Report of the
Defense Science Board Task Force
on
Developmental Test & Evaluation

May 2008

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

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

DEFENSE SCIENCE BOARD

27 May 2008

MEMORANDUM FOR UNDER SECRETARY OF DEFENSE FOR ACQUISITION, TECHNOLOGY & LOGISTICS

SUBJECT: Final Report of the Defense Science Board (DSB) Task Force on Developmental Test and Evaluation (DT&E)

I am pleased to forward the final report of the DSB Task Force on Developmental Test and Evaluation, chaired by Mr. Charles "Pete" Adolph.

As requested in the Terms of Reference the Task Force was asked to assess, from a Test and Evaluation (T&E) perspective, OSD organizational roles and responsibilities; policy and practices in oversight of acquisition programs; assess changes required to establish statutory authority for OSD DT&E oversight; and assess Initial Operational Test and Evaluation (IOT&E) failures due to lack of Operational Suitability.

The final report addresses the taskings in the Terms of Reference and provides findings and recommendations addressing broader programmatic issues stemming from systemic changes to the acquisition process. The report also presents findings and recommendations on program structure, requirements definition, contractual performance requirements, alignment of DoD terminology with systems engineering procedures, Commercial Off-The-Shelf (COTS) products, and Systems of Systems (SoS).

I endorse the Task Force's findings and recommendations and encourage you to review the report.

Dr. William Schneider, Jr.
DSB 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 (DSB) Task Force on Developmental Test and Evaluation (DT&E)

The report of the Defense Science Board Task Force on Developmental Test and Evaluation is attached. The Task Force examined, from a Test and Evaluation (T&E) perspective, Office of the Secretary of Defense (OSD) and Service organizational roles and responsibilities, and policy and practices for oversight of acquisition programs. The Task Force was asked to recommend changes that may contribute to increasing the number of programs undergoing Initial Operational Test and Evaluation (IOT&E) being evaluated as Operationally Effective and Operationally Suitable. An additional objective was to recommend changes that would result in quicker delivery of improved capability and sustainability to Warfighters.

Early in the study, it became obvious that the high suitability failure rates were the result of systemic changes that had been made to the acquisition process; and that changes in developmental test and evaluation alone could not remedy poor program formulation. Accordingly, the Task Force study was expanded to address broader programmatic issues, in addition to the tasking in the Terms of Reference (TOR).

The Task Force principal findings include:

- The high suitability failure rates were caused by the lack of a disciplined systems engineering process, including a robust reliability growth program, during system development
- Sequential workforce cuts in the last ten years had a significant adverse impact on the DoD acquisition capability
- Acquisition personnel reductions combined with loss of guidance documents and retirement of experienced senior industry and government personnel have exacerbated the adverse impact
- Strong OSD and Service leadership commitment is vital to solving the major acquisition problems which include widespread suitability deficiencies
- The implementation of Acquisition Reform provided flexibility but, when combined with an eroding workforce, sometimes resulted in less discipline in program formulation and execution
- DT&E needs improvement but changes in test processes will not remedy systemic program formulation and execution deficiencies

- Reliability, Availability, and Maintainability (RAM) shortfalls are frequently identified during DT, but program constraints (schedule and funding) often preclude incorporating fixes and delaying IOT&E
- Additional emphasis on integrated testing will improve T&E process efficiency as well as allow for program cost reductions

Based on the findings, the Task Force outlined recommendations in the report which address these issues. The single most important step necessary to correct high suitability failure rates is to ensure programs are formulated to execute a viable systems engineering strategy from the beginning, including a robust RAM program, which includes reliability growth, as an integral part of design and development. No amount of testing will compensate for deficiencies in RAM program formulation.

Several other issues were addressed as part of the study. A discussion of each of the following topics, along with findings and recommendations, may be found in the report.

- Program Structure
- Requirements Definition
- Contractual Performance Requirements
- Alignment of DoD Terminology with Systems Engineering Procedures
- Commercial Off-The-Shelf (COTS)
- Systems of Systems (SoS)

The Task Force urges senior leaders in the U.S. government to implement the recommendations in this report at the earliest opportunity.

On a personal note, I would like to express my appreciation to the Task Force members, executive secretaries, government advisors and support staff for their superb efforts in support of this study. I would also like to convey my thanks to the Defense Science Board peer reviewers, Dr. Paul Kaminski and Dr. George Heilmeier, for their many constructive comments and suggestions to improve the report.

Pete Adolph

Charles "Pete" Adolph
Task Force Chairman

TABLE OF CONTENTS

I. Executive Summary ... 1

II. Introduction ... 13

III. Problem Definition ... 13

IV. Background ... 15

A. Major Changes in the Last 15 Years ... 15
B. Recent Army T&E changes ... 16
C. Recent Navy T&E changes ... 17
D. Recent Air Force T&E changes ... 19

V. Principal findings and recommendations ... 21

A. Reliability, Availability, and Maintainability (RAM) ... 21
B. Roles and Responsibilities of Government Test and Evaluation Organizations ... 24
C. Integrated Test and Evaluation ... 28
D. Operational Test Readiness Review (OTRR) ... 31
i. Army Operational Test Readiness Review (OTRR) ... 31
ii. Navy Operational Test Readiness Review (OTRR) ... 32
iii. Air Force Operational Test Readiness Review (OTRR) ... 32
E. OSD Test and Evaluation Organization ... 33

VI. Other Issues ... 38

A. Program Structure ... 38
B. Requirements Definition ... 38
C. Contractual Performance Requirements ... 39
D. Alignment of DoD Terminology with Systems Engineering Process ... 40
E. Commercial Off-The-Shelf (COTS) ... 41
F. Systems of Systems ... 43

