mr. George G. Gleghorn
Space Systems Division
TRW, Inc.

OPTIMIZING SPACE VEHICLE
TEST PROGRAMS

Maj Gen Howard W. Leaf
US Air Force

IMPROVING AIR FORCE
INDEPENDENT
OPERATIONAL TESTING

Mr. John W. McCord
Ofc of the Undersecretary
of Defense

MANAGEMENT OF MAJOR
DOD TEST FACILITIES

INDEX OF AUTHORS, Continued

Lt Col David L. Washington
US Air Force

TEST AND EVALUATION OF TRI-TAC EQUIPMENT

Mr. Howard W. Kreiner
System Planning Corp.

HOW MUCH TESTING IS ENOUGH?

Mr. Charles W. Karns
Ofc of the Undersecretary of Defense

TEST AND EVALUATION POLICY

Lt Col Michael J. Butchko
US Air Force

AIRCRAFT OPERATIONAL TEST AND EVALUATION

Lt Col Martell D. Fritz
US Army

USE OF TRAINING EXERCISES FOR TEST AND EVALUATION

Mr. David T. Love
General Electric Co.

CHALLENGES IN AIRCRAFT ENGINE TEST AND EVALUATION PROGRAMS

Col Ben H. Swett
US Air Force

RELIABILITY AND MAINTAINABILITY IN THE ACQUISITION PROCESS

Mr. John W. Rustenburg
Coauthor, Dept of Air Force

B-1 STRUCTURAL TEST PROGRAM

Mr. Mark A. Owen
Coauthor, Dept of Air Force

B-1 STRUCTURAL TEST PROGRAM

Mr. William L. Geese
Coauthor, Dept of Air Force

B-1 STRUCTURAL TEST PROGRAM

Lt Col Robert B. Machen
US Army

PRODUCT LIABILITY

STAFF/CONTRIBUTORS

RADM R.G. FREEMAN III, USN
Commandant, DSMC
Jul 1977 –

MG JOHN G. ALBERT, USAF
Commandant, DSMC
Jul 1974 – Jul 1977

BG Frank J. Palermo, Jr., USA
OUSD R&E
Contributing Editor
VOL I, No. 3

Mr. David D. Acker, DSMC
Professor of Management
Contributing Editor
VOL I, No. 1,2,3, & 4

Mr. Donald R. Greenlee
Staff Assistant, OUSD R&E
Contributing Editor
VOL I, No. 5

PH2 Raymond S. Nunn, USN
Staff Photographer, DSMC

Mr. Richard Brown
Graphics Chief, DSMC

Nicholas W. Kunz, SP-5, USA
Graphics Designer, DSMC

Mrs. Jeanette Tippie
Editorial Assistant, DSMC

Ms Avonale "Stevie" Stephenson
Managing Editor, DSM Review

CALL FOR MANUSCRIPTS

Manuscripts will be considered for publication in the Defense Systems Management Review. The following topics are of particular interest to the Review readership.

- **Views of professionals on current and pertinent defense systems acquisition and program management**
- **Problems confronting Program and Systems Acquisition Managers**
- **Analysis of approaches to problem solution**
- **Past experiences of responsible authorities**
- **Defense systems management perspectives of the US Congress, the military services, industry, the media, *and* multinational programs.**

To share your knowledge and expertise contact the Managing Editor, Defense Systems Management Review, Defense Systems Management College, Fort Belvoir, VA 22060. Telephone: Area Code 703, 664–5082; AUTOVON 354–5082.

DEFENSE SYSTEMS MANAGEMENT REVIEW

DISTINGUISHED ASSOCIATE EDITORS

DEFENSE SYSTEMS MANAGEMENT COLLEGE

CORRECTIONS VOL I, NO. 4 — AUTUMN, 1977

Page vi, line 5.	Reads: Dr. Albert J. Kelley, President, Author D. Little Program Systems Management Company **Should read**: Dr. Albert J. Kelley, President, Arthur D. Little Program Systems Management Company
Page 51, column 1, line 6.	Reads: ...from the OMB-Base Operations budget; **Should read**: ...from the OMA-Base Operations budget;

NOTES

☆ U. S. GOVERNMENT PRINTING OFFICE : 1978 733 569/461