Report of the
Defense Science Board Task Force

on

TEST AND EVALUATION

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

September 1999

Office of the Under Secretary of Defense
For Acquisition and Technology
Washington, D.C. 20301-3140

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
September 1999

3. REPORT TYPE AND DATES COVERED
Final Technical, 1999

4. TITLE AND SUBTITLE
Report of the Defense Science Board Task Force on Test and Evaluation

5. FUNDING NUMBERS
n/a

6. AUTHOR(S)
Dave Heebner

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Defense Science Board, Ofc of the Under Secy of Def (A&T)
3140 Defense Pentagon, RM 3D865
Washington, DC 20301-3140

8. PERFORMING ORGANIZATION REPORT NUMBER
n/a

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Defense Science Board
3140 Defense Pentagon, RM 3D865
Washington, DC 20301-3140

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
N/A

11. SUPPLEMENTARY NOTES
N/A

12a. DISTRIBUTION AVAILABILITY STATEMENT
Distribution Statement A: Approved for Public Release: Distribution is unlimited

12b. DISTRIBUTION CODE
A

13. ABSTRACT *(Maximum 200 words)*

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
unclassified	unclassified	unclassified	N/A

TABLE OF CONTENTS

Transmittal letter to the Undersecretary of Defense (Acquisition & Technology) ... iii

Transmittal letter to the Chairman, Defense Science Board ... v

EXECUTIVE SUMMARY ... 1

INTRODUCTION ... 10

- Background ... 11
- The Test and Evaluation Process ... 12
- The Impact of T&E on Streamlining the Acquisition of Major Defense Programs ... 15

TEST AND EVALUATION AND THE REQUIREMENTS DETERMINATION PROCESS ... 17

- Tester Involvement in the Acquisition Process ... 17
  - How the "System" Should Work ... 19
  - Recommendations ... 20

DEVELOPMENTAL AND OPERATIONAL TEST PROCESS ... 21

- The Current DT/OT Process ... 21
  - Findings ... 22
  - Recommendations ... 23

MODELING AND SIMULATION IN SUPPORT OF THE TEST AND EVALUATION PROCESS ... 25

  - Findings ... 26
  - Recommendations ... 33

TEST AND EVALUATION FACILITY REENGINEERING ... 36

- Overall Facility Management ... 36
  - Findings ... 36
- Reengineering Individual T&E Facilities ... 37
  - Findings ... 38
  - Recommendations ... 39

SPECIAL ISSUES ... 42

- Advanced Concept Technology Demonstrations (ACTDs) ... 42
  - Recommendation ... 43
- Systems of Systems T&E ... 43
  - Findings ... 44
  - Recommendation ... 45
- Information Systems Software T&E ... 45
  - Recommendations ... 46
- Test and Evaluation of Commercial Off-the-Shelf Software (COTS) ... 46
  - Recommendation ... 47

SUMMARY OF RECOMMENDATIONS ... 48

- Test and Evaluation Requirements Determination Process ... 48
  - Recommendations ... 48

Developmental and Operational Test Process 48
Recommendations 48
Modeling and Simulation in Support of the Test and Evaluation Process 49
Recommendations 49
Test and Evaluation Facility Reengineering 50
Recommendations 50

ANNEX A: TERMS OF REFERENCE A-1

ANNEX B: TASK FORCE MEMBERSHIP B-1

ANNEX C: MAJOR RANGE & TEST FACILITY BASE (MRTFB) OPERATIONS FY99 ... C-1

ANNEX D: GLOSSARY D-1

OFFICE OF THE SECRETARY OF DEFENSE
3140 DEFENSE PENTAGON
WASHINGTON, DC 20301-3140

DEFENSE SCIENCE BOARD

MEMORANDUM FOR UNDERSECRETARY OF DEFENSE (ACQUISITION AND TECHNOLOGY)

SUBJECT: Final Report of the Defense Science Board Task Force on Test and Evaluation

I am pleased to forward the Final Report of the Defense Science Board Task Force on Test and Evaluation.

The Task Force examined facilities, processes, policies, contractor testing, government development testing, joint T&E, and operational T&E, both before and after full production and in the continuing evolution of force capability. The Terms of Reference directed that the Task Force undertake a broad review of all activities relating to T&E. Specifically, the Task Force was asked to:

- Examine new and innovative ways that the T&E community can better support its users;
- Find new ways to integrate operational testing into the overall system development process;
- Consider the special problems associated with T&E of the "systems of systems" which are increasingly comprising critical parts of our military capability;
- Identify and quantify the current and future needs of the Department's T&E capabilities and resources.

The Task Force made the following major recommendations:

DoD should establish a stable team made up of users, developers, testers and appropriate contractors called a Combined Acquisition Force (CAF) to streamline the acquisition process for ACAT 1 programs.

Each of the Service Developmental Test and Operational Test (DT&OT) organizations should be consolidated, to include integrated planning, use of models, simulation, and data reduction.

Establish Oversight and Direction of Modeling and Simulation (M&S) development and employment for T&E to provide needed investment in M&S for T&E, discipline to the Verification, Validation, and Accreditation (VV&A) process, and understanding of where M&S can and cannot contribute to T&E.

Implement the plan developed by OSD to create a central T&E Resource Management Organization to own, fund and operate all DoD T&E resources.

Ensure the participation of a tester in the ACTD process to maximize the value of this concept.

Create a systems of systems design authority (system designer) with a joint T&E component which uses the Service T&E resources for joint testing.

DoD should conduct functional and interface testing of Commercial Off-the-Shelf Software (COTS) software as a prerequisite to its use in DoD programs.

The Task Force proposed clear and concise recommendations that can be implemented. I concur with those recommendations and recommend you forward the study to the Secretary of Defense.

Craig I. Fields
Chairman

OFFICE OF THE SECRETARY OF DEFENSE
3140 DEFENSE PENTAGON
WASHINGTON, DC 20301-3140

DEFENSE SCIENCE BOARD

MEMORANDUM FOR THE CHAIRMAN DEFENSE SCIENCE BOARD

SUBJECT: Final Report of the Defense Science Board Task Force on Test and Evaluation

Attached is the report of the Defense Science Board Task Force on Test and Evaluation (T&E).

The Terms of Reference directed that the Task Force undertake a broad review of all activities relating to T&E. Specifically, the Task Force was asked to:

- Examine new and innovative ways that the T&E community can better support its users;
- Find new ways to integrate operational testing into the overall system development process;
- Consider the special problems associated with T&E of the "systems of systems" which are increasingly comprising critical parts of our military capability;
- Identify and quantify the current and future needs of the Department's T&E capabilities and resources.

To perform this review the Task Force examined facilities, processes, policies, contractor testing, government development testing, joint T&E, and operational T&E, both before and after full production and in the continuing evolution of force capability.

Based on our study of the T&E community, the Task Force made the following major recommendations. Additional supporting recommendations can be found in the body of the attached report.

DoD should establish a stable team made up of users, developers, testers and appropriate contractors called a Combined Acquisition Force (CAF) to streamline the acquisition process for ACAT 1 programs. The purview of the CAF need not be limited to T&E but should be oriented toward all aspects of improving the efficiency and effectiveness of the acquisition process.

Each of the Service Developmental Test and Operational Test (DT&OT) organizations should be consolidated, to include integrated planning, use of models, simulation, and data reduction. Planning should be totally integrated, and the OSD T&E

organizations consolidated. There should be integrated use of models, simulation, and data reduction.

Establish Oversight and Direction of Modeling and Simulation (M&S) development and employment for T&E to provide needed investment in M&S for T&E, discipline to the Verification, Validation, and Accreditation (VV&A) process, and understanding of where M&S can and cannot contribute to T&E.

Implement the plan developed by OSD to create a central T&E Resource Management Organization to own, fund and operate all DoD T&E resources. This organization should be responsible to the Secretary of Defense for planning, budgeting, operations and maintenance of the Department's T&E resources.

Ensure the participation of a tester in the ACTD process to maximize the value of this concept. The most useful role for the T&E community in the ACTD program is helping with transition to acquisition status.

Create a systems of systems design authority (system designer) with a joint T&E component which uses the Service T&E resources for joint testing. There is, at this time, no overall architect of the joint system although many efforts are being made to cope with the resulting problems.

DoD should conduct functional and interface testing of Commercial Off-the-Shelf Software (COTS) software as a prerequisite to its use in DoD programs. Military application may stress the COTS software in unusual ways, a problem that may not appear until the system has been completely assembled and tested under realistic conditions.

The Task Force believes that the changes recommended in the attached report can help T&E to become an integral and beneficial part of the acquisition process and will put in place processes to streamline T&E and make it increasingly cost effective.

The Task Force would like to express its appreciation for the extensive support provided by the O.S.D. staff.

We would also like to thank the members and government advisors of the Task Force for their very helpful contributions and advice.

David R. Heebner

David R. Heebner
Task Force Chair

EXECUTIVE SUMMARY

The Defense Science Board task force on Test and Evaluation was chartered to conduct a broad review of the entire range of activities relating to test and evaluation (T&E). This included examining new and innovative ways that the T&E community can better support its users and to find better ways to integrate operational test into the overall systems development process. The task force was asked to consider the special problem of "systems of systems" testing. In addition, the task force was to identify and quantify the current and future needs of the Department's T&E capability and resources, and recommend specific and quantified changes to those capabilities and resources.

To perform this review the task force examined facilities, processes, policies, contractor testing, government development testing, joint T&E, and operational T&E, both before and after full production and in the continuing evolution of force capability. The focus of the task force examination was on the T&E part of the acquisition process.

The task force formed sub-panels to focus on particular subjects related to T&E. These sub-panels studied requirements determination in the T&E process, the developmental testing and operational testing (DT/OT) process, the role of modeling and simulation in T&E, and facility reengineering. The task force also looked at special issues affecting T&E including T&E for Advanced Concept Technology Demonstrations (ACTDs), T&E of Commercial Off the Shelf (COTS) software, T&E of Software, and T&E of Systems of Systems.

Each of these sub-panels developed a set of findings and recommendations to improve the T&E process.

Findings

- **The focus of T&E should be on how to best support the acquisition process.** The focus of T&E should be on optimizing support to the development/acquisition process, not on minimizing (or even "optimizing") T&E capacity. T&E is an integral part of system design, development and acquisition. The two crucial aspects of T&E in supporting acquisition are testing to learn and testing to confirm. For large projects, as investments of funds and time grow, and more is at stake, the need for testing to confirm increases.
- **T&E planning with operational test personnel should start early in the acquisition cycle.** Testing must contribute earlier in the development process. The early involvement of testers and users, especially through the Service Operational Test Agencies is key to this optimization. Early testing to learn about new systems is equally important and requires early tester involvement. Early involvement of testers (and also users), who are independent of the development, provide the feedback essential for the design refinements that lead to truly excellent procurements.
- **Distrust remains between the development and test communities.** There has been reluctance to involve the test and evaluation community early by some program

offices hoping to maintain control of early test results. There has also been reluctance in some testing organizations to be involved early out of fear of losing their independence. In addition, the operational test organizations have not had the resources for early involvement. This has led to polarization within what should be an integrated venture. The resulting tensions have reduced the effectiveness of both T&E and acquisition by interfering with the potential contributions from early involvement. Recapturing the early involvement of testers, while preserving independence of evaluation is a critical objective.

- **Contractor Testing, Developmental Testing, and Operational Testing have some overlapping functions.** A key element in improved support to the users of test and evaluation is improved integration of operational testing into the overall system development process. The current system forces program managers to minimize the learning function of test and evaluation – in both developmental and operational contexts. The Integrated Product Teams that the department uses do not go far enough, and do not start early enough.
- **Independence of evaluation of test data is the essential element, not the taking of the data itself.** The concept of independence of test and evaluation is an important principle. However, is has been interpreted in such a way as to separate T&E from the development and acquisition process of which it is an integral part.
- **Response to perceived test "failures" is often inappropriate and counterproductive.** A test failure often uncovers important information that allows major steps of progress in system design but is too often accompanied by withholding funding, costly rescheduling and threats of cancellation. Test "failures", especially in the early phases of system development, should be received as important learning opportunities and chances to solve problems as they are uncovered.

Recommendations

TEST AND EVALUATION AND THE REQUIREMENTS DETERMINATION PROCESS

1. Establish a stable team made up of users, developers, testers and appropriate contractors called a Combined Acquisition Force (CAF) to streamline the acquisition process for Acquisition Category I (ACAT I) programs. The CAF should be formed once a need is identified and remain in place throughout the acquisition process.

The task force studied the current requirements determination process to identify ways to reduce cost and development time of systems with emphasis on streamlining the testing processes. Early tester involvement in the acquisition process is key to improving testing and reducing cycle time in developing and acquiring defense equipment.

In order to foster a closer relationship among user, developer, and tester, and to capture the benefits of a multidiscipline approach throughout the acquisition cycle, we urge the formation of a Combined Acquisition Force (CAF) for each proposed new start that meets ACAT criteria. We envision that the CAF would be composed of members from the user, the developer, and the tester communities. It could have multi-service representation if appropriate. A contractor representative can be added when a contractor is selected. The duties of the CAF are to assist in the development of the Mission Needs Statement (MNS) and the Operational Requirements Document (ORD). It also assists in the Test and Evaluation Master Plan (TEMP) design. Upon program approval, the CAF continues to act in its role as advisor and reviewer throughout the development process to include Engineering and Manufacturing Development (EMD). For this CAF concept to work the team members must be handpicked by their commands and then remain relatively stabilized on the team. They must have the confidence and the ear of their commanders, and be empowered to speak for their commands in matters relevant to the system being sought. Clearly, the purview of the CAF need not be limited to T&E but should be oriented toward all aspects of improving the efficiency and effectiveness of the acquisition process.

***Additional Requirements Determination Process Recommendations*:**

2. **Change 5000.1 to incorporate the CAF concept.**

3. **Require the development of a Preliminary Test and Evaluation Plan in conjunction with the MNS.**

4. **Conduct an early operational assessment based on the ORD and an operational scenario before milestone I (no actual hardware would be involved).**

5. **Make the CAF a key participant in the Requirements Development Process.**

DEVELOPMENTAL AND OPERATIONAL TEST (DT/OT) PROCESS

1. **Each of the Service DT&OT organizations should be consolidated, to include integrated planning, use of models, simulation, and data reduction. Planning should be totally integrated, and the OSD T&E organizations consolidated. There should be integrated use of models, simulation, and data reduction. Except for limited dedicated Operational Test and Evaluation (OT&E), contractor and government testing should also be integrated.**

Bureaucratic barriers to cooperation and efficiencies are contributors to an increasingly protracted weapons system development process.