Appendices:
Appendix A: Terms of Reference ... 44
Appendix B: Task Force Members & Government Advisors ... 46
Appendix C: Briefings Received ... 47
Appendix D: Test and Evaluation Policy Revisions Memo ... 50
Appendix E: RAM Definitions ... 52
Appendix F: OUSD (A) Organizational Chart (April 1992) ... 53
Appendix G: OUSD (AT&L) Organizational Chart (June 2007) ... 54
Appendix H: Retirement Eligible Major Range and Test Facility Base Civilians ... 54
Appendix H: Retirement Eligible Major Range and Test Facility Base Civilians ... 55
Appendix I: OSD CAIG Life Cycle Costs ... 56
Appendix J: Acronym List ... 57

I. EXECUTIVE SUMMARY

A Defense Science Board (DSB) Task Force on Developmental Test and Evaluation (DT&E) was convened in the summer of 2007 to investigate the causal factors for the high percentage of programs entering Initial Operational Test and Evaluation (IOT&E) in recent years which have been evaluated as both not operationally effective and not operationally suitable. The following are the specific issues which the Task Force was asked to assess:

- Office of the Secretary of Defense (OSD) organization, roles, and responsibilities for Test and Evaluation (T&E) oversight. Compare organization, roles, and responsibilities in both DT&E and Operational Test and Evaluation (OT&E). Recommend changes that may contribute to improved DT&E oversight, and facilitate integrated T&E.
- Changes required to establish statutory authority for OSD DT&E oversight. Title 10 United States Code (USC) has an OT&E focus, and does not address OSD authority in oversight of DT&E. Recommend changes to Title 10 or other U.S. statutes that may improve OSD authority in DT&E oversight.
- Many IOT&E failures have been due to lack of operational suitability. Specific problems have been in the materiel readiness sustainment areas of reliability, maintainability, and availability. Recommend improvements in DT&E process to discover suitability problems earlier, and thus improve likelihood of operational suitability in IOT&E.

PROBLEM DEFINITION

In recent years, there has been a dramatic increase in the number of systems not meeting suitability requirements during IOT&E. Reliability, Availability and Maintainability (RAM) deficiencies comprise the primary shortfall areas. DoD IOT&E results from 2001 to 2006 are summarized in Figures 1 through 3. These charts graphically depict the high suitability failure rates during IOT&E resulting from RAM deficiencies. Figure 4 is a comparison of Army systems that met or did not meet reliability requirements.

Program	Service	ACAT	IOT&E Result		Reason
FY 2001					
F-15 TEWS	USAF	II	Effective	Not Suitable	Reliability, Maintainability, Availability
V-22 Osprey	Navy	1D	Effective	Not Suitable	Reliability, Availability, Maintainability (RAM), Human Factors, BIT
Joint Direct Attack Munitions (JDAM)	USAF	1C	Effective only with legacy fuses	Not Suitable	Integration with delivery platforms
M2A3 Bradley Fighting Vehicle	Army	1D	Effective	Suitable	
FY 2002					
Joint Primary Aircraft Training System (JPATS)	USAF	1C	Effective with deficiencies	Not Suitable	RAM, Safety, Human Factors
Cooperative Engagement Capability (CEC)	Navy	1D	Effective	Suitable	
Multiple Rocket Launcher System (MLRS)	Army	1C	Effective	Suitable	
MH-60S	Navy	1C	Effective	Not Suitable	RAM, excessive administrative and logistic repair time impacted RAM
FY 2003					
B-1B Block E Mission Upgrade Program	USAF	1D	Effective	Not Suitable	16% decrease in weapons release rate, reduction in accuracy of Mark 82 low drag weapons, 14% hit rate on moving targets
Sea wolf Nuclear Attack Submarine	Navy	1D	Effective	Suitable	Several requirement thresholds were not met but overall system effective and suitable

Figure 1. DoD IOT&E Results FY 2001-2003.

Program	Service	ACAT	IOT&E Result		Reason
FY 2004					
Evolved Sea sparrow Missile	Navy	II	Effectiveness unresolved	Suitable	Testing was not adequate to determine effectiveness.
Stryker	Army	1D	Effective	Suitable	
Advanced SEAL Delivery System (ASDS)	Navy	1D	Effective with restrictions	Not suitable	Effective for short duration missions; not effective for all missions and profiles. Not suitable due to RAM.
Tactical Tomahawk	Navy	1C	Effective	Suitable	
Stryker Mortar Carrier-B (MC-B)	Army	1D	Effective	Not Suitable	RAM and safety concerns.
FY 2005					
CH-47F Block I	Army	1C	Effective	Not Suitable	RAM; communications system less suitable than CH-47D; did not meet Information Exchange Requirements for Block I.
F/A-22	USAF	1D	Effective	Not Suitable	RAM; needed more maintenance resources and spare parts; BIT
Joint Stand-Off Weapon-C	Navy	1C	Not Effective		Not effective against moderately hardened targets; mission planning time was excessive.
Guided-MLRS	Army	1C	Effective	Suitable	
High Mobility Attack Rocket System (HMARS)	Army	1C	Effective	Suitable	
V-22 Osprey	Navy	1D	Effective	Suitable	
EA-6B (ICAP III)	Navy	II	Effective	Suitable	

Figure 2: **DoD IOT&E Results FY 2004-2005.**

Program	Service	ACAT	IOT&E Result		Reason
CY 2006					
Common Missile Warning System (CMWS)	Army	1C	Effective	Suitable	Effective and suitable in the OIF/OEF environment but needs further testing outside of the OIF/OEF environment.
Deployable Joint Command and Control (DJC2)	Navy	1AM	Effective	Not Suitable	Operational Test Agency, COTF, reported effective, not suitable. BLRIP not complete.
Integrated Defensive Electronic Countermeasures	Navy	II			Test suspended due to reliability problems.
Surface Electronic Warfare Improvement Program (SEWIP) Block 1A	Navy	II	Not Effective	Not Suitable	Block 1A Upgrade does not make the AN/SLQ-32 EWS operationally effective and suitable but does enhance ability to protect ships
C-130J	USAF	1C	Effective single ship; Not effective in formation	Suitable with shortfalls	Effective single ship; not effective in formation air land / air drop; not effective in non-permissive threat environment. Shortfalls in suitability due to maintainability issues
Small Diameter Bomb (SDB) Increment 1	USAF	1D	Effective with limitations	Suitable with limitations	Limited effectiveness and suitability due to bomb rack reliability and deficiencies in software used to predict optimum fuzing solutions. Oct 2006 flight operations suspended

Figure 3: DoD IOT&E Results for 2006.