Contractor and government developmental testing (DT), government operational testing (OT), government user evaluations, and experiments have become increasingly stratified. Test organizations involved in a typical DoD major acquisition program include the prime contractor (original equipment manufacturer), a Service DT and OT organization, a System Program Office test management activity and two OSD test oversight organizations. All of the above, as well as the Primary Service Command, are involved to some degree in test planning, conduct, analysis and reporting. The Service DT and OT organizations actively participate in the test planning, conduct, analysis and reporting process. In addition, there are a number of government oversight organizations that typically are involved to some degree in test plan review and approval and in test result assessment.

Rigid DT/OT stovepipes are, for the most part, artificial. Barriers, real and perceived, discourage cooperation and in some instances prohibit cooperation. The vast majority of test objectives (80%) provide developmental insights as well as operationally relevant information.

Additional DT/OT Recommendations:

2. **For each development program and its associated test and evaluation effort, special attention should be directed early in the planning cycle (and periodically throughout program development) toward compressing the developmental test schedule wherever practical.**

3. **Facilities within the DOD's Major Range and Test Facility Base should be required to conduct periodic, systematic reviews to determine where data acquisition, reduction and analysis procedures could be improved to increase the efficiency of the T&E process.**

MODELING AND SIMULATION (M&S) IN SUPPORT OF THE TEST AND EVALUATION PROCESS

1. **Establish Oversight and Direction of M&S Development and Employment for T&E.**

   Assign responsibility to Director, Defense Research and Engineering (DDR&E) (at an appropriately high level) for oversight and direction of all aspects of <u>research</u> relevant to M&S, particularly that which supports T&E.

Responsibilities for oversight of M&S research would fit into three broad categories:

- Phenomenology aimed at measurements and basic data collection to support M&S development and employment across the board, to include T&E applications.
- Conceptual model development to include not only the traditional engineering focus or legacy-based improvements, but with more attention to the insertion of revolutionary computational and simulation technology to handle non-linear conditions and uncertainties that influence weapon system effectiveness in realistic combat situations.
- Development and maintenance of a catalog of M&S capabilities and accredited applications, to include categories such as hardware-in-the-loop simulations, man-in-the-loop simulators, physics-based models, and constructive models.

M&S is being used extensively and successfully for training, design trade-offs, and basic engineering work in both the commercial and the defense sectors. Within defense acquisition, and in T&E more specifically, there are scattered areas, such as hardware-in-the-loop testing, with long traditions of high quality work. Opportunities for the use of M&S are most obvious in the developmental testing that is an inherent part of engineering development. Advancing technology also permits M&S to be used in the technical and operational evaluation of weapons systems at various stages in their development, and there have been some exceptionally well-run programs in this regard.

Overall, however, opportunities for M&S in support of T&E have not been well exploited. For the most part, M&S techniques have been developed not to support T&E, but for use in training and in engineering and design efforts in the pursuit of excellence and improved product quality. Interest in the use of M&S to support T&E is a more recent phenomenon. This recent, high level interest in M&S for T&E specifically, and for acquisition more generally, appears to be in pursuit of reductions in acquisition times and attendant cost savings.

2. Needed investment in M&S for T&E.

> ***Modify the directives governing the acquisition process to require identification and funding of an M&S plan for evaluating a program's progress and operational effectiveness/suitability at the earliest practical point in the program.***

Substantial early investment is often needed to capitalize on the potential M&S benefits in the T&E process. A comprehensive M&S plan should be required as part of the Request for Proposal (RFP) process for development of a new system and should be considered as a major evaluation criterion for each contractor's proposal.

3. Understanding of where M&S can contribute to T&E and where it cannot contribute.

> ***Require the users of M&S in their evaluations to state model assumptions, limitations, and uncertainties explicitly, as well as sources for input data in the presentations of results to decision makers.***

The disciplined accreditation process called for in the recommendation above would accommodate this problem, which has contributed to the perception that the benefits of M&S are oversold.

4. Near-Term Priority on M&S Development and Use in Reliability, Availability, and Maintainability (RAM) Evaluations

> ***Give near-term priority, with required funding, to the development of more state-of-the-art RAM models, and to the update of existing models with legacy data from our vast experience with weapon systems in the field.***

These efforts should be aimed not only at an ability to credibly evaluate the suitability of systems in development, but at reducing the cost of ownership of both existing and planned systems. The emphasis on reducing life cycle costs of systems should be translated into increased attention to predicting the reliability, maintainability, and availability of systems in development.

NOTE: A much more detailed examination of M&S in support of T&E is located in the body of this report.

TEST AND EVALUATION FACILITY REENGINEERING

1. **The task force was briefed on concepts for bringing together the management of T&E resources. OSD and the Services should work together to develop a plan whereby T&E resource management is strengthened and brought under coherent control.**

Physical T&E of weapon systems requires large, sophisticated facilities, such as those at Edwards Air Force Base, Patuxent River Naval Air Station and Aberdeen Proving Grounds. In the course of this study, the task force reviewed the overall management of these facilities and visited a cross-section of DoD's T&E facilities to assess their readiness to meet current and future defense requirements. The task force views these facilities as indispensable to continued development of new weapons systems; however, it concludes that management of T&E facilities must adapt to the changing test and evaluation environment to retain their vitality and capability. Broadly stated, the task force concludes that the operation of the Defense-wide system of T&E facilities should be carefully rethought in the light of changing defense procurement and management practices, advancing technology, budget constraints, and an ability to represent actual and potential enemy capabilities in some circumstances.

2. **Apply the reengineering process, employed at Arnold Engineering and Development Center (AEDC), to all test ranges in order to update operations and maintenance and capability investment plans and strategy.**

DoD should undertake a thorough inventory of the Department's T&E facilities and skills and those of other government and non-governmental organizations. This inventory should look beyond DoD needs and capabilities to take into account those facilities elsewhere in government and in the private sector. The inventory should clearly identify DoD facilities needed to meet future near and long term needs, both for test and evaluation capacity and for unique capabilities.

Additional Facility Reengineering Recommendations:

3. **Military departments responsible for acquiring specific weapons systems should continue to test those systems and be responsible for evaluating and reporting results of these tests.**

4. **DoD should develop a T&E Facility investment strategy, based on the inventory and future needs, to assure ability to meet DoD T&E needs through the most effective and efficient combination of all national facilities.**

5. **DoD should establish a set of figures of merit to assess, compare and contrast reengineering progress of each of its major T&E facilities.**

6. **A searching examination of management practices should be undertaken by all test and evaluation centers to address changing test and evaluation technical and business practices, aging facilities, cost and test cycle time reduction and customer service.**

7. **This study limited itself to DoD owned and operated facilities but would observe that a serious look at the T&E facilities should include NASA, DOE, and other agencies test facilities.**

SPECIAL ISSUES

T&E for Advanced Concept Technology Demonstrations (ACTDs)

1. **Direct the participation of a tester in the ACTD process to maximize the value of this concept.**

The purpose of the ACTD process is to introduce new technological capability to the warfighter for their evaluation and development of appropriate tactics. In particular, the technical community gets feedback from the warfighter, while the warfighter gets to check out new technologies and their attendant operational concepts without an upfront commitment to buy. This provides for the possibility of introducing technologies, especially for joint capabilities or other "out-of-the-box" programs that might otherwise not get examined. The requirement of JROC review and CINC sponsorship favors capabilities that have a real, near term mission to perform.

The most useful role for the T&E community in the ACTD program is helping with transition to acquisition status. "Leave behind" products of ACTDs should have adequate operational testing to understand the operational effectiveness before fielding.

Systems of Systems T&E

1. **Create a systems of systems design authority (system designer) with a joint T&E component that uses the Service T&E resources for joint testing.**

"Systems of systems" is a contemporary phrase that can mean anything from a relatively simple integration task that attempts to make components work together to a land, air, sea and space multi-platform, and multi-service joint campaign level interdependent system. For this report systems of systems is defined as those systems that require interoperability or interdependency between functional entities for mission success.

There is currently concern about integrating the information systems of the services and agencies, many of which are not well integrated within themselves, into a joint system which can serve the CINC's effectively in joint operations. There is, at this time, no overall architect of the joint system although many efforts are being made to cope with the resulting problems. Since the joint systems of systems problem is so large and complex and involves so many organizations and so many components of all ages and origins it will be difficult to exercise control over the systems of systems design and test process. And yet, the joint system must be made to work.

Information Systems Software T&E

1. **Insist on Capability Maturity Model levels appropriate to the size and complexity of the system**

2. **Employ alpha and beta testing as employed in the private sector to the extent possible**

3. **Consider machine testing of non-real time software**

4. **Ask DARPA to look into developing machine testing technology for real-time and non real-time software**

There is no exact definition of what constitutes an information system. In general, any information system is made up of a number of subsystems or components and, in turn, is a component of a still larger system sometimes called a system of systems. Furthermore, many military information systems already exist in some form and rarely provide an opportunity for starting over. They are made up of a number of different components, built at different times by different people who may belong to different agencies. The problem is one of upgrade and modification with occasional replacement of a component. There is rarely a central design authority for the system as a whole.

Nonetheless, new components must be built to protocols or standards that permit them to be integrated with already existing systems. T&E plays an important role in the process. The fundamental need is for a well disciplined engineering process for designing, building and testing information systems.

Test and Evaluation of Commercial Off-the-Shelf Software (COTS)

1. **Conduct functional and interface testing of COTS software as a prerequisite to its use in DoD programs.**

The functional requirements for systems in the commercial world have moved closer to the requirements for military systems. This has led to more extensive use of COTS products, including hardware and software, in military systems. The DOD, in policy and practice, has stated a strong preference for COTS products and for a reduction in custom-built software and hardware. However, the use of COTS products can significantly change the process by which systems are built.

Except in trivial cases (such as a word processing or e-mail system), military systems include many software packages. Even if these are COTS products, they will probably come from different vendors and may use different languages, different interfaces and other standards. Even if the individual packages are carefully chosen by the system designers for performance and quality, integration of the software into a working whole may not be easy and may require building substantial amounts of interface software, drivers, etc. This is true for commercial systems as well.

Presumably the COTS software will already have been extensively tested by commercial users and should work well as long as the military application is similar to the usual commercial application and is well within the performance limits of the package. In some cases, however, the military application may stress the COTS software in unusual ways, a problem that may not appear until the system has been completely assembled and tested under realistic conditions.

In Summary

- **Start T&E planning early — very early**
- **Make T&E part of the acquisition process — not adversarial to it**
- **Consolidate DT and OT**
- **Provide joint test leadership**
- **Fund M&S support of T&E in Program Budgets**
- **Maintain independence of evaluation process while integrating all other activities**
- **Establish range ownership and operation structure separate from the Service DT/OT organizations**

INTRODUCTION

The Defense Science Board task force on Test and Evaluation (T&E) was formed by the Under Secretary of Defense (Acquisition and Technology) (USD(A&T)) to undertake a broad review of all activities relating to T&E. The task force is co-sponsored by the Director of Operational Test and Evaluation and the USD(A&T)'s Director of Test, Systems Engineering and Evaluation.

The T&E Structure of the DoD has been built up over a number of years, and today two-thirds of the T&E structure of DoD is over 30 years old. Now, under the guidance of the QDR, procurement is planned to increase 50% by 2003 from a low point in 1996. These procurements will encompass major systems and "systems of systems," upgrades to existing systems, and commercial and non-developmental items. Further, a number of new acquisition techniques have increased early and extensive involvement of the warfighter in acquisition activities. Inclusion of the warfighter as a user will affect future T&E by providing the opportunity to address operational issues early in the development process. Additionally, the variety and ever-changing nature of threat systems that the U.S. and its coalition partners will face in the future requires a new look at the role of T&E in the rapid and responsive acquisition cycles that will be needed to meet these threats.

The task force was asked to:

- Examine new and innovative ways that the T&E community can better support its users.
- Find new ways to integrate operational testing into the overall system development process (follow selected industrial principles and practices).
- Consider the special problems associated with T&E of the "systems of systems" which are increasingly comprising critical parts of our military capability.
- Identify and quantify the current and future needs of the Department's T&E capabilities and resources.
- Recommend specific and quantified changes.

To perform this review the task force examined facilities, processes, policies, contractor testing, government development testing, joint T&E, and operational T&E, both before and after full production and in the continuing evolution of force capability.

The task force formed sub-panels to focus on particular subjects related to T&E. These sub-panels studied requirements determination in the T&E process, the developmental testing and operational testing (DT/OT) process, the role of modeling and simulation in T&E, and facility reengineering. The task force also looked at special issues affecting T&E of Advanced Concept Technology Demonstrations (ACTDs), Commercial Off the Shelf (COTS) products (especially software), and Systems of Systems.

This is a particularly opportune time for such a review. Acquisition reform has been underway for several years to identify opportunities for improving the way we develop and field weapons and support systems. In the forefront of such initiatives are actions that reduce the system development and acquisition cycle time. This is critical because reduced cycle time can reduce costs and put improved capabilities into the hands of our warfighters earlier—thereby providing greater military advantage. T&E tends to be viewed as a "pacing item" in the cycle time for most military systems. We perform developmental tests first, complete system design, and then use operational tests to determine the military utility of the system. That is the perception—and while it is not fully valid, there is clearly a division of focus that probably causes T&E to yield less knowledge early in the program and to take longer than may be necessary. Our DSB review of Integrated Test and Evaluation is clearly timely from the perspective of acquisition reform.

There are other considerations that make this review timely. Joint Vision 2010 is the Department's conceptual framework for how U.S. forces will fight in the future. According to the Secretary's Defense Reform Initiative Report, the ability to achieve the capabilities envisioned in Joint Vision 2010 is contingent upon a Revolution in Military Affairs, to enable our forces to attack enemy weaknesses directly throughout the battlespace with great precision; to better protect themselves from enemy attack; and to receive the right supplies in the right place at the right time. Reducing overhead and support structures will be critical to achieving the Revolution in Military Affairs.

Test and evaluation ranges and facilities are among the major elements of the Department's acquisition infrastructure. They are viewed in many circles as still sized for the Cold War and resistant to downsizing. While this infrastructure has in fact reduced its facilities this review gives us an opportunity to assess the potential improvements to the T&E support structure that might be available though better integration of DT and OT. Improved integration of operational tests into the system development process will not only provide earlier information regarding system effectiveness and suitability, but will also achieve improved efficiency from the infrastructure.

BACKGROUND

The feasibility of reducing acquisition process time through improving the efficiency of T&E was an important consideration of the task force. The role of T&E in DoD acquisition is described in DoD Regulation 5000.2-R under the section of Program Structure, as "part of a strategy to provide information regarding risk and risk mitigation, to provide empirical data to validate models and simulations, to permit an assessment of the attainment of technical performance specifications and system maturity, and to determine whether systems are operationally effective, suitable, and survivable for intended use."

Current DoD policy is that "test and evaluation objectives for each phase of an Acquisition Category (ACAT) I and ACAT IA program shall be designed to allow assessment of system performance appropriate to each phase and milestone. For ACAT I and II programs for conventional weapons systems designed for use in combat, a beyond low-rate initial production decision shall be supported by completed independent initial operational test and evaluation and by completed live fire test and evaluation." Additionally, "operational test and evaluation does not include an operational assessment based exclusively on computer modeling, simulation, or an analysis of system requirements, engineering proposals, design specification, or any other information contained in program documents." The DSB task force reviewed this guidance to determine if any changes are appropriate, and in particular the value of using modeling and simulation.

The task force looked at the roles of both government and contractors in performing test and evaluation, as well as the role of systems engineering and test and evaluation. The task force reviewed how DoD 5000.2R is being followed, how it would apply to ACTD programs, and whether T&E personnel should be involved in the requirements development process.

The DSB task force reviewed constraints by and on the test and evaluation community and the T&E infrastructure that, if removed, might result in reducing acquisition cycle time. We reviewed several acquisition and systems of systems programs to determine how T&E contributed to the acquisition cycle time and to the costs of the programs. The task force looked at commercial and defense industry T&E practices in support of both their development and acquisition programs.

THE TEST AND EVALUATION PROCESS

T&E results are essential information for acquisition decision-makers. The process, depicted in the following figure, is based on defining the information needed, designing an approach to collecting the information required and predicting the results, collecting the data, and then comparing outcomes with predictions, and providing the information back to the acquisition decision-maker. This process applies to DT&E and OT&E, regardless if it is performed by the government or by contractors. The acquisition decision-makers may use this information to assess program risks for costs and schedule, whether the program will reach threshold performance criteria, or as early warning for whether the program will achieve required effectiveness, suitability, and survivability requirements. However T&E data is used, it is integral to the DoD acquisition process.

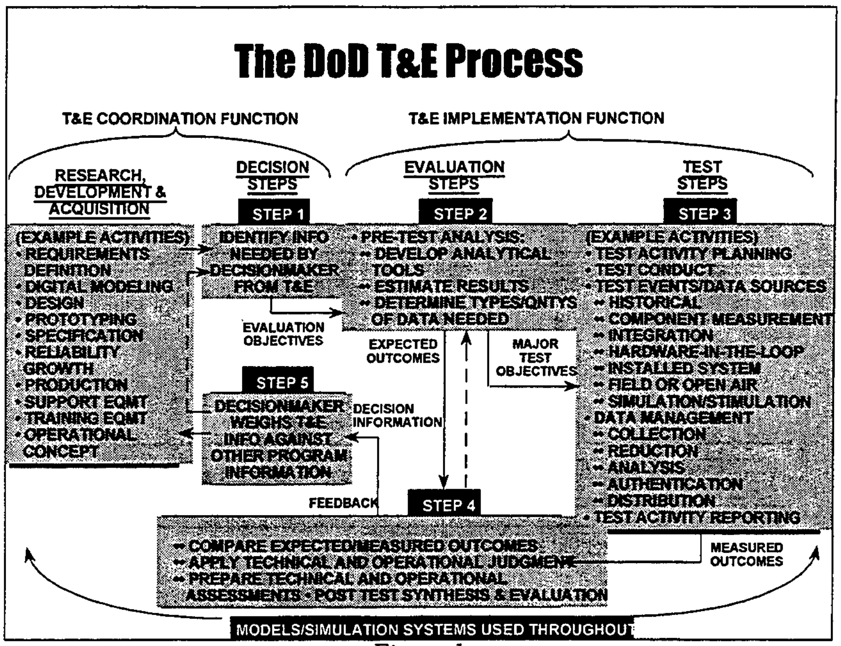

Figure 1

The T&E process, Figure 1, is usually depicted as a parallel process to the acquisition cycle, with emphasis on Developmental Test and Evaluation (DT&E) during the early phases of a program, then gradually shifting to OT&E during later phases. This idea apparently reflects the notion that DT&E is used to select from alternative designs, refine design performance, and to provide information to help assess and manage the risks of the program. In the systems engineering process, it is normally DT&E that is associated with providing the design verification feedback. Emphasis on OT&E is normally associated with preparation for an Initial Operational Test and Evaluation (IOT&E) to seek a Low Rate Initial Production (LRIP) decision, and it is during this period when concerns about effectiveness, suitability, and survivability tend to get the most attention. Figure 2, on the following page, depicts the usual process and the relative emphasis of DT&E and OT&E.

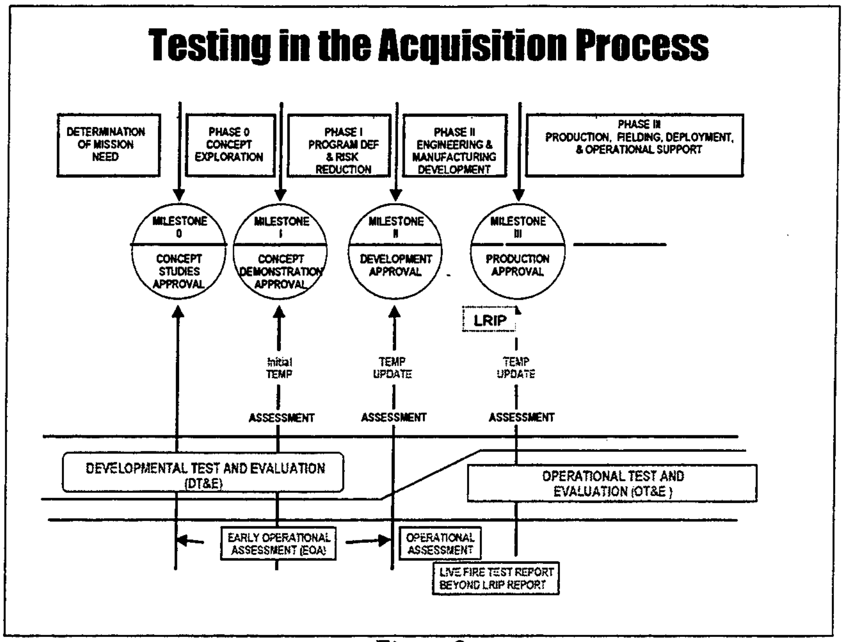

Figure 2

The government, its contractors, and subcontractors all have had responsibilities for performing DT&E, which might also include qualification and acceptance tests as well. Traditionally, DT&E in an acquisition program has been performed by both the government and the contractor. There appears to be a desire to increase contractor responsibility for DT&E. Figure 3, on the following page, depicts the usual involvement of government and contractor personnel.

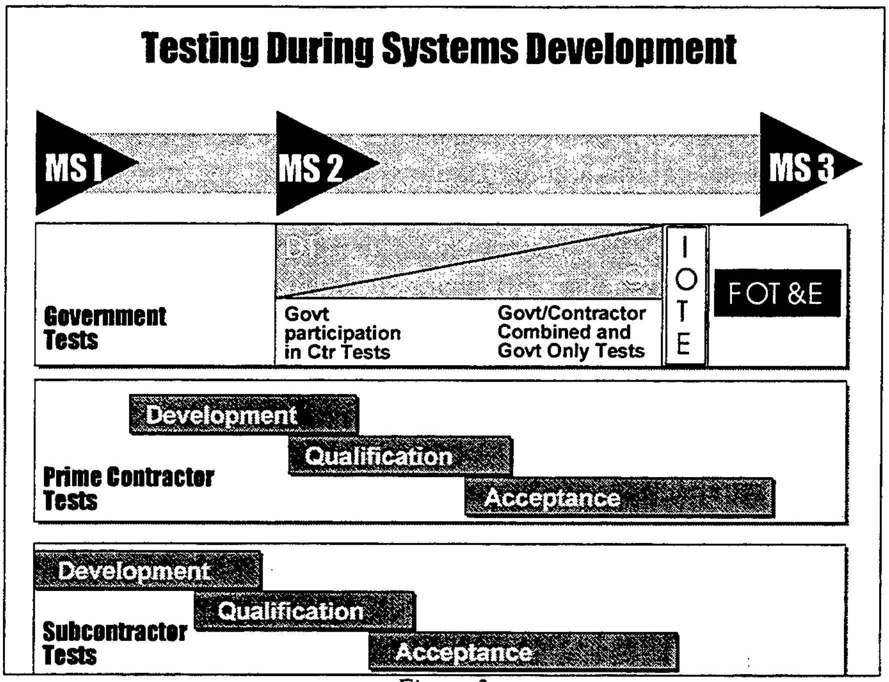

Figure 3

Depending on the type of program, T&E may not follow the process described above. For example, space platform programs vary vastly from the approach for an airframe program with DT&E and OT&E being performed by the contractor in a ground-based space-simulated environment. No directive prescribes OT&E of ACTD programs, therefore, some ACTD programs may have no OT&E involvement. Software-intensive systems and systems of systems programs may call for other non-traditional approaches.

THE IMPACT OF T&E ON STREAMLINING THE ACQUISITION OF MAJOR DEFENSE PROGRAMS

The department's efforts to streamline the acquisition process derive from a long history of attempts to procure modern military hardware faster and at lower cost. This ongoing concern has merged with a much broader desire to make government more efficient. The desire for streamlining has many elements, and there are many reasons why it has become prominent again recently. One factor has been the technology-driven revolution in commercial business affairs. Growing awareness of how American business has become more efficient has fueled the desire to make government more efficient. To some, this means incorporating best business practices into government operations, while to others it means making government operate more like commercial enterprises in more fundamental ways. Acquisition streamlining seeks both to capitalize on better business practices, and to eliminate selected regulations that have accreted over time, but might no longer be useful.

The challenges of introducing new state of the art materiel into the U.S. inventory are daunting. In most cases a new military capability rests on a new technology – or at least great advances in established technology. This implies high technical risk and an associated uncertainty in meeting schedules, especially given limited resources. If maturation of new technologies of a system take longer than expected, other aspects of system design can be put on hold, causing the "technology in hand" to obsolesce as the technology being developed is delayed. Schedules that are technically feasible might not be executable because of funding constraints or other uncertainties. Stretching the program for budgetary reasons can exacerbate total program costs and uncertainties that were high to start with. The long development times that result mean that concepts of operations and threat assessments will change, which in turn forces an evolution of essential requirements.

Test and evaluation can play two important roles in these situations. First, an early DT&E program can identify high-risk technical areas at the outset. Second, what is discovered can be linked to operational assessments of the concept or the design. Mapping technical risk early in a program, and determining the operational impact of not achieving particular technical goals, can have great influence on shrinking the acquisition cycle. In addition, specific assessments of the linkages between technical risk and operational performance can support the execution of Cost as An Independent Variable (CAIV) in a disciplined, rational fashion. Test and evaluation that is both technically and operationally oriented is an essential part of any well-run program. High value comes from learning as early as possible, in as much detail as possible, how the technical challenges and the operational effectiveness are linked.

The key element in acquisition streamlining so far as T&E is concerned is increased early involvement of the T&E community. This involvement needs to include thorough developmental test programs focused on identifying and characterizing the technical risks. These need not be government-conducted tests; contractor tests that aggressively pursue technical risk, with program management office (PMO) and oversight office visibility, would be fine. Also needed are early operational assessments of the system under development. These assessments should not be unrelated but should be closely linked to the testing to identify technical risk. Such a combined program is essential if the PMO is to concentrate on the operationally critical but technically challenging features and appeal for relief on the technically challenging features that promise little operational pay-off.

In short, what is needed is a T&E program that is comprehensive enough to relate operational capability and technical risk, early enough to eliminate low pay-off risk, and focused enough to concentrate on the high pay-off technical challenges.

TEST AND EVALUATION AND THE REQUIREMENTS DETERMINATION PROCESS

The task force studied the current requirements determination process to identify ways to reduce cost and development time of systems with emphasis on streamlining the testing processes. The Secretary of Defense asked for a determination of how further improvements could be made in the testing process to reduce cycle time in developing and acquiring defense equipment. As one of five themes to reform the acquisition process, the Secretary called for "early tester involvement" in system development.

TESTER INVOLVEMENT IN THE ACQUISITION PROCESS

To understand where and how testers are currently involved in the acquisition cycle, depicted in figure 2 on page 14, is helpful to show the DoD scheme as portrayed by today's regulations.

The first tester benchmark is the publishing of an initial test plan, the Test and Evaluation Master Plan (TEMP), at milestone I. In practice, testers are brought into the concept exploration phase too late to properly formulate the initial TEMP. By this time in the acquisition cycle, the performance parameters and the formal user requirements have been through several reviews and hard fought staffing. The late arriving testers are forced into a reactive mode to understand the operational concept, how it will fit into the future warfighting doctrine of the users, decide which of the performance parameters should be tested, and, of these, which are even capable of being tested. It is a steep "learning curve" for the testers, and it fosters the beginning of an adversarial relationship between developers and testers. The initial TEMP, therefore, becomes a negotiated compromise between testers and developers. Moreover, this late inclusion of testers tends to put them in the role of disrupters rather than helpers whose insights could have served to anticipate problem areas and seek remedies to avoid them.

The fundamental purpose for testing is to discover and to learn. The testers as well as the developers and users are the learners in this process. In order to build mutual trust and confidence, and to focus on system operating characteristics that can be tested in a way that makes sense, the expertise of the testers should be sought by the users and developers as the system requirements are being formulated. This means that the testing community should be a part of the *requirements development process.* It is not too early to bring in testers during the "determination of mission needs" sequence and make them full members at this phase. The Army's document titled "Requirements Determination" is a case in point. It directs the formation of an Integrated Concept Team (ICT) which is charged with developing formal requirements for new hardware through the synergy of a multidiscipline team. It falls short of calling for testers to be mandatory team members, however. Nevertheless, the concept is valid. If the Army follows its own directive and adds a tester to the ICTs that it forms, then it will have established a model that could serve DoD wide.

These streams of activity apply not only to major new systems but also to the evolutionary improvement in small steps of the joint military capability.

Three changes to the process are recommended:

1. Bring testers (DT and OT) into the requirements generation process as early as the initiation of Mission Need Statement (MNS).
2. Keep the user (both service and joint) and testers involved throughout system developments.
3. Relate DT and OT in time, space, data, and test criteria. User exercises can produce data useful in both DT and OT without having to do separate exercises.

Any modifications in the current acquisition process (to include the operational requirements development/documentation process) will not significantly improve the process unless the individuals involved are exposed to each other on essentially a daily basis. The challenge to be met is elimination of the isolation of the individuals representing the different organizations involved in the acquisition process.