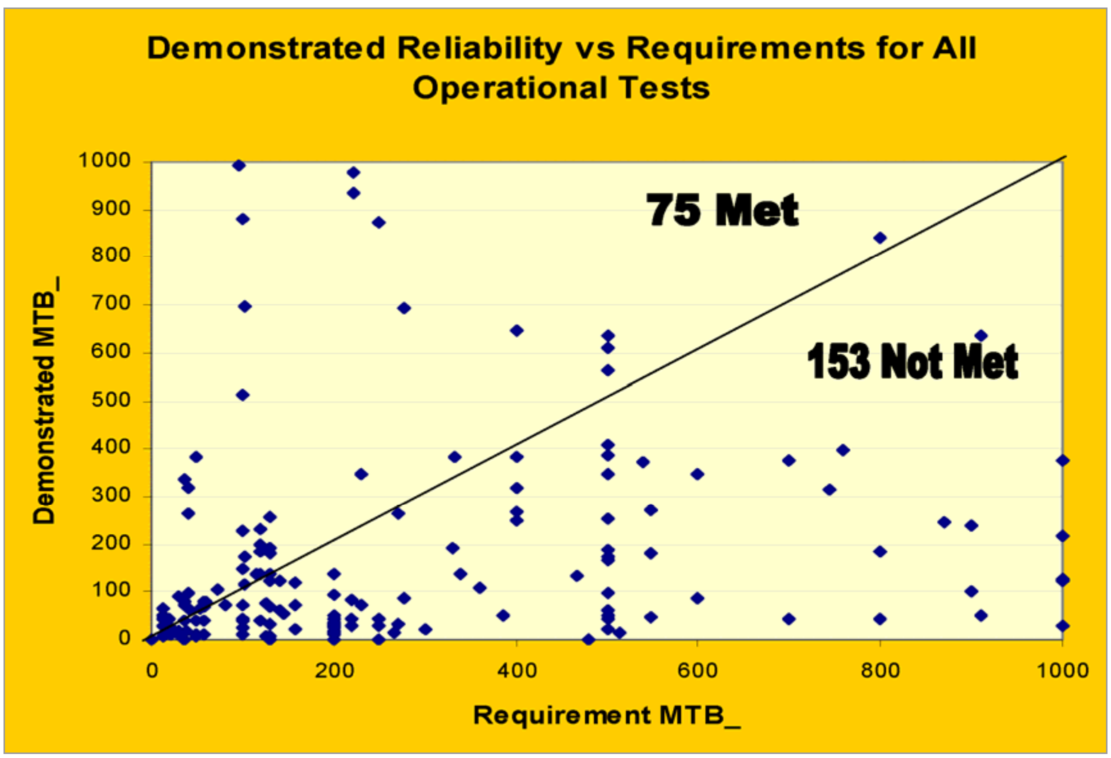


Figure 4: **Army Systems Failing Reliability during Operational Testing (1997-2006).**

Early in the DSB study, it became obvious that the high suitability failure rates were the result of systemic changes that had been made to the acquisition process; and that changes in developmental test and evaluation could not remedy poor program formulation. Accordingly, the Task Force study was expanded to address the broader programmatic issues, as well as the above issues identified in the Terms of Reference (TOR).

A number of major changes in the last 15 years have had a significant impact on the acquisition process. First, Congressional direction in Fiscal Year (FY) 1996, 1997, 1998 and 1999 Defense Authorization Acts reduced the acquisition workforce (which includes developmental test and evaluation). Several changes resulted from the implementation of Acquisition Reform in the late 1990s. The use of existing commercial specifications and standards was encouraged, unless there was justification for the use of military specifications. Industry was encouraged to use commercial practices. Numerous military specifications and standards were eliminated in some Service acquisition organizations. The requirement for a reliability growth program during development was also deemphasized, and in most cases, eliminated. At the same time, systems became more complex, and systems-of-systems integration became more common. Finally, there was a loss of a large number of the most experienced management and technical personnel in government and industry without an adequate replacement pipeline. The loss of personnel was compounded in many cases by the lack of up-to-date standards and handbooks, which had been allowed to atrophy, or in some cases, eliminated. It should be noted that Acquisition Reform included numerous beneficial initiatives. There have been many programs involving application of poor judgment in the last 15 years that can be attributed to acquisition/test workforce inexperience and funding reductions. It is probable that these problems would have occurred independently of most Acquisition Reform initiatives.

All Service acquisition and test organizations experienced significant personnel cuts, the magnitude varying from organization to organization. Over time, in-house DoD offices of subject matter experts (who specialized in multiple areas, such as promoting the use of proven reliability development methods) were drastically reduced, and in some cases, disestablished. A summary of reductions in developmental test personnel follows. The Army essentially eliminated their military Developmental Testing (DT) component and declared the conduct of DT by the government to be discretionary in each program. The Navy reduced their DT workforce by 10% but no shift of "hands-on" government DT to industry DT occurred. The trend within the Air Force gave DT conduct and control to the contractor. Air Force test personnel have been reduced by approximately 15% and engineering personnel supporting program offices have been reduced by as much as 60% in some organizations. The reduction of DT personnel in the Services occurred during a time when programs have become increasingly complex **(**e.g., significant increases in software lines of code, off-board sensor data integration, and systems of systems testing**)**. Congressional actions to cut the DOD acquisition workforce are also discussed in a recent National Research Council sponsored study.[1]

[1] Paul Kaminski, et al, Pre-Milestone A and Early Phase Systems Engineering: A Retrospective Review and Benefits for Future Air Force Acquisition, Washington, D.C., National Research Council, 2008.

PRINCIPAL FINDINGS AND RECOMMENDATIONS

RELIABILITY, AVAILABILITY, AND MAINTAINABILITY (RAM)

As a result of industry recommendations in the early 1970's, the Services began a concerted effort to implement reliability growth testing as an integral part of the development process. This implementation consisted of a reliability growth process wherein a system is continually tested from the beginning of development, reliability problems are uncovered, and corrective actions are taken as soon as possible. The Services captured this practice in their reliability regulations, and the Department of Defense (DoD) issued a new military standard on reliability, which included reliability growth and development testing as a best practice task. The goal of this process from 1980 until the mid-1990's was to achieve good reliability by focusing on reliability fundamentals during design and manufacturing rather than merely setting numerical requirements and testing for compliance towards the end of development.

The general practice of reliability growth was discontinued in the mid-to-late 1990's, concurrent with the implementation of Acquisition Reform. This discontinuance may not be a direct result of Acquisition Reform, but may be related instead to the loss of key personnel and experience, as well as short-sighted attempts to save acquisition funds at the expense of increased life cycle costs. With the current DoD policy, most development contracts do not include a robust reliability growth program. The lack of failure prevention during design, and the resulting low initial Mean Time Between Failure (MTBF) and low growth potential are the most significant reasons that systems are failing to meet their operational reliability requirements.

An OSD Cost Analysis Improvement Group (CAIG) study[2] shows operations and sustainment account for two-thirds or more of a system's life-cycle cost. According to Army studies[3], almost 90% of the sustainment costs are directly correlated with the reliability of the system. Given the amount of resources consumed during sustainment, investments in reliability enhancements can provide a very large return on that investment. A case study[4] conducted by the Logistics Management Institute (LMI), provided data that indicated an investment in total program reliability equal to twice the average production unit cost would yield an approximate 35% reduction in support costs.

FINDINGS

- Acquisition personnel reductions combined with acquisition system changes in the last 15 years had a detrimental impact on RAM practices
    - With some exceptions, the practice of reliability growth methodologies was discontinued during System Design and Development (SDD)
    - Relevant military specifications, standards and other guidance were not used
    - Suitability criteria, including RAM, were de-emphasized

[2] See Appendix I. Costs based on data reported in recent DoD Visibility and Management of Operating and Support Costs (VAMOSC) for programs, projected over the probable service lives of the systems.
[3] Michael Cushing, David Mortin, and Steve Yuhas, Improving Army Materiel Reliability: A Business Case Approach, Washington, D.C., AEC and AMSAA, 2007.
[4] Jim Forbes, Presentation on Empirical Relationships Between Reliability Investments and Life-Cycle Support Costs, Washington, D.C., LMI Government Consulting, 2007.

- Improved RAM decreases life cycle costs and reduces demand on the logistics system
- The Deficiency Report (DR) can be a valuable tool for early identification of RAM-related suitability problems, when used in conjunction with an adequately resourced deficiency correction system

RECOMMENDATIONS

The single most important step necessary to correct high suitability failure rates is to ensure programs are formulated to execute a viable systems engineering strategy from the beginning, including a robust RAM program, as an integral part of design and development. No amount of testing will compensate for deficiencies in RAM program formulation. To this end, the following RAM-related actions are required as a minimum:

- Identify and define RAM requirements during the Joint Capabilities Integration Development System (JCIDS), and incorporate them in the Request for Proposal (RFP) as a mandatory contractual requirement
- During source selection, evaluate the bidders' approaches to satisfying RAM requirements
  - Ensure flow-down of RAM requirements to subcontractors
  - Require development of leading indicators to ensure RAM requirements are met
- Make RAM, to include a robust reliability growth program, a mandatory contractual requirement and document progress as part of every major program review
- Ensure that a credible reliability assessment is conducted during the various stages of the technical review process and that reliability criteria are achievable in an operational environment
- Strengthen program manager accountability for RAM-related achievements
- Develop a military standard for RAM development and testing that can be readily referenced in future DoD contracts
- Ensure a adequate cadre of experienced RAM personnel are part of the Service acquisition and engineering office staffs

ROLES AND RESPONSIBILITIES OF GOVERNMENT TEST AND EVALUATION ORGANIZATIONS

The role of the government in the DT process has evolved over the past 50 years. Historical catalysts for change have included technological advances, acquisition policy changes, government resource availability and, in recent years, the Global War on Terrorism (GWOT). The most significant acquisition policy changes in the past several decades were made as a part of Acquisition Reform in the mid-to-late 1990's. With some exceptions, there has been a significant decrease in government involvement in test planning, conduct and execution.

The traditional role of the government during the DT planning phase included the identification of the test resource requirements and government test facilities, the development of the test strategy and detailed test and evaluation plans, as well as the actual conduct of T&E. When a program moved from the planning phase to the test execution phase, the government traditionally participated in test conduct and analysis; performing an evaluation of the test results for the program office. With some exceptions, this is no longer the case. Until recently, it was recognized that there should be some level of government involvement and oversight even when the contractor has the primary responsibility regarding planning and execution of the DT program.

In addition to the reduction in the number of government acquisition and test personnel, the experience level of both government and industry personnel has steadily diminished in recent years. A significant percentage of the workforce became eligible to retire since 2000, and due to prior downsizing, there has not been a steady pipeline of younger technical personnel to replace them. As an example, Appendix I is a chart depicting near-term retirement eligibility for Major Range and Test Facility Base (MRTFB) personnel. Two-thirds or more of the senior civil service personal are eligible for retirement.

FINDINGS

The changes in the last 15 years, when aggregated, have had a significant negative impact on DoD's ability to successfully execute increasingly complex acquisition programs. Major contributors include massive workforce reductions in acquisition and test personnel, a lack of up-to-date process guidance in some acquisition organizations, acquisition process changes, as well as the high retirement rate of the most experienced technical and managerial personnel in government and industry without an adequate replacement pipeline.