In order to foster a closer relationship among user, developer, and tester, and to capture the benefits of a multidiscipline approach throughout the acquisition cycle, we urge the formation of a Combined Acquisition Force (CAF) for each proposed new start that meets ACAT I criteria. We envision that the CAF would be composed of members from the user, the developer, and the tester communities. It could have multi-service representation if appropriate, and a contractor representative is added after a contractor is selected. The duties of the CAF are to assist in the development of the Mission Needs Statement (MNS) and the Operational Requirements Document (ORD). It also assists in the TEMP design. Upon program approval, the CAF continues to act in its role as advisor and reviewer throughout the development process to include Engineering and Manufacturing Development (EMD). For this CAF concept to work, the team members must be handpicked by their commands and then remain relatively stabilized on the team. They must have the confidence and the ear of their commanders, and be empowered to speak for their commands in matters relevant to the system being sought.

The matter of chain of command for a CAF is critical to its success. As we see it, the CAF reports to the service commander responsible for each phase of the development. For example, a CAF appointed to assist in the development of a shoulder fired anti-tank weapon for the Army would report to the Army's Training and Doctrine Command (TRADOC) commander until Milestone I. At Milestone I, the CAF would then report to the acquisition executive. In all phases, the CAF will keep appropriate commanders abreast of developments, and be a sounding board for the staffs and commands who play a role in funding and guiding the acquisition system. Although it is not envisioned that membership on a CAF will be a full-time duty, it is clear that such duty must be a number one priority for each of its members.

How the "System" Should Work

Figure 4 below displays the process as the task force sees it — a continuous stream of activity leading to improvements in joint military capability.

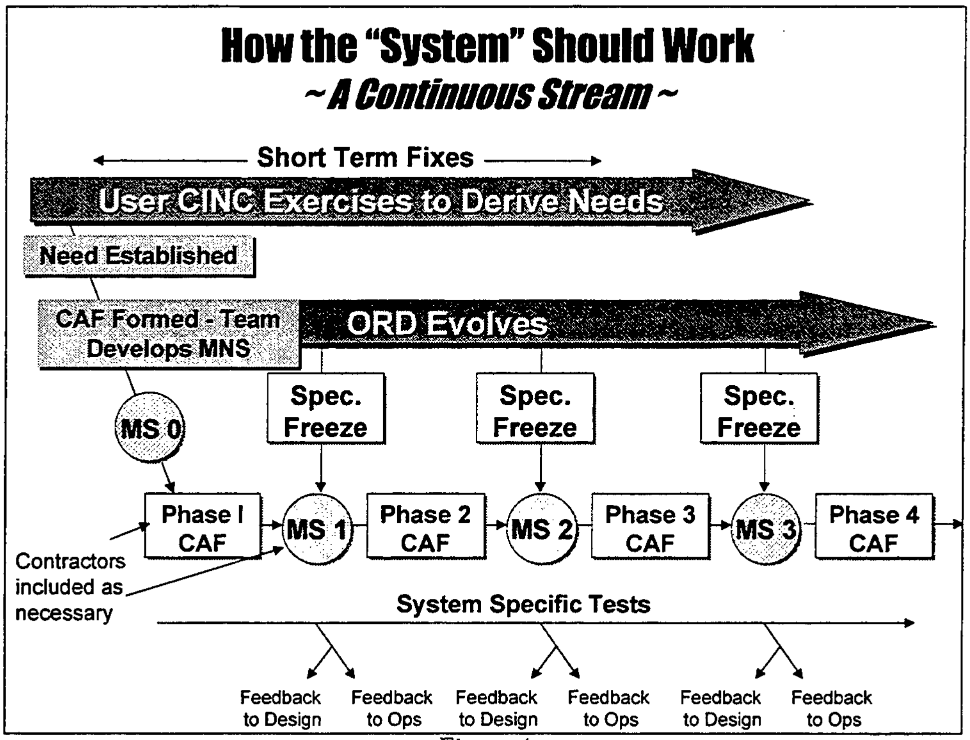

Figure 4

The first stream is one of CINC or user exercises that lead both to short-term fixes and to recognized needs for improvement. The developers and testers participate in these exercises in order to derive common understanding of the needs.

Once a need is recognized, a CAF is formed to assist in the MNS and subsequently in the ORD. The team assures that the proposed system is both doable and testable.

As the project proceeds through its various Milestones and Phases, the CAF stays in force. The ORD and system specifications are assumed not to be fixed but to evolve as information about the need, the development, and the testing warrant changes. The CAF is there to participate in this process.

Associated with the project is a continuous stream of system tests. In the beginning these are largely analysis, simulation, component, and breadboard tests. As the project proceeds and the system comes into being, testing becomes more formal including DT&E and OT&E. The testing provides continuous feedback both to the users to make sure they know what they are getting, to gain their approval, and to the developers to make sure they are building the right thing and ensure that it works properly.

To complete these modest revisions to the acquisition process and perhaps to have the greatest impact on reaching the goal, we should change the basic philosophy of military hardware and software development. Namely, focus new development on making evolutionary improvements to warfighting systems rather than revolutionary changes to individual parts of warfighting systems. Each of the services has in some stage of R&D today a large number of systems, many more than can be afforded by current and projected procurement funds. Each of the new systems can be shown to be better than the systems they are to replace; however, only some can be afforded. Therefore, the strategy is to refocus R&D on those items that do most for evolving warfighting system improvements, do not start R&D on projects that offer revolutionary gain to a sub system, but offer little total system improvement; and use the resulting cost avoidance to realize shorter procurement times for the evolutionary improvements.

Recommendations

1. **Establish a stable team made up of users, developers, testers, and appropriate contractors called a Combined Acquisition Force (CAF) to streamline the acquisition process for ACAT I programs. The CAF should be formed once a need is identified and remain in place throughout the acquisition process.**

Rather than accepting or rejecting the Solution Option out-of-hand, the construct of a CAF could be tested. Pick an emerging weapon program and mandate collocation of all acquisition representatives. Test the concept for two years and make a cross-DoD decision regarding the utility and benefits of the approach. Such a field 'test' would also provide significant data to either justify, modify, or reject use of the CAF construct in the future and be a source of important 'lessons learned' for its future application.

The Joint Staff should provide a vision of the Acquisition Process for Joint Systems to include how the voids mentioned in the previous section will be addressed. This action should be accomplished before the U.S. Atlantic Command (ACOM) 'experimentation' activity takes final shape. A poorly constructed strategic process could wreak havoc on Acquisition (T&E) budgets across the DoD.

2. **Change 5000.1 to incorporate the CAF concept.**

3. **Require the development of a Preliminary Test and Evaluation Plan in conjunction with the MNS.**

4. **Conduct an early operational assessment based on the ORD and an operational scenario before milestone I (no actual hardware would be involved).**

5. **Make the CAF a key participant in the Requirements Developments Process.**

DEVELOPMENTAL AND OPERATIONAL TEST PROCESS

The task force examined Developmental and Operational T&E processes and recommended improvements and reengineering that will reduce test cycle time and cost.

THE CURRENT DT/OT PROCESS

Over the last 25 years the DoD test management, and test oversight organizations have evolved and new organizations have been created. In the early to mid-1970's, DoD created AFOTEC and OPTEC (OPTEVFOR already existed) in response to the President's Blue Ribbon Defense Panel's recommendation that independent operational test agencies be established in all the Services. In addition, an office in OSD was set up with responsibility for oversight of T&E activities throughout the Department. In 1983, Congress passed a statute that established an independent Director of Operational Test and Evaluation reporting directly to the Secretary of Defense.

Contractor and government developmental testing (DT), government operational testing (OT), government user evaluations, and experiments have become increasingly stratified. Test organizations involved in a typical DoD major acquisition program include the prime contractor (original equipment manufacturer), Service DT and OT organizations, a System Program Office test management activity and two OSD test oversight organizations. All of the above, as well as the Lead Service Using Command, are involved to some degree in test planning, conduct, analysis, and reporting. The Service DT and OT organizations actively participate in the test planning, conduct, analysis, and reporting process. In addition, there are a number of government oversight organizations that typically are involved to some degree in test plan review and approval and in test result assessment. These organizations include the SPO, OSD (A&T), DOT&E, and FFRDC's (IDA, MITRE, Aerospace Corporation). The service DT and OT organizations are involved in test conduct, and other government entities play a role in planning, analysis and reporting. With its Title 10 oversight and reporting responsibility, DOT&E plays the dominant role in operational test oversight for major development programs. In addition, there is a large test range infrastructure, managed and operated by the services, that supports system developmental and operational testing and to some degree subsystem development and operational testing.

The Service developmental and operational test organizations work together to varying degrees. The Army is currently in the process of combining the government test management and, with some notable exceptions, the test range infrastructure within that service. For aircraft flight test programs, the Air Force employs a combined test force concept, which consolidates some elements of the contractor, government DT and OT organizations that are directly involved with test program planning and execution.

Within the DoD, Developmental Test and Evaluation (DT&E) is defined as that T&E conducted throughout the acquisition process to assist in the engineering design and development process and to verify the attainment of technical performance specifications and supportability objectives. The Government developmental test organization is a part of the Service Acquisition Command. In a typical DoD program, the government and the manufacturer are both heavily involved in developmental testing and analysis. (In contrast, in a typical commercial program there is little, if any, direct customer involvement in the product developmental process.)

Operational Test and Evaluation (OT&E) within DoD is defined as the field test under realistic conditions to determine the tested item's operational effectiveness and suitability for use in combat by typical users. The Service operational test agencies report to the Service Chiefs of Staff and, with the exception of the Army, are not involved in the management of test or training ranges.

Findings

Bureaucratic barriers to cooperation and efficiencies are contributors to an increasingly protracted weapons system development process.

The rigid DT/OT stovepipes previously described are, for the most part, artificial. Barriers, real and perceived, discourage cooperation and in some instances prohibit cooperation. For example, the vast majority of test objectives (80%) provide developmental insights as well as operationally relevant information. The stovepipes create a climate that fosters separate, duplicative testing, separate data bases, isolated development and use of models and simulations, and data processing beyond first-generation engineering units. The separation of governmental test organizations and duplication of management overhead is counterproductive. In addition, the government acquisition community needs to focus on evaluation and not use a high percentage of its declining resources on detailed test planning and test conduct.

More emphasis is needed on reducing time required for DoD test programs.

DoD test programs undergo extensive technical and safety reviews, but little attention is paid explicitly to test cycle-time reduction. There is the direct benefit to be realized in completing testing earlier or compressing the test schedule. There are several basic mechanisms by which test cycle-time can be reduced:

- Reduce test program content
- Accomplish testing more effectively/efficiently
- Use test facilities and resources more intensively (e.g., multiple shifts, seven-day weeks, etc.)
- Eliminate duplicative testing
- Budget and fund testing and test planning earlier in the program

In most DoD test programs, the content is already at or near a minimum. It could be argued that, for some programs, test content has been reduced below an acceptable minimum, and in fact increasing test content could reduce total acquisition cycle time. A 6- or 7-day workweek is the norm on commercial aircraft development test programs, but military aircraft developmental programs usually operate at a more leisurely pace (except for competitive fly-offs). While a review of program content should not be overlooked, more opportunities for cycle-time reduction probably exist in the area of test process effectiveness/efficiency. Efficiencies can be realized by exploiting advances in information processing technology. So-called "fast-track" T&E programs are common in commercial practice, where lengthy "time-to-market" can have an enormous price tag, up to and including survival of the company. Lengthy DoD development programs also result in significant increases in both direct and indirect costs. Under the current management structure there is little, if any, incentive on the part of government developmental test personnel to accelerate the developmental test process.

> ***The duplication that exists in many programs between contractor and government developmental and operational testing also adds to system development time.***

The Integrated Product Team (IPT) process has improved the dialog among the numerous agencies involved in the test execution and oversight process, but much more streamlining can be done.

> ***For many major weapon system development programs, the fixed overhead cost of the test support personnel and associated infrastructure can be a greater cost driver than the variable per test or per mission cost.***

This latter cost includes the direct cost associated with such items as range support and data processing. One exception is a test program where a high-cost test asset is destroyed in every full-up test, e.g., most missile flight tests. Adding test support personnel to accelerate a test program can be less expensive over the duration of a development program than conducting a program that is stretched out over several years. More importantly, decreasing the time needed for the test program, particularly during early development, allows for the identification and correction of deficiencies earlier in the development process, and consequently lowers overall development costs.

Recommendations

1. **Each of the Service DT&OT organizations should be consolidated, to include integrated planning use of models, simulation, and data reduction. Planning should be totally integrated, and the OSD T&E organizations consolidated. There should be integrated use of models, simulation, and data reduction. Except for limited dedicated OT&E, contractor and government testing should also be integrated.**

2. **For each development program and its associated test and evaluation effort, special attention should be directed early in the planning cycle (and periodically throughout program development) toward compressing the developmental test schedule wherever practical.**

3. **Facilities within the DoD's Major Range and Test Facility Base should be required to conduct periodic, systematic reviews to determine where data acquisition, reduction, and analysis procedures could be improved to increase the efficiency of the T&E process.**

MODELING AND SIMULATION IN SUPPORT OF THE TEST AND EVALUATION PROCESS

For decades, Modeling and Simulation (M&S) has been used in various forms to predict and evaluate the performance of weapon systems. Many of the simulation tools most often used today—computer-based models of complex weapon functions and the even more complex human interactions with those weapons and the surrounding environment—evolved from models used in the early 1960s. Currently, the large number of simulations in use includes very complex models, known as physics-based or engineering models, that attempt to capture the basic physics and engineering inherent in the design of a system and its operating environment, to include materials, structure, aerodynamics, propulsion, sensors, electronics, and so forth. They also include so-called higher order models or constructive simulations, which are digital computer models that attempt to represent combat elements and their functions in force-on-force engagements, along with the environment in which they operate.

Other important and widely used simulations known as hardware-in-the-loop involve actual hardware as an integral part of the simulation itself. Sometimes the hardware includes a complete system such as a missile, but more typically included are weapon subsystem elements such as sensors, guidance and control units, aerodynamic surfaces, propulsion systems, etc. Finally, there are the man-in-the-loop simulations that use displays and sensory stimulants for feedback, some with actual hardware-in-the-loop, including both fixed and motion-based simulators.

Models and simulations are being used extensively and successfully for training, design trade-offs, and basic engineering work in both the commercial and the defense sectors. Within defense acquisition, and in T&E more specifically, there are scattered areas, such as hardware-in-the-loop testing, with long traditions of high-quality work. Opportunities for the use of M&S are most obvious in the developmental testing that is an inherent part of engineering development. Advancing technology also permits M&S to be used in the technical and operational evaluation of weapons systems at various stages in their development, and there have been some exceptionally well-run programs in this regard.

Overall, however, opportunities for M&S in support of T&E have not been well exploited. For the most part, M&S techniques have been developed not to support T&E, but for use in training and in engineering and design efforts in the pursuit of excellence and improved product quality. Interest in the use of M&S to support T&E is a more recent phenomenon. This recent, high-level interest in M&S for T&E specifically, and for acquisition more generally, appears to be in pursuit of cost savings.

In general, those who trumpet the usefulness and benefits of M&S for defense acquisition programs have been advocates within the M&S community, or acquisition officials under increasing pressure to reduce both development time and costs of their programs. However, several recent studies of the benefits and cost savings to be gained through the use of M&S tools in the T&E process have raised substantial issues about the validity of advocates' claims. For example, the November 1998 Air Force Scientific Advisory Board

(SAB) Review of Air Force Test and Evaluation concluded that "[t]he high and unrealistic expectations for Digital Modeling and Simulation (DMS) contributions to T&E processes are underwriting a rationale that the Air Force and DoD can ignore ongoing and planned T&E infrastructure investments." Also, a 1996 SAIC study funded by OSD, entitled "Study of the Effectiveness of Modeling and Simulation in the Weapon system Acquisition Process," concluded that "cost savings are especially difficult to quantify and reported cost savings are often illusionary."

One problem has been that M&S developments and applications that are "top down" driven frequently focus on eye-catching graphical displays of information and user interfaces rather than on valid representations of the complex physical and human interactive processes inherent in the design and employment of weapon systems. This approach often results in models that are unsupported by data or analytic expertise.

There also appears to be a heavy focus on improving current legacy-based simulation approaches at the expense of inserting advanced computational technologies. For instance, there appears to be no ongoing substantive relationship between DoD agencies and Department of Energy laboratories aimed at employing DOE's new TERAFLOP (trillion operations per second) computer technology. A good example of the enhanced capability provided by this new technology is the Los Alamos Laboratory's Transportation Analysis Simulation System (TRANSIMS) transportation model, which can accommodate an order of magnitude more objects in the simulation than DARPA's Synthetic Theater of War (STOW).

Clearly, a better coordinated and more disciplined process is needed for the development and use of models and simulations, and for their Verification, Validation and Accreditation (VV&A). This more disciplined approach must also be a more visible and transparent process, particularly if use in test and evaluation is planned.

Findings

M&S has been very beneficial and cost effective in many applications, particularly at the engineering design level, and especially when there is a clearly defined application.

Successes that begin at the engineering design level are generally the result of organic developments embedded in a program where both expertise and data are available, and where the developing and using organizations have a common mission. Most often, successful applications of engineering or physics-based models are planned and carried out by contractors who have a clear stake in their validity and accuracy. These models tend to be most useful when their parameter values are based on solid research, experience, or the laws of physics.

The aerospace industry has used sophisticated M&S tools for many years. NASTRAN is used for basic structural design, computational fluid dynamics (CFD) codes are routinely used for aerodynamic and aerophysics analyses, and UNIGRAPHICS and CATIA are used for air vehicle design. Similar M&S tools have been developed and used for other weapon systems. But while there have been many beneficial applications of these tools, especially by industry, not all have been successful.

For example, existing CFD tools failed to predict the vortex problems behind the C-17 that interfered with the airlifter's ability to carry out its airdrop of paratroopers mission. Flight testing eventually uncovered this problem. Likewise, these tools were not capable of predicting the F/A-18E/F wing drop problems, or even of successfully analyzing proposed fixes after the phenomenon was discovered in flight testing. In another case, NASTRAN-based analyses of the C-5A wing resulted in a wing design providing only about one-fourth the required wing durability.

Despite such failures, M&S appears to provide an overall valuable evaluation tool where there is a clearly defined application, an adequate knowledge of required input data and its appropriate use, and a forthright appraisal of the model in question, leading to an understanding of its limitations.

Models and simulations tend to be least useful when the inputs are largely unknown or uncertain.

Among the most overrated M&S tools are large-scale constructive force-on-force combat simulations and models used to predict weapon platform contributions to combat outcomes. To a large extent, the parameters used in these models are not based on solid experimental data or research. In many cases, the inputs are simply guesses by the model's builders or users, or they have been altered from known values in an attempt to compensate for perceived or known model limitations.

Many of the constructive legacy models used for combat analyses, such as BRAWLER, TACWAR, ESAMS, TRAP, and SUPPRESSOR, were developed years ago. Although these have been updated, they are not the comprehensive evaluation tools that today's technology might support if significant funding was available. These models are considered useful only for relative trade-off studies, but not for predicting actual combat outcomes. The few cases where comparisons of weapon system combat capabilities have been attempted have resulted in highly variant outcomes. For example, the F-22 Cost and Operational Effectiveness Analysis (COEA) used the TAC BRAWLER model and displayed F-22 exchange ratio advantages over the F-15 ranging from 8-to-1 to 20-to-1, depending on the combat mission and scenario. In the end, the Air Force agreed at Milestone II to demonstrate a 2-to-1 effectiveness advantage on the part of the F-22 as one of the criteria for entering production. More recently, the use of TACWAR as a force-on-force combat simulation was criticized in the Quadrennial Defense Review (QDR) for its inability to accord adequate consideration to air power contributions to campaign outcomes.

Much of the extensive investment in M&S by DoD seems to emphasize model architecture, interfaces, graphical displays, and code writing at the expense of conceptual model development and basic data collection.

Most of the high profile, force-on-force, constructive modeling efforts lack the underpinnings of basic science and engineering, both physical and behavioral. Interactive, operationally oriented models have not replicated the more complex interactions that are at the core of military combat. The emphasis on animated, high-resolution graphics provides an opportunity for "easy and ready-made analyses," but conveys to the decision-maker no understanding of the basic engineering, physics, or input assumptions. Clearly, more investment is needed in conceptual modeling, phenomenology, and experimentation to gather realistic input data in order to provide decision-makers with confidence in the results of M&S efforts. The application of cutting-edge technology, such as that demonstrated in the TRANSIMS modeling effort, is an area that promises significant payoff.

Much of the emphasis by the Defense Modeling and Simulation Organization (DMSO) and the Simulation Interoperability Standards Organization (SISO) has been on model architecture (for example, the High Level Architecture (HLA) program), and on interfaces, interoperability, and code writing. DMSO has funded a major effort involving many DoD organizations to develop a Conceptual Model of Mission Space (CMMS) describing the air/land/sea environment for use by all model developers. The project is ambitious and expensive. Another Department priority is the ongoing development of the Joint Modeling and Simulation System (JMASS), intended to develop and facilitate interfaces between legacy tools and new models. This latter effort has been underway now for several years but has undergone at least one major restructuring, so that its success is uncertain at this time.

Some progress has been made, however, by the Joint Advanced Distributed Simulation (JADS) test force in demonstrating distributed simulation interfaces. Sponsored by OSD's Joint Test and Evaluation (JT&E) office, the joint test force has carried out several demonstrations in its investigation of the utility of advanced distributed simulation technologies for test and evaluation purposes. For example, the so-called system integration test successfully linked air-to-air missile and fighter aircraft simulators on the ground with instrumented aircraft flying on the test range. Even this modest success, however, has not been rewarded with follow-on use of this simulation technology by any major weapon system acquisition program. Technical issues such as latency, as well as the cost of developing and maintaining such advanced simulation capabilities, no doubt dampen enthusiasm on the part of potential users.

> ***Verification, Validation and Accreditation (VV&A) as presently practiced with respect to M&S techniques in general is not sufficiently disciplined to inspire confidence in their use in the T&E process.***

VV&A does not receive the attention or resources needed to provide decision makers with even minimum levels of confidence in M&S tools for most applications. The use of data from tests and operational exercises to evaluate or validate model outputs appears to be minimal. The use of such "real world" data as model inputs appears to be equally rare. A disciplined method to ensure feedback of test and exercise results in the continuous upgrade of M&S tools should be an integral part of any approval process for investing in a model's development or employment in a particular application. This iterative process is often promised but seldom followed. Planned VV&A activities are often cancelled for cost and schedule reasons, just as testing itself often is cancelled or postponed.

A good example of how VV&A can be overlooked was provided in connection with the Army's state-of-the-art computer model, SQuASH (Stochastic Quantitative Analysis of System Hierarchies). SQuASH had been used since the late 1980s to assess the vulnerability of armored vehicles. In 1993, an independent validation effort compared SQuASH predictions of M-1A1 tank vulnerabilities to actual results obtained in earlier live fire tests. This effort was undertaken in response to an Army proposal to use this model as the primary data source for evaluating the M-1A2 tank's vulnerability. The disturbing results indicated gross mismatches between many of the model's predictions and the actual field test data. As a result, the proposal for assessing the new tank's vulnerability was modified to place more reliance on live fire testing. The Army also developed a plan to establish a "Red Team" and to undertake a full-blown VV&A of the model. (Unfortunately, the latter plans never came to fruition, primarily because of funding limitations.)

The Joint Accreditation Support Activity (JASA) can be a potentially significant resource to assist in the VV&A process. This small group has provided VV&A consulting support to several programs, including the F/A-18E/F, Tomahawk, Joint Strike Fighter, and AIM-9X. However, JASA is severely limited in that the organization has no institutional budget and must rely on customer programs for operating funds. This results in fluctuating staff size and reliance on short-term contracting support to accomplish their VV&A tasks, a very inefficient approach for what could be an important M&S resource.

There appears to be no authoritative single catalog or list of available models or simulations that includes descriptions, applications, level of VV&A achieved, limitations, input data requirements, or other information needed to inspire confidence in their use.

Various types of M&S are used in the T&E process, including large constructive force-on-force models, real time missile or projectile trajectories used on test and training ranges, hardware-in-the-loop and man-in-the-loop simulations, engineering or physics-based models, etc. Each has different applicability to test and evaluation activities, different development and maintenance costs, and different VV&A requirements.

Keeping track of this plethora of tools is not easy. Several repositories of information on models and simulations exist, some of which include the VV&A status of models and their uses. For example, DMSO maintains its Modeling and Simulation Resource Repository (MSRR), while Survivability/Vulnerability Information Analysis Center (SURVIAC) maintains a repository for many constructive models. However, most of the existing repositories tend to be quite specialized and oriented to their specific sponsors' needs. What is needed is a comprehensive single repository or catalog maintained and adequately funded at a higher level than is presently the case. The availability of such an authoritative catalog could assist in the VV&A discipline process and could preclude unnecessary duplication of effort by programs requiring M&S techniques.

For some weapon systems and warfare situations, open air testing is not feasible.

Despite the overselling of M&S and the potential for misuse, practically every mission or warfare area appears to be dependent to some degree on M&S techniques for meaningful evaluations of candidate weapon systems. Many situations and/or weapons systems that cannot be fully evaluated through open air testing alone require access to alternative evaluation approaches. In many cases, it isn't cost or schedule considerations alone that create reliance on means other than live testing.

Evaluating the survivability of an aircraft system, to include the effectiveness of its defensive avionics subsystems, is a good example of such a situation. Clearly, we cannot fire live missiles (even with inert warheads) against manned aircraft during testing. We are thus forced to rely on different forms of M&S. These include surface-to-air or air-to-air missile flyout models, often incorporated into the test range's instrumentation or data reduction routines; Hardware-in-the-loop (HWIL) simulations, such as the Air Combat Environment Test and Evaluation Facility (ACETEF) facility at Patuxent or the RFSS (Radio Frequency Simulation System) at Huntsville; and even constructive digital models that extrapolate (or expand) the one-on-one results from open air tests and HWIL simulations to many-on-many or campaign level evaluations.

But even in the case of aircraft survivability tests that depend almost exclusively on missile flyout models, the models should replicate the engineering and physics associated with the specific missile/target combinations and environmental conditions, and/or demonstrate good correlation with key results, such as miss distances, from actual missile firings against live targets. Such live firing data should be collected from both test and training exercises, such as Combat Archer, to include live shots against target drones maneuvering aggressively to simulate piloted aircraft; these results should be folded back into the flyout models as part of a continuous update program.

The F-22 operating in the modern integrated air defense threat environment for which it was designed cannot be tested realistically on any existing range. Nor can we hope to fully test the operational effectiveness of space-based sensor systems, such as the Space-Based Infra-Red Sensor (SBIRS), or sophisticated missile defense systems, such as Theater High Altitude Area Defense (THAAD) or National Missile Defense (NMD), across either the spectrum of operational scenarios or against the numbers or types of threats for which they are being developed. Full system effectiveness and suitability evaluations depend on an array of models and simulations that must be developed and made available to analysts and decision makers to augment data gathered during limited live firings in support of system acquisition milestones.

Likewise, evaluations of the lethality of our weapons and the vulnerability of our weapons platforms as part of the Live Fire Test and Evaluation (LFT&E) program depend on the use of M&S to extrapolate from affordable live firings, component testing, and the use of surrogates. In this context, government agencies as well as industry employ various physics-based models, finite element models, as well as empirical models to assess the complex interactions inherent in the lethality and vulnerability calculations used in LFT&E.

A potentially fruitful area for effective use of M&S in support of T&E is in the test design and planning process.

M&S techniques can be employed at the outset of a program to determine the critical technical characteristics or operational capabilities that a new system must possess and be tested against in order to justify its development and production. These results can then be used to guide technical and operational test designs in order to focus the scenarios, tactics, threat representations, countermeasures, statistical confidence levels, etc., on the critical technical parameters or operational issues during the detailed test planning and execution phases. However, full realization of M&S techniques' potential benefits for scenario and operational concept development will require more capable self-learning ("adaptive") tools not currently in use.

A notable recent successful use of M&S for test design and planning occurred in the Apache Longbow OT&E. The contractor's full-mission simulator was used to train aircrews and to develop tactics prior to field tests, while U.S. Army Missile Command's (MICOM) HWIL simulator was used for missile flyout predictions prior to live Longbow Hellfire firings on the test range. VV&A of the simulators was generally successful, with simulator fidelity judged to be adequate for aircrew training and pre-test planning.

> ***Another potentially high payoff area for the application of M&S techniques is the Reliability, Availability, and Maintainability (RAM) field.***

Improved forecasting of RAM through M&S applications could build on the successes in the engineering design field. The Defense Department's emphasis on reducing operating and support costs will not bear fruit unless the logistics system can be simulated well enough early in a program to identify problem areas and permit tradeoffs to be carried out before designs are concrete.

Logistics Composite Model (LCOM), developed in the 1960s, is still the primary suitability/logistics analysis model. Although it is a legacy model, LCOM has been updated on a near continuous basis and extensively modified over the years. But in spite of efforts to remain current, there are still some problems associated with the use of such M&S techniques in system design and subsequent RAM evaluations. Typically, these logistics models contain many factors difficult to quantify in algorithms, such as the system's concept of operations (CONOPS), motivation and training maturity of the maintenance personnel, allocated manning and skill levels, and the efficiency of the supply chain. There appears to be no systematic effort to collect data on operational systems in a variety of environments and at different operational tempos for feedback into these models to ensure they accurately reflect real world situations.