- Major personnel reductions have strained the pool of experienced government test personnel
- A significant amount of developmental testing is currently performed without a needed degree of government involvement or oversight and in some cases, with limited government access to contractor data

RECOMMENDATIONS

- As a minimum, government test organizations should develop and retain a cadre of experienced T&E personnel to perform the following functions:
  - Participate in the translation of operational requirements into contract specifications, and in the source selection process, including RFP preparation
  - Participate in DT&E planning including Test and Evaluation Master Plan (TEMP) preparation and approval
  - Participate in technical review processes
  - Participate in test conduct, data analysis, and evaluation and reporting; with emphasis on analysis and reporting
- Utilize red teams, where appropriate, to compensate for shortages in skilled, experienced T&E domain and process experts
- Develop programs to attract and retain government personnel in T&E career fields so that the government can properly perform its role as a contract administrator and as a “smart buyer”

INTEGRATED TEST AND EVALUATION

Integrated testing is not a new concept within the Department of Defense, but its importance in recent years has been highlighted, due in part to the growth of asymmetric threats and the adoption of net-centric warfare. The December 2007 OSD Test and Evaluation Policy Revisions memorandum reinforces the need for integrated testing.[5] Implementation of integrated test concepts has been allowed to evolve on an ad-hoc basis. The time has come to pursue more consistency in integrated test planning and execution.

Collaboration between developmental and operational testers to build a robust integrated test program will increase the amount of operationally relevant data that can be used by both communities. DT and Operational Test (OT) planning is separate and this inhibits efforts by the Services to streamline test schedules, thereby increasing the acquisition timeline and program test costs.

Additionally, there is a widely held assumption by many in the OT community that only data from independent OT is acceptable for operational evaluation purposes. While not all information from DT may be useable by the Operational Test Agency (OTA) to support IOT&E, a significant amount of developmental test data can be used to partially satisfy OT requirements. More importantly, an operational perspective earlier in the developmental process has often proven to be a catalyst to early identification and correction of problems.

DoD policy should mandate integrated test planning and execution on all programs to the extent possible. To accomplish this, programs must establish a team made up of all relevant organizations (including contractors, developmental and operational test and evaluation communities) to create and manage the approach to incorporate integrated testing into the T&E Strategy and the TEMP.

FINDINGS

- Service acquisition programs are incorporating integrated testing to a limited degree through varying approaches
- Additional emphasis on integrated testing will result in greater T&E process efficiency and program cost reductions

RECOMMENDATIONS

- Implement OSD and Service policy[6] mandating integrated DT&E/OT&E planning and execution throughout the program
  - Require sharing and access to all appropriate system-level and selected component-level test and model data by government DT and OT organizations, as well as the prime contractor, where appropriate
  - Integrate test events, where practical, to satisfy OT and DT requirements

OPERATIONAL TEST READINESS REVIEW (OTRR)

Each Service has an Operational Test Readiness Review (OTRR) process. Although it varies from Service to Service, the process generally results in in-depth reviews of readiness to undergo an IOT&E event.

[5] See Appendix D.
[6] See Appendix D.

FINDINGS

- Department of Defense Instruction (DoDI) 5000.2 requires that “the Service Acquisition Executive (SAE) shall evaluate and determine materiel system readiness for IOT&E"
  - Decision authority is frequently delegated to the appropriate Program Executive Officer (PEO)
  - Materiel developer is also required to furnish DT&E report to the Under Secretary of Defense for Acquisition, Technology and Logistics (USD[AT&L]) and Director, Operational Test and Evaluation (DOT&E)
- Shortcomings in system performance, suitability, and RAM are usually identified during the OTRR
- In most cases, the operational test readiness certifying authority is well aware of the risk of not meeting OT criteria when major shortcomings exist
- Because of funding constraints, the low priority given to sustainment, as well as the urgency in recent years to get new capabilities to the Warfighter, major suitability shortcomings have rarely delayed the commencement of dedicated IOT&E

RECOMMENDATIONS

- Conduct periodic operational assessments to evaluate progress and the potential for achieving pre-determined entrance criteria for operational test events
- Conduct an independent Assessment of Operational Test Readiness (AOTR) prior to the OTRR (included in latest draft DODI 5000.2)
- Include a detailed RAM template in preparation for the OTRR
- Require the Command Acquisition Executive (CAE) to submit a report to OSD that provides the rationale for the readiness decision

OSD TEST AND EVALUATION ORGANIZATION

The Task Force was asked to assess OSD roles and responsibilities for T&E oversight. T&E has been a visible part of OSD since the early 1970’s, reporting to the Research and Engineering command section when it was in charge of acquisition oversight and subsequently to the Under Secretary of Defense for Acquisition (now AT&L). The early T&E office was responsible for all T&E, ranges, resources oversight, and policy. In 1983, Congress established an independent Director, Operational Test and Evaluation (DOT&E) organization, reporting directly to the Secretary of Defense (SECDEF), responsible for operational test and evaluation policy, budget review, and assessments of operational effectiveness and suitability. The Live Fire Test (LFT) oversight function was created and added to the DT&E office responsibilities in the mid 1980’s. Later, the LFT oversight function was moved to the DOT&E organization.

In 1999, the DT&E organization was dismantled by DoD. Many functions were moved to DOT&E, including test ranges and resources, and joint T&E oversight. Some of the remaining T&E personnel billets were eliminated to comply with a congressionally mandated (AT&L) acquisition staff reduction. The residual DT&E policy and oversight functions were separated and moved lower in the AT&L organization.

A 2000 DSB Task Force Study on Test and Evaluation Capabilities recommended that DoD create a test and evaluation resource enterprise within the office of the DOT&E to provide more centralized management of T&E facilities. This recommendation ultimately led to removing the test ranges and resources oversight from DOT&E, abandoning the notion of centralized management, and the establishment of the Test Resource Management Center (TRMC) in AT&L (as directed by the National Defense Authorization Act for Fiscal Year 2003).