Unfortunately, not enough emphasis is placed on logistics analyses early in development. Availability models are too often used to show that weapon systems are supportable and suitable, instead of determining realistic support requirements and attempting to identify and fix support problems.

More credible and accurate forecasting of RAM will allow for improved systems engineering, improved operational testing, and reduced ownership costs for new systems. Credible RAM modeling and simulation capabilities are necessary in that, during the testing phases of a new system, measurements of these parameters are typically made for relatively immature hardware/software, requiring extrapolation to predict the life cycle costs of the system.

It is extremely difficult to measure the cost and time benefits associated with the use of M&S in the T&E process.

Claims of substantial program cost savings attributable to the increased use of M&S, with a concomitant reduction in testing, cannot be verified. In addition to the Air Force SAB and SAIC reports referred to previously, a White Paper prepared by the AIAA Flight Test Technical Committee (FTTC) in late 1998 entitled, "Seeking the Proper Balance between Simulation and Flight Test," states "the members of the FTTC are unaware of any study that has supported the claim of substantial program cost savings realized by a significant expansion of the use of M&S with a concomitant reduction in testing."

The 1996 SAIC report documents numerous claims of cost savings attributed to M&S use, some of which border on the incredible. In many of the cases reported, the so-called cost savings are actually cost avoidances where the use of M&S tools permits a program to forego extensive testing, some of which would not be carried out in any event. Examples of large savings reported to have been achieved (or planned for achievement) include $40 million saved by the F-16 program by using M&S to replace flight testing of new avionics, $673 million saved by the Comanche program, and up to 3 percent of life cycle costs saved for the Joint Strike Fighter, which would equate to $5 billion.

In many cases, large savings forecasts have not been realized, even though budgetary and personnel reductions have been levied on both infrastructure and programs based on these unrealistic forecasts. Unfortunately, not even modest up-front investments have been made to attain savings in software, hardware, and personnel. Credible value analyses of the contributions of M&S to the cost, cycle time, and effectiveness of the system to be tested and evaluated need to be carried out to justify these investments and presented to decision makers early in a program's acquisition.

The lack of up-front funding is often a critical problem for M&S.

The Air Force SAB panel states that it "could find little information on the cost of ongoing Digital Modeling & Simulation (DMS) development programs," but did conclude that funding of DMS projects is substantial and it doesn't appear to have any coordinated focus or central control.

The sizeable costs associated with the development and employment of M&S capabilities can be daunting to the program manager who requires such capabilities as an adjunct to physical testing of his weapon system. This often leads to a promised M&S capability not being available when needed for T&E. A good example of this problem is the F-22 Air Combat Man-in-the-Loop simulator, which early on became a linchpin in the Air Force strategy to minimize open air OT&E. It costs around $140 million, and it has been a frequent target during budget drills, resulting in a challenging schedule in danger of not meeting the OT&E requirement.

There is serious concern about the ability of the Department to attract and maintain the intellectual capital required to achieve the potential benefits M&S appears to offer.

The entertainment industry and its pioneering and innovative applications of simulation tools are apparently far more glamorous than government work, and appear to be attracting many of the best and brightest in the field. Clearly, the government has failed to compete with the commercial world in attracting or maintaining the talent necessary to provide a first class M&S development capability. This concern applies to both industry and in-house government agencies responsible for M&S, and affects the long-term prospects for the development and use of a first-rate, credible M&S capability to support decision making.

At present the center-of-gravity for cutting edge advances in computational and simulation technology currently resides in the Department of Energy (DOE) Defense Program Laboratories. Their TERAFLOP computing capabilities provide the potential for large, adaptive simulations involving millions of objects. Yet there appears to be little, if any, significant DoD involvement on a strategic level aimed at the use of these emerging tools.

Recommendations

1. Oversight and Direction of M&S Development and Employment

Assign responsibility to Director, Defense Research and Engineering (DDR&E) (at an appropriately high level) for oversight and direction of all aspects of <u>research</u> relevant to M&S, particularly that which supports T&E.

Responsibilities for oversight of M&S research would fit into three broad categories:

- Phenomenology aimed at measurements and basic data collection to support M&S development and employment across the board, to include T&E applications.
- Conceptual model development to include not only the traditional engineering focus or legacy-based improvements, but with more attention to the insertion of revolutionary computational and simulation technology to handle non-linear conditions and uncertainties (including the impact of human factors, combat behavior, and environmental factors) that influence weapon system effectiveness in realistic combat situations.
- Development and maintenance of a catalog of M&S capabilities and accredited applications, to include categories such as hardware-in-the-loop simulations, man-in-the-loop simulators, physics-based models, and constructive models.

Critical to the exercise of these responsibilities would be a shift from the current emphasis on "code-writing" and simulation connectivity to conceptual model building and basic data collection and use.

Establish an independent accreditation process for M&S use in support of the acquisition process, to include input data for models.

This process is needed to instill much-needed discipline, as well as to increase the level of confidence in the development and application of M&S as an evaluation tool in the acquisition process. For the specific use of M&S tools in operational assessments or system evaluations, the Joint Chiefs of Staff J-8 office should be assigned the responsibility for accreditation of those tools for particular applications. In a similar vein, the DOT&E should accredit input data for use in these models and simulations.

Encourage the use of M&S techniques to improve the test design and planning process, to rehearse tests, and to predict the outcome of test trials. In the latter case, require the modification of models and simulations to reflect the actual outcome of tests when predictions miss the mark.

Require more balanced cost consciousness, to include independent evaluations, in the cost benefit analyses used to justify the development and use of M&S as a means to reduce open air testing.

2. Up-Front Investment in M&S for T&E Applications

Modify the directives governing the acquisition process to require identification and funding of an M&S plan for evaluating a program's progress and operational effectiveness/suitability at the earliest practical point in the program.

Substantial early investment is often needed to capitalize on the potential M&S benefits in the T&E process. A comprehensive M&S plan should be required as part of the Request for Proposal (RFP) process for development of a new system and should be considered as a major evaluation criterion for each contractor's proposal. Approval and commitment of funding for the M&S plan should occur as early as Milestone I, if possible, and certainly by Milestone II at the latest, and should be a necessary condition for proceeding with the program.

Assign funding responsibility for the development of M&S tools required for the evaluation of acquisition programs to a central agency such as the DDR&E agency discussed in the first recommendation above or a Defense Agency that might be assigned the responsibility for the ranges and test facilities.

Typically, program managers lack the incentive to fully fund the M&S tools required for thorough evaluation of their systems and to protect that funding through the life of the program. This situation is analogous to the overall funding of the T&E infrastructure in general as well as system-specific requirements (e.g., targets, unique instrumentation).

3. Identification of Model/Simulation Assumptions and Limitations

Require the users of M&S in their evaluations to state model assumptions, limitations, and uncertainties explicitly, as well as sources for input data in the presentations of results to decision makers.

The disciplined accreditation process called for in the recommendation above would accommodate this problem, which has contributed to the perception that the benefits of M&S are oversold.

4. Near-Term Priority on M&S Development and Use in RAM Evaluations

Give near-term priority, with required funding, to the development of more state-of-the-art reliability, maintainability, and availability models, and to the update of existing models with legacy data from our vast experience with weapon systems in the field.

These efforts should be aimed not only at an ability to credibly evaluate the suitability of systems in development, but at reducing the cost of ownership of both existing and planned systems. The emphasis on reducing life cycle costs of systems should be translated into increased attention to predicting the reliability, maintainability, and availability of systems in development.

TEST AND EVALUATION FACILITY REENGINEERING

Physical T&E of weapons systems requires large, sophisticated facilities, such as those at Edwards Air Force Base, Patuxent River Naval Air Station and Aberdeen Proving Grounds. In the course of this study, the task force reviewed the overall management of these facilities and visited a cross-section of DoD's T&E facilities to assess their readiness to meet current and future defense requirements. The task force views these facilities as indispensable to continued development of new weapons systems; however, it concludes that management of T&E facilities must adapt to the changing test and evaluation environment to retain their vitality and capability. Broadly stated, the task force concludes that the operation of the Defense-wide system of T&E facilities should be carefully rethought in the light of changing defense procurement and management practices, advancing technology, and budget constraints.

OVERALL FACILITY MANAGEMENT

Testing must drive improvement in the management system. Both investment and operating costs can be significantly reduced through management reengineering. Initial savings of the order of $1 billion per year are possible.[1] While this result is welcome, near-term cost of testing savings must not be the primary goal for management changes.

Savings possible from improved testing far exceed what is achievable from concentrating on simple cost reductions of current operations. Testing is second only to design in leveraging early investment to reduce total program cost. Finding problems early through testing and incorporating fixes while in development saves precious cycle time, precludes expensive fixes to fielded systems and provides combat systems to the warfighter much faster. These are the high payoff results of better testing. Once initial designs are completed, there is no better return on investment than funds and time spent doing comprehensive, well-planned testing.

Findings

The Army has recently consolidated its organizations responsible for evaluating results obtained during development and operational testing. This is a possible model for Department of Defense test and evaluation management consolidation.

The Army operational test and evaluation agency is now responsible for independent evaluation of all testing results accomplished by the Army throughout the life of a

[1] Major Range and Test Facility Base (MRTFB) operating budgets for FY '99 are listed in Annex C of this report.

program. The next step in the Army's T&E reengineering process is to create a single management organization for all Army test resources.

> ***Unnecessary test facility duplication among the Services, defense contractors, and other government agencies could be eliminated by centralizing T&E management in a single DoD T&E organization.***

Studies conducted of OSD's T&E organizations have examined alternative management structures. Significant improvements in ownership and utilization of T&E resources can result from:

- centralized financial management and administration
- distributed functional activities
- common strategic resource investment planning
- integrated data acquisition, analysis, archiving, and processing operations
- common research, development, test, and evaluation initiatives
- T&E community-wide human resource management
- sharing of best T&E practices across the Department
- distributed test facility management and test execution.

> ***Outsourcing without reengineering, however, can only reduce the cost of business as usual.***

The task force also considered other possible T&E management models. Contracting out represents nearly 73 percent of the total R&D T&E budget. Only reengineering of the management, ownership, and operation of T&E resources as shown in the DoD studies can provide "new and innovative ways that the T&E community can better support users."

REENGINEERING INDIVIDUAL T&E FACILITIES

The T&E task force believes that a business as usual approach to T&E facility management is insufficient to maintain these critical systems development support facilities to meet future DoD needs. Coping with today's technical and business environment requires more than short-term palliative measures; the management approaches, supporting business systems and test designs must adapt to ensure continued assurance that U.S. weapons systems meet defense needs. The task force concludes that most DoD T&E facilities lack such essential change processes; this short-coming jeopardizes U.S. T&E capability and affordability.

Findings

Physical space required for flight, vehicle, sea, and live-fire testing and facilities with sophisticated or unusual capabilities are a major consideration for T&E facility reengineering.

Facilities for weapon system physical T&E must often be large (e.g., Edwards AFB), expensive, complex (wind tunnels, rocket and jet engine test stands and flight test facilities), sophisticated (well instrumented, supported by high-performance communications and computing systems) and have unusual capabilities (e.g., anechoic chambers, terrestrial and space environment simulators, vehicle test tracks). These aspects must be taken into account in any T&E facility reengineering effort.

Some major test and evaluation facilities are obsolete and inoperable as a result of cannibalization, whereas other heavily used test facilities are limited by time needed for essential repairs.

Many underlying test capabilities witnessed by the task force are old, some constructed during the 50's and using World War II vintage components, such as some of the wind tunnels at Arnold AFB. In response to declining budgets and demands for less costly testing, preventative maintenance for many facilities has been postponed; remediation of acute facility problems now dominates facility upkeep.

Test and evaluation technology is rapidly advancing, propelled by tight test budgets, short T&E schedules and advancements in sensors, communications and computer technology.

Today's programs budget for fewer live fire test vehicles and less flight test time than earlier efforts. High-fidelity modeling and simulation, combined with sophisticated ground testing of reusable vehicles, must replace some field tests. Remote real time, or near real time, control of tests and data evaluation is changing historical relationships between test operators and system developers.

Test and evaluation business practices are changing, forcing profound changes in the way facilities are managed and costs are accounted.

Increased contractor performance test and evaluation, use of government facilities by commercial developers, activity-based accounting, the requirement that system developers pay test costs, and a substantial drop in institutional funding are forcing profound changes in the way facilities are managed and costs are accounted.

DoD has many unique, irreplaceable test and evaluation facilities, some of which, while not required to meet immediate needs, will be needed for future weapons developments. Comparable capability is not available elsewhere, including in the private sector.

Cost-effective ways must be found to deactivate these unique, irreplaceable facilities in ways that preserve their capabilities to meet future needs.

Duplicate T&E capabilities exist among the Services and between the Department and non-DoD organizations.

Test and evaluation facilities within DoD were developed in response to Service needs. As a result, duplicate T&E capabilities exist among the Services and between the Department and non-DoD organizations such as NASA. Examples of this are seen at Patuxent River and Edwards AFB. Many of these facilities appeal to the same outside customers for business.

The task force found some test and evaluation organizations that are effectively addressing changes in the T&E environment and technology.

Arnold Engineering Development Center's (AEDC) reengineering program is one notable example. The task force examined the Center's process of managing operations to meet the shifting T&E environment and followed up with a site visit to learn more. The task force found a sustained, center-wide, top-down driven activity focused on adapting AEDC operations to long-term trends in support, project, and facilities funding and personnel strength. This activity, closely patterned on industrial reengineering successes, places heavy emphasis on measures of cost-effectiveness, response, activity-based costing, and activity-based management. Continuous process improvement, (for both businesses and T&E processes), qualified by metrics, is central to this reengineering program. This commitment to reengineering, which has persisted throughout the tenures if several AEDC commanders, resulted in a 50% reduction in allocated costs in the first year, with additional significant savings coming from improved logistics support and business processes in subsequent years. This has been accomplished while sustaining the center's missions cost structure, organization, and T&E processes.

Such successful efforts have resulted from a thorough rethinking of the test and evaluation process and shifts in business pressures. The need to compete more strongly for test and evaluation customers through cost, cycle time, service and capabilities drives these long-term transformations to ensure organizational survival.

Recommendations

1. **The task force was briefed on concepts for bringing together the management of T&E resources. OSD and the Services should work together to develop a plan whereby T&E resource management is strengthened and brought under coherent control.**

This control organization should be responsible to the Secretary of Defense for planning, budgeting, operations and maintenance of the Department's T&E resources.

2. **Military departments responsible for acquiring specific weapons systems should continue to test those systems and be responsible for evaluating and reporting results of these tests.**

This new management organization should make every effort to outsource activities that can be performed in the private sector. Decisions concerning outsourcing and facility conversion to non-government operation should be made first, on the basis of fielding the best possible weapons systems to warfighters in the most expeditious manner; second, to exploit the best applicable business practices to achieve better, faster, and cheaper test and evaluation; and third, to ensure the preservation of unique facilities needed to meet future defense test and evaluation needs.