FINDINGS

Current policy as of December 2007 mandates that developmental and operational test activities be integrated and seamless throughout the system life cycle. There must be enough experts in OSD with the ability to understand and articulate lessons learned in early testing and the ability to execute the new T&E policy. That policy is to "take into account all available and relevant data and information from contractors and government sources" in order to "maximize the efficiency of the T&E process and effectively integrate developmental and operational T&E."[7]

- Currently there is not an OSD organization with comprehensive DT oversight responsibility, authority or staff to coordinate with the operational test office
  - The historic DT organization has been broken up and residual DT functions were moved lower in organization in 1999, and lower yet in 2002
  - Programmatic DT oversight is limited by staff size and often performed by generalists vice T&E experts
  - Recruitment of senior field test personnel is hampered by DT's organizational status
  - Existing residual organizations are fragmented and lack clout to provide DT guidance
  - System performance information and DT lessons learned across DoD has been lost
  - DT is not viewed as a key element in AT&L system acquisition oversight
  - Documentation of DT results by OSD is minimal
- Access to models, data, and analysis results is restricted by current practice in acquisition contracting, and the lack of expertise in the DT organization
- TRMC has minimal input to program-specific questions or interaction with oversight organizations on specific programs
  - Organizational separation is an impediment

RECOMMENDATIONS

- Implementation of integrated and seamless DT and OT will require, at a minimum, greater coordination and cooperation between all testing organizations
- Consolidate DT-related functions in AT&L to help reestablish a focused, integrated, and robust organization[8]
  - Program oversight and policy, and Foreign Comparative Test (FCT)
  - Have Director, DT&E directly report to Deputy Under Secretary of Defense, Acquisition and Technology (DUSD[A&T])

[7] See Appendix D.
[8] Three Task Force members out of fourteen voted against consolidation.

    - Restore TEMP approval authority to Director, DT&E
- Integrate TRMC activities early into DT program planning
    - Make TRMC responsible for reviewing the resources portion of the TEMP
- If such an organization is established and proves itself effective, consider as part of a future consolidation moving LFT back to its original DT location (this would require congressional action and DOT&E concurrence)

Most of the organizational changes recommended above are within the purview of AT&L. The LFT change requires the concurrence of DOT&E and a legislative change to Title 10 because of the change in reporting official. All the other recommendations made throughout the report can be implemented within current DoD authority.

OTHER ISSUES

Several other issues were addressed as a part of the study. A discussion of each of the following topics, along with findings and recommendations, may be found in the body of the report.

- Program Structure
- Requirements Definition
- Contractual Performance Requirements
- Alignment of DoD Technology with Systems Engineering Procedures
- Commercial Off-The-Shelf (COTS)
- Systems of Systems (SoS)

II. INTRODUCTION

The Defense Science Board Task Force on Developmental Test and Evaluation (DT&E) was established by the Under Secretary of Defense for Acquisition, Technology and Logistics[9] to examine T&E roles, responsibilities, policy, and practices; and to recommend changes that may contribute to improved success in IOT&E as well as quicker delivery of improved capability and sustainability to the Warfighter. The Task Force study was the result of a dramatic increase in recent years in the number of programs that have been evaluated as not operationally effective and/or operationally suitable at the completion of IOT&E[10]. Approximately 50% of the programs completing IOT&E since 2000 have been assessed as not operationally effective and/or suitable. Problems in the suitability area predominate with reliability deficiencies serving as the major shortcoming[11].

The Task Force was asked to assess:

- OSD organization, roles, and responsibilities for T&E oversight. Compare organization, roles, and responsibilities in both DT&E and OT&E. Recommend changes that may contribute to improved DT&E oversight, and facilitate integrated T&E.
- Changes required to establish statutory authority for OSD DT&E oversight. Title 10 USC has an OT&E focus and does not address OSD authority in oversight of DT&E. Recommend changes to Title 10 or other U.S. statutes that may improve OSD authority in DT&E oversight.
- Many IOT&E failures have been due to lack of operational suitability. Specific problems have been in the materiel readiness sustainment areas of reliability, maintainability, and availability. Recommend improvements in DT&E process to discover suitability problems earlier, and thus improve the likelihood of operational suitability in IOT&E.

III. PROBLEM DEFINITION

In recent years, there has been a dramatic increase in the number of systems not meeting suitability requirements during IOT&E. RAM deficiencies comprise the primary shortfall areas. DoD IOT&E results from 2001 to 2006 are summarized in Figures 1 through 3[12]. These charts graphically depict the high suitability failure rates during IOT&E resulting from RAM deficiencies. Figure 4 is a comparison of Army systems that met or did not meet reliability requirements..[13]

[9] See Appendix A.
[10] See Figures 1-3 (pages 2-3).
[11] See Figure 4 (page 3).
[12] See Executive Summary (pages 2-3).
[13] See Executive Summary (page 3).

IV. BACKGROUND

A. MAJOR CHANGES IN THE LAST 15 YEARS

Starting with the post-Cold War world of the early 1990's, many initiatives have been undertaken to reduce the defense budget. In the aggregate, these initiatives have resulted in both beneficial and detrimental changes to the acquisition process.

Congress directed several sequential cuts to the DoD acquisition workforce in the late 1990's. Congressionally mandated reductions started in FY1996, with additional cuts in the FY1997, 1998, and 1999 Defense Authorization Acts. The reductions mandated by Congress put the DoD acquisition workforce on a precipitous path and are highlighted in a recent National Research Council study.[14] Personnel reductions, combined with implementation of new contracting approaches (e.g., performance-based contracting and Total System Performance Responsibility [TSPR]) reduced close government oversight. Over time, in-house DoD offices of subject matter experts (who specialized in multiple areas, such as promoting the use of proven reliability development methods) were drastically reduced, and in some cases, disestablished. The invoking of military specifications and standards in defense contracting was replaced by strong encouragement for industry to use what were thought to be less costly commercial practices and Commercial Off-The-Shelf (COTS) products.

Organizational efficiencies have been pursued through Base Realignment and Closure (BRAC) decisions. Also, efforts have been underway to bring a closer integration of the developmental and operational T&E organizations.

At the same time, the emergence of asymmetric threats have driven the need for the Services to collaborate closely in warfighting and the requirement for new joint Service operational capabilities and acquisition programs.