3. **Apply the reengineering process, employed at Arnold Engineering and Development Center (AEDC), to all test ranges in order to update operations and maintenance and capability investment plans and strategy.**

DoD should undertake a thorough inventory of the Department's T&E facilities and skills and those of other government and non-governmental organizations similar to that recently completed for radar cross section measurement facilities[2]. This inventory should look beyond DoD capabilities to take into account those facilities elsewhere in government and in the private sector. The inventory should clearly identify DoD facilities needed to meet future near- and long-term needs, both for test and evaluation capacity and for unique capabilities.

4. **DoD should develop a T&E Facility investment strategy, based on the inventory and future needs, to assure ability to meet DoD T&E needs through the most effective and efficient combination of all national facilities.**

Investment alternatives should be formulated using scenario- or statistics-based modeling tools to provide the basis for decisions. Options including increased use of contractors to conduct operations should be considered.

5. **DoD should establish a set of figures of merit to assess, compare, and contrast reengineering progress of each of its major T&E facilities.**

The test and evaluation community as a whole (including both DoD and contractor organizations) must develop more efficient test procedures, based on computer-based modeling and simulation, that achieve more insight into system performance from ground-based tests. The virtual test project at Arnold Engineering Development Center is one good example of creative ways to improve the fidelity and coverage of wind tunnel testing.

Test and evaluation is no longer limited to the later stages of system development and operational deployment. Future test and evaluation planning must reach back to early systems development process. The complexity of modern systems, especially those based on Very Large Scale Integration (VLSI) electronics, are impractical to thoroughly test from outside the system. Design for Test and Evaluation and Built-In Self Test and Evaluation must become a routine part of DoD systems development to assure developmental and operational T&E can be adequately performed.

---

[2] National Test Facility Advisory Council, Radar Cross Section (RCS) Measurement Facilities Assessment, December 1998.

6. **A searching examination of management practices should be undertaken by all test and evaluation centers to address changing test and evaluation technical and business practices, aging facilities, cost and test cycle time reduction, and customer service.**

The result of this examination should be a sustained reengineering effort that is disciplined and enterprise-wide. Coping with the future test and evaluation environment requires new thinking about business processes, not merely new computer or business programs. Experience shows that such a change will require dedicated leadership and resources; outside help from groups with proven records of success in planning and facilitating similar changes elsewhere will be needed.

The critical spiral of diversion of scarce preventative maintenance resources to solve short-term operating problems while delaying routine maintenance must be broken. Once started, this spiral leads to operations where all problems are emergencies. The result is inevitably longer test and evaluation schedules and compromised capabilities.

Sustained leadership dedicated to achieving real change is essential to success in the kind of major shift in business and technical practices envisioned by the Panel. Leadership must achieve organization-wide buy-in and enthusiasm for the change. The time required will last throughout the terms of several facility military commanders.

Test and evaluation reengineering will not succeed without incentives to facility management to achieve dramatic long-term results. The current financial system actively penalizes improvements by decrementing operating budgets based on forecasted savings at the outset of reengineering. In the near term, facilities should be allowed to reinvest a portion of savings achieved to promote still further test and evaluation improvements.

7. **This study limited itself to DoD owned and operated facilities but would observe that a serious look at the T&E facilities should include NASA, DOE, and other agencies test facilities.**

SPECIAL ISSUES

This special issues section includes discussion and recommendations on topics that are of special interest to the Department and include T&E involvement in advanced concept technology demonstrations (ACTDs), T&E of systems of systems, T&E of information system software, and T&E of commercial off-the-shelf (COTS) software.

To address these issues the task force looked at how each of these areas can benefit from T&E. The task force examined how the transition of ACTDs can be improved by adding a tester to the process and how T&E should be conducted on military systems that are actually systems of systems. Lastly, T&E of software was studied, to include both information systems and COTS software.

ADVANCED CONCEPT TECHNOLOGY DEMONSTRATIONS (ACTDS)

The Advanced Concept Technology Demonstration (ACTD) Program, initiated in 1994, supports a set of the department's programs aimed at quickly capturing technology and getting it into the hands of the warfighter. Through 1998, nine ACTDs had been completed, including the Predator medium altitude Unmanned Aerial Vehicle (UAV), which saw operational use in Bosnia and Kosovo, and the Precision/Rapid Counter Multiple Rocket Launcher System in Korea. Currently, just over 50 ACTDs are underway, including 11 accepted for FY99.

ACTDs are designed to directly foster alliances between technologists and warfighters. In particular, the technical community gets feedback from the warfighter, while the warfighter gets to examine new technologies and their attendant operational concepts without an upfront commitment to buy. This provides for the possibility of introducing technologies, especially for joint capabilities or other "out-of-the-box" programs, that might otherwise not get examined. The requirement of JROC review and CINC sponsorship favors capabilities that have a real, near-term mission to perform.

The intent of the ACTD programs, if successful, is to create and "leave behind" some operational capability when the program concludes in a two- to four-year time period. An operational assessment based on effectiveness in an anticipated operational environment with expected threats must accompany a "leave behind" operational capability. At this point, it should be possible to decide whether or not to convert the effort into an acquisition program to equip more of the force.

The most useful role for the T&E community in the ACTD program is helping with transition to acquisition status. "Leave behind" products of ACTDs should have adequate operational testing against threats to understand the operational effectiveness before fielding. CINCs must know both the capabilities and limitations for combat employment. The T&E community can help by aiming the demonstration and the data collected at a particular milestone. For example, a demonstration that was intended to proceed directly to production would require resolving issues ordinarily settled in a dedicated operational test, whereas a demonstration aimed at MS II insertion would have much more modest learning requirements. In addition, in the spirit of "early involvement" the T&E community can work with and archive the data to minimize the redundancy of information gathered in formal testing later on.

T&E participation has provided early operational insights that have influenced the direction taken by the demonstration. In the case of the Medium Endurance Unmanned Aerial Vehicle (Predator), the testers identified lack of system reliability as a driver preventing the continuous coverage that would become one of the Key Performance Perameters.

However, Predator was one of the first ACTDs, and the testers were added to the demonstration late in the process. Inclusion of testers at the beginning of an ACTD is becoming more common. This will go a long way toward easing transition pain. The adoption of the CAF concept will be the remedy for better integration of testers in ACTDs as well as in the standard acquisition process.

Recommendation

> ***Direct the participation of a tester in the ACTD process to maximize the value of this concept.***

SYSTEMS OF SYSTEMS T&E

The integration of information systems from Services and Agencies, many of which are not well integrated themselves, into joint systems which can effectively serve the CINCs in joint operations presents difficult T&E challenges. In most instances there is no overall architect of the joint system although efforts are made to cope with the resulting lack of overall concepts. Since the joint systems of systems problem is so large and complex and since it involves so many organizations and so many components of all ages and origins it is difficult to exercise control over the systems of systems design and test process.

The term "Systems of systems" means different things to different people. It covers anything from a simple component integration to a land, air, sea and space multi-platform, multi-service joint campaign-level interdependent system. For this report, systems of systems are defined as those systems that require interoperability or interdependency between functional entities for mission success. In nearly all cases, the systems being integrated do not have a single responsible design agent and extensive coordination is required across Service and agency lines.

There are three classes of components:

- Legacy systems
- New systems
- Interoperability/interdependency enabling systems

The joint systems of systems must work. Yet, very large systems tend to evolve; they are not created in a single action. The usual technique is to assemble the system and run it under realistic conditions, learn how to use it, and find out what doesn't work and fix it. This is how very complex civil systems come into being.

Systems of systems T&E differs from the form of T&E where the tests are run to determine whether the system meets specifications. Instead, the idea is to make the best of the situation, to find out how to use what exists, to fix what is possible to fix, and to encourage suppliers to create new and useful components as well as to improve old components. The emphasis is on evaluating the fixes.

Joint learn-and-fix tests and exercises, both real-world and simulated, place new demands on the DoD's T&E resources. Systems of systems testers must have comprehensive knowledge of all the systems of systems components in the joint environment. The tester must have a truly joint perspective. Further, the systems of systems tester must be able to mount a test environment that accurately reflects joint operations. The environment must be able to support interoperability/interdependency testing.

Historically, interoperability/interdependency testing has been optional. Yet system interoperability weaknesses are often first uncovered during operational testing, when component systems are operated together in a field test. Solving problems thus discovered are key to operational mission success. Operating evaluations and innovative work around procedures can overcome many interoperability deficiencies. Today's shift to interdependency will create new systems that have limited autonomous capability, thereby introducing risk to the potential success of these systems and will lead to "dependent" systems that cannot be operationally tested without the presence of the integrated "resource" systems.

The rapidly improving technology that is associated with the systems that provide the "information linking" must deal with both new systems having comparable technology and legacy systems that are often comprised of dated technology

DoD has an organization devoted to the T&E of joint systems of systems. The Joint Interoperability Test Command (JITC), is tasked with assuring joint interoperability, yet JITC involvement in systems of systems development is optional. It has no funds of it's own, Program Managers must bring cross-service budget agreements with them when they enlist JITC's' help, and thus often skip T&E aimed at interoperability issues.

Findings

There is no clear, common, definition of "systems of systems."

Many of the briefings presented to the DSB T&E task force were, in the eyes of the briefer, "systems of systems" programs. Representatives from DOT&E, the Military Services OT&E agencies organizations responsible for modeling and simulation and testing of joint systems and interoperability, presented many views of systems of systems.

There is no single responsible entity that has both the financial and performance responsibility for the operational success of the particular system of systems.

Joint Interoperability Test Command (JITC) should be assigned responsibility for assuring systems of systems interoperability and be provided limited program funds to carry out this mission.

Recommendation

Create a systems of systems design authority (system designer) with a joint T&E component that uses the Service T&E and JITC resources for joint testing.

Joint systems of systems testing will likely be difficult, as it is not well defined.

INFORMATION SYSTEMS SOFTWARE T&E

Information systems are frequently systems of systems, incorporating new as well as legacy hardware and software and lacking an overall systems architect; they often share many of the systems of systems T&E challenges indicated in the previous section. The organization of development into separate hardware and software design activities often gets in the way of achieving efficient marriage of these two major systems aspects.

A well-disciplined engineering process is essential for designing, building and testing information systems to deal with system complexities.

Test and evaluation considerations must be part of the initial system development plan and must be an important consideration at every step of the process. Serious problems not found by conventional T&E until late in the development process can be totally disruptive and compromise the entire development.

Although T&E of the integrated hardware and software components of information systems has many challenges, software is a major T&E problem in its own right. Postponing software T&E until late in the development process wastes valuable opportunities to uncover problems at periodic system "builds" and to correct them easily early in the development.

Modular software design that allows incremental testing of pieces of the system as the development proceeds is essential. Even when modules are validated individually, integration testing can uncover major flaws, but at least the elements of the system are understood. Sound software system architecture and rigorous configuration control are essential to assure that interface problems can be identified and tested early. Each module should have built-in test capabilities to assist in diagnosing T&E and operating problems.

Software testing must start early in the program to uncover problems as they arise. Software "inspections" and peer reviews, where a small, independent team of software and applications experts reviews the design, the developed code and documentation are important. Such reviews detect existing flaws and potential problems as well, saving time and money later in the development process. Development organizations themselves, not separate organizations, are responsible for carrying out these early T&E activities.

Effective software design and development hinges on having a competent, experienced software team (contractor or government). The proven Capability Maturity Model (CMM) helps to assess a software organization's capabilities. The process involves a team of 5 or 6 people lead by a certified expert in software appraisal that spends 5- 10 days performing a CMM assessment. A successful assessment qualifies the ability of the organization to carry out disciplined software development and to track software performance to specifications.

Late in the software development process, commercial companies employ alpha and beta testing of their products by sample users to uncover unanticipated problems and assure a robust product at introduction to the general market. DoD should study the private sector practices and employ them in the development of military information system software to the maximum extent possible.

Automated testing of non real-time software that automatically identifies design flaws in developmental software is showing some promise. Automated test technology should be developed and employed by software testers to the maximum extent possible.

Recommendations

1. **Insist on Capability Maturity Model qualification of software development teams appropriate to the size and complexity of the information system.**

2. **Employ alpha and beta testing, as practiced in the private sector, to the fullest extent possible.**

3. **Consider automated testing of non real-time software; ask DARPA to consider developing machine testing technology for both real-time and non real-time software.**

4. **Insist that regular, independent software "inspections" be routinely incorporated into information system development programs.**

TEST AND EVALUATION OF COMMERCIAL OFF-THE-SHELF SOFTWARE (COTS)

The functional requirements for systems in the commercial world have moved closer to the requirements for military systems. For example, both electronic commerce and military systems require network security. This has led to more extensive use of COTS products, including hardware and software, in military systems. The DoD, in policy and practice, has stated a strong preference for COTS products and for a reduction in custom-built software and hardware. However, the use of COTS products can significantly change the process by which systems are built. For example, the system must be designed to use existing COTS products even if they are merely satisfactory and not optimum for the military purpose. One expectation is that the COTS product has been thoroughly tested and proven and that its characteristics have been demonstrated in commercial applications, therefore, testing for military applications can be greatly reduced. This is a reasonable assumption for individual software packages but is less so for systems of packages. In general, it seems likely that the use of COTS software in military systems will not have a major effect on reducing the effort required for system testing, in either DT or OT.

Except in trivial cases (such as a word processing or e-mail system), military systems include many software packages. Even if these are COTS products, they will probably come from different vendors and may use different languages and interfaces and other standards. Even if the individual packages are carefully chosen by the system designers for performance and quality, integration of the software into a working whole may not be easy and may require building substantial amounts of interface software, drivers, etc. This is true for commercial systems as well.

Presumably the COTS software will already have been extensively tested by commercial users and should work well as long as the military application is similar to the usual commercial application and is well within the performance limits of the package. In some cases, however, the military application may stress the COTS software in unusual ways; a problem may not appear until the system has been completely assembled and tested under realistic conditions. The integrating contractor may find that his lack of knowledge of the COTS details, the lack of source code, and the limited leverage he will have on the vendor to make changes and corrections will offset the relative lack of bugs in the COTS software. These problems also exist in system level testing where discovering which package is at fault may be difficult.

COTS software can have many advantages including savings in time and money, fewer bugs, and built-in vendor support and upgrades. At the same time, the frequent new versions of COTS software will create new integration problems. Existing software in the system may have to be changed because the vendors no longer support the old versions while the new versions may require related upgrades in hardware and operating systems which will affect the other packages. Software maintenance is a major problem with all software systems, but maintenance of COTS systems presents somewhat different, but not necessarily easier, problem than maintenance of custom systems.

Recommendation

Conduct functional and interface testing of COTS software as a prerequisite to its use in DoD programs.

Summary of Recommendations

Test and Evaluation Requirements Determination Process

Recommendations

1. **Establish a stable team made up of users, developers, testers, and appropriate contractors called a Combined Acquisition Force (CAF) to streamline the acquisition process for ACAT I programs. The CAF should be formed once a need is identified and remain in place throughout the acquisition process.**

2. **Change 5000.1 to incorporate the CAF concept.**

3. **Require the development of a Preliminary Test and Evaluation Plan in conjunction with the MNS.**

4. **Conduct an early operational assessment based on the ORD and an operational scenario before milestone I (no actual hardware would be involved).**

5. **Make the CAF a key participant in the Requirements Developments Process.**

Developmental and Operational Test Process

Recommendations

1. **Each of the Service DT&OT organizations should be consolidated, to include integrated planning, use of models, simulation, and data reduction. Planning should be totally integrated, and the OSD T&E organizations consolidated. There should be integrated use of models, simulation, and data reduction. Except for limited dedicated OT&E, contractor and government testing should also be integrated.**

2. **For each development program and its associated test and evaluation effort, special attention should be directed early in the planning cycle (and periodically throughout program development) toward compressing the developmental test schedule wherever practical.**

3. **Facilities within the DoD's Major Range and Test Facility Base should be required to conduct periodic, systematic reviews to determine where data acquisition, reduction, and analysis procedures could be improved to increase the efficiency of the T&E process.**

Modeling and Simulation in Support of the Test and Evaluation Process

Recommendations

1. Oversight and Direction of M&S Development and Employment

Assign responsibility to DDR&E (at an appropriately high level) for oversight and direction of all aspects of research relevant to M&S, particularly that which supports T&E.

Establish an independent accreditation process for M&S use in support of the acquisition process, to include input data for models.

Encourage the use of M&S techniques to improve the test design and planning process, to rehearse tests, and to predict the outcome of test trials. In the latter case, require the modification of models and simulations to reflect the actual outcome of tests when predictions miss the mark.

Require more balanced cost consciousness, to include independent evaluations, in the cost benefit analyses used to justify the development and use of M&S as a means to reduce open air testing.

2. Up-Front Investment in M&S for T&E Applications

Modify the directives governing the acquisition process to require identification and funding of an M&S plan for evaluating a program's progress and operational effectiveness/suitability at the earliest practical point in the program.

Assign funding responsibility for the development of M&S tools required for the evaluation of acquisition programs to a central agency such as the DDR&E agency discussed in the first recommendation above or a Defense Agency that might be assigned the responsibility for the ranges and test facilities.

3. Identification of Model/Simulation Assumptions and Limitations

Require the users of M&S in their evaluations to state model assumptions, limitations, and uncertainties explicitly, as well as sources for input data in the presentations of results to decision makers.

4. **Near-Term Priority on M&S Development and Use in RAM Evaluations**

   Give near-term priority, with required funding, to the development of more state-of-the-art reliability, maintainability, and availability models, and to the update of existing models with legacy data from our vast experience with weapon systems in the field.

TEST AND EVALUATION FACILITY REENGINEERING

Recommendations

1. **The task force was briefed on concepts for bringing together the management of T&E resources. OSD and the Services should work together to develop a plan whereby T&E resource management is strengthened and brought under coherent control.**

2. **Military departments responsible for acquiring specific weapons systems should continue to test those systems and be responsible for evaluating and reporting results of these tests.**

3. **Apply the reengineering process, employed at Arnold Engineering and Development Center (AEDC), to all test ranges in order to update operations and maintenance and capability investment plans and strategy.**

4. **DoD should develop a T&E Facility investment strategy, based on the inventory and future needs, to assure ability to meet DoD T&E needs through the most effective and efficient combination of all national facilities.**

5. **DoD should establish a set of figures of merit to assess, compare, and contrast the reengineering progress of each of its major T&E facilities.**

6. **A searching examination of management practices should be undertaken by all test and evaluation centers to address changing test and evaluation technical and business practices, aging facilities, cost and test cycle time reduction, and customer service.**

7. **This study limited itself to DoD owned and operated facilities but would observe that a serious look at the T&E facilities should include NASA, DOE, and other agencies test facilities.**

ANNEX A: TERMS OF REFERENCE

THE UNDER SECRETARY OF DEFENSE
3010 DEFENSE PENTAGON
WASHINGTON, D.C. 20301-3010

UISITION AND
TECHNOLOGY

18 MAY 1998

MEMORANDUM FOR CHAIRMAN, DEFENSE SCIENCE BOARD

SUBJECT: Terms of Reference--Defense Science Board Task Force on Test and Evaluation

You are requested to form a Defense Science Board (DSB) Task Force to undertake a broad review of the entire range of activities relating to Test and Evaluation (T&E). The T&E structure of the Department of Defense has been built over a number of years, and today two-thirds of the T&E structure of the Department is over 30 years old. While some new T&E capability has been added, generally T&E did not share in the buildup in the 1980s when both procurement and Defense R&D increased considerably. Now under the guidance of the QDR, procurement is planned to increase 50% by 2003 from its low point in 1996. These procurements encompass major systems as large as any in the history of the Department as well as small upgrades, commercial, and nondevelopmental items. The costs and cycle times of all these defense acquisitions can be reduced, and the T&E structure cannot only be responsive to all these needs, it can help lead the way to positive change. Further, a number of new acquisition techniques, such as the ACTD, have formalized the department's increasing emphasis on early and extensive involvement of the warfighter via CINC participation in acquisition activities. Inclusion of the warfighter as a user will effect future T&E, in part by providing the opportunity to address operational issues and gain operational insights early in the development process.

Your study should include an examination of new and innovative ways that the T&E community can better support these users. For example, the variety and everchanging nature of the threat systems which U.S. and coalition forces face calls for a new look at the role of T&E in rapid and responsive acquisition cycles to meet these threats. There are a number of test processes that merit thoughtful review in light of these new demands and the new technologies with which they are associated. Major advances can be realized by applying selected industrial principles and practices to operational testing and the associated information gathering and evaluation process in the development of military systems. In addition, it is desirable to find new ways to integrate operational testing into the overall system development process to provide as much information as possible as soon as possible on operational effectiveness and suitability so that improvements to the system and decisions about continuing system development or passing to full-rate production can be made in a timely manner.

In view of the above, the DSB Task Force review should be comprehensive, including facilities, processes, policies, contractor testing, government development testing, joint test and evaluation, and operational test and evaluation, both before and after full production and in the continuing evolution of force capability. Further, the Task Force should consider the special problems associated with Test and Evaluation of the "Systems of Systems" which are comprising increasingly critical parts of our military capability. From the information gathered the Task Force should, first, identify and quantify the current and future needs of the Department's T&E capabilities and resources and, second, should recommend specific and quantified changes to those capabilities and resources.

In structuring this Task Force, it is recommended that the DSB bring to bear individuals experienced in T&E, knowledgeable of the weapons system development process as it is evolving, and individuals familiar with the Congressional initiatives which have formed much of the T&E structure over the past years.

This study should provide an interim report to the sponsors by November 1998 and complete its work by April 1999.

The Under Secretary of Defense (Acquisition and Technology) and the Director, Operational Test and Evaluation will co-sponsor this Task Force. Mr. David Heebner will serve as Chairman of the Task Force. Mr. Bill Meyer from the office of the Director, Test Systems Engineering and Evaluation, USD(A&T), and Dr. Dave Sparrow from the office of the Director, Operational Test and Evaluation will serve as co-Executive Secretaries. CDR David Norris, USN, will be the Defense Science Board Secretariat Representative.

The Task Force will operate in accordance with the provisions of P.L. 92-463, the "Federal Advisory Committee Act," and DoD Directive 5105.4, the "DoD Federal Advisory Committee Management Program." It is not anticipated that this Task Force will need to go into any "particular matters" within the meaning of Section 208 of Title 18, U.S. Code, nor will it cause any member to be placed in the position of acting as a procurement official.

J. S. Gansler

ANNEX B: TASK FORCE MEMBERSHIP

CHAIRMAN

Mr. David R. Heebner

MEMBERS

Mr. Charles E. Adolph
Mr. Thomas P. Christie
Mr. Robert R. Everett
Mr. Donald N. Fredericksen
Dr. William G. Howard, Jr.
The Honorable John E. Krings
Mr. Frank Marchilena
Mr. Ken Morrelly
GEN Glenn K. Otis, USA (Ret)
Maj. Gen. Cecil W. Powell, USAF (Ret)
MG Richard W. Tragemann, USA (Ret)
RADM Jack Zerr, USN (Ret)

EXECUTIVE SECRETARIES

Dr. David A. Sparrow
Mr. Bill Meyer

ANNEX C: MAJOR RANGE & TEST FACILITY BASE (MRTFB) OPERATIONS FY99

ALL INSTALLATIONS
SOURCE: JUNE 99 MRTFB
($000)

MRTFB	INSTITUTIONAL*	FUNDING USER**	TOTAL
Aberdeen Test Center (ATC), Aberdeen Proving Ground, MD	29,356	70,680	100,036
Dugway Proving Ground (DPG), UT	9,983	22,450	32,433
High Energy Laser Systems Test Facility (HELSTF), NM	14,442	286	14,728
U.S. Army Kwajalein Atoll/Kwajalein Missile Range (KMR)	53,153	70,821	123,974
White Sands Missile Range (WSMR), NM including EPG at Ft Huachuca, AZ	57,143	151,461	208,604
Yuma Proving Ground (YPG), AZ including Tropic and Arctic test centers	19,022	58,538	77,560
Sub Total	**183,099**	**374,236**	**557,335**
Atlantic Fleet Weapons Training Facility (AFWTF), Roosevelt Roads, PR	20,533	3,891	24,424
Atlantic Undersea Test and Evaluation Center (AUTEC), Andros Islands, BA	46,134	16,453	62,587
Naval Air Warfare Center-Aircraft Division (NAWC-AD), Patuxent River, MD	71,074	115,790	186,864
Naval Air Warfare Center- Weapons Division (NAWC-WD) Pt Mugu & China Lake	119,931	109,964	229,895
Sub Total	**257,672**	**246,098**	**503,770**
30TH Space Wing, Vandenberg AFB, CA	59,446	63,218	122,664
45TH Space Wing, Patrick AFB, FL	137,048	168,605	305,653
46TH Test Group at Holloman AFB, NM	18,717	34,234	52,951
Arnold Engineering Development Center (AEDC), Arnold AFS, TN	99,053	114,030	213,083
Air Armament Center (AAC), Eglin AFB, FL	61,357	211,456	272,813
Air Force Flight Test Center (AFFTC), Edwards AFB, CA	104,745	232,430	337,175
Air Force Air Warfare Center (AWC) at Nellis Range Complex, NV	37,800	15,200	53,000
Utah Test and Training Range (UTTR), Hill AFB, UT	17,584	8,475	26,059
Sub Total	**535,750**	**847,648**	**1,383,398**
Joint Interoperability Test Command (JITC), Defense Information Systems Agency Activity	7,549	72,110	79,659
GRAND TOTAL	**984,070**	**1,540,092**	**2,524,162**

* Institutional funds to cover indirect cost of Test & Evaluation support.

** User funds charged to the customers for direct cost

ANNEX D: GLOSSARY

ACAT	Acquisition Category
ACETEF	Air Combat Environment Test and Evaluation Facility
ACOM	U.S. Atlantic Command
ACTD	Advanced Concept Technology Demonstration
AEDC	Arnold Engineering and Development Center
BRAWLER	Air Combat Simulation (not an acronym)
CAF	Combined Acquisition Force
CAIV	Cost as An Independent Variable
CFD	Computational Fluid Dynamics
CINC	Commander-in-Chief
CMM	Capability Maturity Model
CMMS	Conceptual Model of Mission Space
COEA	Cost and Operational Effectiveness Analysis
CONOPS	Concept of Operations
COTS	Commercial Off-The-Shelf
DARPA	Defense Advanced Research Projects Agency
DDR&E	Director, Defense Research & Engineering
DMS	Digital Modeling and Simulation
DMSO	Defense Modeling and Simulation Organization
DoD	Department of Defense
DOE	Department of Energy
DOT&E	Director, Operational Test and Evaluation
DSB	Defense Science Board
DT	Developmental Test
DT&E	Developmental Test & Evaluation
EMD	Engineering and Manufacturing Development
ESAMS	Enhanced Surface-to-Air Missile Simulation
FFRDC	Federally Funded Research and Development Center
FTTC	Flight Test Technical Committee
HLA	High Level Architecture

HWIL	Hardware-in-the-Loop
ICT	Integrated Concept Team
IDA	Institute for Defense Analysis
IOT&E	Initial Operational Test & Evaluation
IPT	Integrated Product Team
ITF	Integration Task Force
JADS	Joint Advanced Distributed Simulation
JASA	Joint Accreditation Support Activity
JCS	Joint Chiefs of Staff
JMASS	Joint Modeling and Simulation System
JT&E	Joint Test & Evaluation
JTIC	Joint Interoperability Test Command
LCOM	Logistics Composite Model
LFT&E	Live Fire Test & Evaluation
LRIP	Low Rate Initial Production
M&S	Modeling & Simulation
MICOM	U.S. Army Missile Command
MNS	Mission Needs Statement
MSRR	Modeling and Simulation Resource Repository
NMD	National Missile Defense
ORD	Operational Requirements Document
OSD	Office of the Secretary of Defense
OT	Operational Test
OT&E	Operational Test & Evaluation
PMO	Program Management Office
QDR	Quadrennial Defense Review
RAM	Reliability, Availability, and Maintainability
RCS	Radar Cross Section
RFP	Request for Proposal
RFSS	Radio Frequency Simulation System
SUPRESSOR	An air, ground, naval and space combat simulation (not and acronym)

SAB	Scientific Advisory Board
SBIRS	Space-Based Infra-Red Sensor
SISO	Simulation Interoperability Standards Organization
SPO	Sponsored Projects Office
SQuASH	Stochastic Quantitative Analysis of System Hierarchies
STOW	Synthetic Theater of War
SURVIAC	Survivability/Vulnerability Information Analysis Center
T&E	Test & Evaluation
TACWAR	Tactical Warfare Model
TEMP	Test and Evaluation Master Plan
THAAD	Theater High-Altitude Area Defense
TRADOC	U.S. Army Training and Doctrine Command
TRANSIMS	Transportation Analysis Simulation System
TRAP	Trajectory Analysis Program
UAV	Unmanned Aerial Vehicle
USACOM	U.S. Atlantic Command
USD(A&T)	Under Secretary of Defense (Acquisition & Technology)
VLSI	Very Large Scale Integration
VV&A	Verification, Validation and Accreditation