The attacks of September 11, 2001, ushered in a new era of warfighting with the Global War on Terrorism. Significant priority was given to finding more efficient ways to deliver new capabilities to the Combatant Commanders for use against quickly adapting threats. Rigorous T&E before deployment was sometimes sacrificed to meet schedule demands.

Major changes, some of which had detrimental impacts, are summarized below:

- Congressionally mandated cuts to the acquisition workforce
  - Decreased government management of acquisition program formulation and execution
- Acquisition Reform
  - Strong encouragement to use commercial specifications and standards in lieu of military specifications
  - Emphasis on commercial practices and products, including COTS equipment and components

[14] Paul Kaminski, et al, Pre-Milestone A and Early Phase Systems Engineering: A Retrospective Review and Benefits for Future Air Force Acquisition, 2008.

  - Performance-based contracting
  - Reduced government oversight
  - Emphasis on survivability and lethality testing below end item
- Reliability growth during system development was deemphasized or eliminated
- Emphasis on evolutionary acquisition and spiral development
- Increased weapon systems complexity
- Reduced government resources and workforce
- Integrated developmental and operational test emphasis
- Emphasis on joint programs
- BRAC realignment
- Loss of strong Service advocates for RAM (Willoughby, Lorber, Goodell)
- More emphasis on systems engineering
- Initiated Systems of Systems (SoS) acquisition approach

B. RECENT ARMY T&E CHANGES

Developmental testing within the U.S. Army has experienced changes in five principal areas over the last 15 years. They are:

- Organizational
- Personnel
- OPTEMPO (Operational Tempo)
- Legal
- Acquisition

In 1999 the Army consolidated the U.S. Army Test and Evaluation Command (TECOM), the Army's principal developmental tester, with the Army's U.S. Army Operational Test and Evaluation Command (OPTEC). This consolidation was driven by recommendations from the Army Science Board (ASB) which asserted that a single Army T&E command would provide more effective and efficient support to the materiel acquisition process and to the Warfighter. This organizational consolidation of testers, along with the previous consolidation of developmental and operational evaluators in 1996, resulted in the first fully-consolidated Service T&E organization within DoD; known as the U.S. Army Test and Evaluation Command (ATEC). Particularly noteworthy is the creation of a single ATEC system team chairperson to develop the early T&E concept with the ability to tailor the testing to eliminate rigid DT and OT stovepipes and minimizes the unnecessary use of resources. This required a significant cultural change but was essential to ATEC's successful integration of DT and OT. Additional benefits have been gained by the development of a single integrated investment strategy across DT and OT organizations along with a single Information Technology (IT) Enterprise architecture. This architecture allows internal and external customers to view all T&E information on a particular acquisition system. Previous barriers, whether real or perceived, have been broken and cooperation has increased.

Personnel levels within the U.S. Army Developmental Test Command (formerly TECOM) decreased by 45% throughout the 1990's to a level of 5,834 in 2000. During this period the military strength was reduced by 97%, virtually eliminating the Army's Soldier, Operator, Maintainer, Test and Evaluation (SOMTE) personnel. The SOMTE soldiers, who were trained as testers and had recent field experience, influenced the planning and test conduct by providing material interface for early input on possible tactics, techniques, and procedures. Their involvement was more efficient than attempting to borrow troops from operational units that are already taxed by their existing missions.

Although military levels have not returned to the Army DT, there has been recognition of the value of a dedicated Army Evaluation Task Force (AETF) for the upcoming Army Future Combat Systems (FCS). A Brigade Combat Team (BCT) has been formed at Fort Bliss, Texas, to provide military test support for the FCS program.

There was a major shift from a government civilian to a contractor workforce from 1991 to 2007. Civilians were reduced by 45% while contractors were increased by 50%. Although contracting is an excellent strategy to augment the current civilian workforce, it should not replace the civilian workforce. Civilians are the stable part of the workforce. Contractors are employed and released based on workload demands. It is impossible to maintain the required testing experience and institutional knowledge necessary without maintaining a core cadre of government civilian personnel. The civilian workforce should be deemed critical in order to maintain the required institutional knowledge and skills to support current and future testing needs.

Major changes are summarized below:

- Testing functions consolidated into ATEC (1999)
- Army designated government DT as discretionary (late 1990s)
- Major shift in DT military, civilian, and contractor workforce mix (1991 to 2007)
  - Military strength reduced from 1836 to 61 soldiers
  - Civilian manpower reduced from 5610 to 3076
  - Contractor full-time equivalent increased from 3072 to 4599
- Rapid Fielding Initiatives (RFIs) reduced test times at the expense of:
  - Increased workloads (up to 24 hours a day, 6 days a week)
  - Decreased traditional subtests such as limited E3, cold and tropic testing, etc.
  - Suitability requirements

C. RECENT NAVY T&E CHANGES

The Navy conducted 264 operational test events between FY1999 and FY2007 (May 2007). The averages over this time found programs 85.6% effective and 67.4% suitable. The figure on the next page displays the Navy's operational test success rates over the past 8 years:

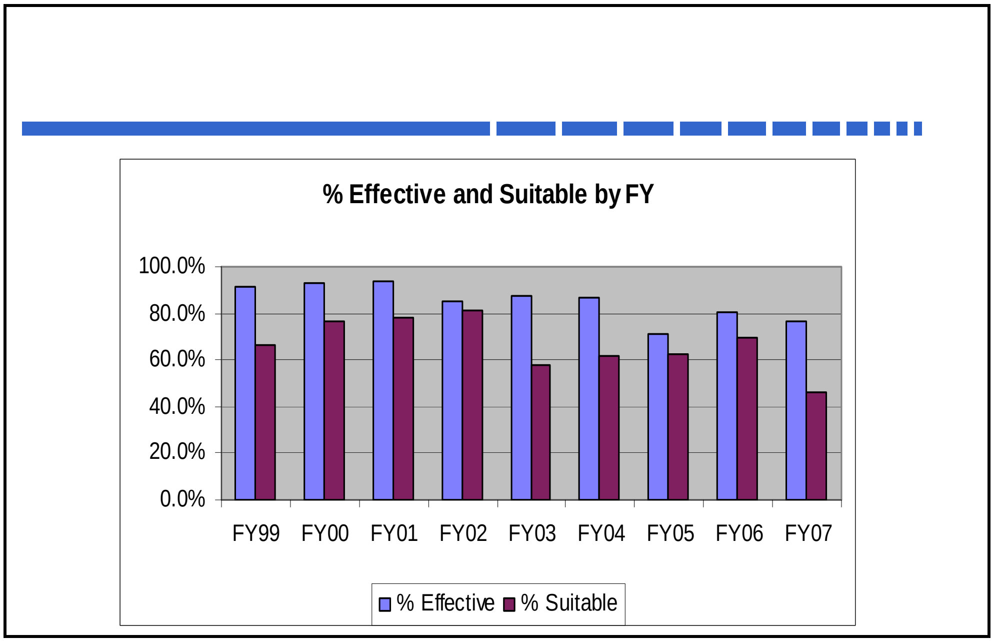


Figure 5: Navy Operational Test Success Rates FY 1999-2007.

Naval Aviation DT has always been "hands on." No shift to oversight has ever occurred. The Naval Air Systems Command (NAVAIR), Naval Sea Systems Command (NAVSEA), and Space and Naval Warfare Systems Command (SPAWAR), generally support the Integrated Test Team (ITT) concept. Integrated DT&E and OT&E provides an opportunity for the Commander, Operational Test and Evaluation Force (COMOPTEVFOR) to participate in DT, provide information to the Decision Authority (DA) in advance of independent OT, enlarge the OT database, and reduce the amount of operational testing. Integrated testing is encouraged and is governed by the TEMP. Program managers have discretion to determine where integrated testing will improve overall test process effectiveness.

In June 2005, the Navy Operational Test Agency formally recognized that testing needs to be considered as a continuum of events. COMOPTEVFOR formally adopted the following goals:

- Seek integration of testing, where appropriate
- Eliminate redundant events
- Recognize the distinct disciplines in developmental and operational test
- Use both to better support the systems engineering process.

A truly integrated approach takes a significant investment of intellectual capital upfront. The OT community is obliged to develop a more sophisticated understanding of the system under development. The requirements community is challenged to articulate Concept of Operations (CONOPS) and to more clearly define system requirements. A more robust definition of the threat environment is needed. The development of a common database that assures the pedigree of the data is essential but may not be compatible with many of the contractual vehicles developed in the late 1990's (e.g., some acquisition contracts awarded during that period did not contain the necessary contract data requirements clauses to obtain access to contractor test data).

Some variations of DT approaches exist within the Navy. Aviation has a robust and organizationally identifiable DT community. The surface community treats DT as a subset of systems engineering. The Command, Control, Communications, Computers and Intelligence (C4I) community has been challenged to find realistic venues for full scale developmental and operational tests. The common approach tends to be to demonstrate new technology and capability while taking older technology as a given (e.g., new long range radar performance versus short range modes, shallow water torpedo modes tested, deep water assumed to be good). Integration with legacy systems is not always adequately tested. Even the best software test facilities seem to be challenged to accurately estimate the time and effort required. Challenges of integrating with older, poorly documented systems still exist.

Major changes are summarized below:

- Navy enterprise T&E board of directors established in 2007
  - Encourages integration of priorities across Navy Warfighting Enterprises
  - Facilitates continuous process improvement initiatives
  - Shares and leverages best practices
- Personnel levels in Navy DT reduced 10% across the board
  - No shift from government "hands-on" DT to oversight has occurred
- Integrated prime contractor/government DT/ OT (Integrated Test Teams)
- Navy DT closely coupled with systems engineering process

D. RECENT AIR FORCE T&E CHANGES

Policies implemented as a result of budgetary pressures from the end of the Cold War through the early 1990's affected resources flowing into DT&E capability development and sustainment. The divergence of policy governing requirements definition, acquisition, and test, along with Air Force acquisition's transition to Total System Performance Responsibility (TSPR) in the later part of the decade, collectively led to a "hollowing out" of the test capability infrastructure and the perception of government "rent-a-ranges" by weapon system Original Equipment Manufacturers (OEMs). In addition, there was a major reduction in the in-house engineering workforce. Some of these trends have begun to reverse. Accordingly, Air Force policy documents were updated after 2002 to re-link requirements, acquisition and test planning, and management and execution. However, additional effort is needed to correctly define, resource, and sustain current test capability needs; plan for future test capability requirements; and realign or divest test capabilities and infrastructure that are minimally used or no longer required, in order to create an optimum balance of DT&E test capabilities.

Budgetary pressures have historically influenced both the policy and resources governing the utilization and sustainment of test capabilities retained by the Air Force. These budgetary pressures and the resulting challenges of implementing long-range test capability strategies remain today. The DT&E portion of the budget has declined, in proportion of Research, Development, Test and Evaluation (RDT&E) funding, from approximately 9.8% in 1996 to approximately 7.3% in 2005. This divergent trend has challenged the Air Force's ability to maintain current DT&E capabilities across test centers and has limited investments in new capabilities necessary for testing advanced technologies (e.g., hypersonics and directed energy weapon systems).

A healthy DT&E capability requires a trained technical workforce that supports current test requirements, adjusts to meet tactical changes, and yet is also prepared to support testing of advanced technologies and future test requirements. The current Air Force DT&E workforce is a composite of military (officer and enlisted), government civilians, and contractors. The Air Force DT&E workforce has decreased approximately 15% since FY1992 and has shifted towards a much reduced organic (civilian, officer and enlisted) government T&E workforce. In 1994, contractors made up approximately 20% of the total workforce; while in 2003 contractors comprised approximately 50% of the workforce. During the same period, the organic government T&E workforce has generally declined. This trend, combined with policy shifts towards increasing OEM developmental test activities, suggests that a significant amount of DT&E is performed without the benefit of government insight or oversight.

Major changes are summarized below:

- Current trend is to give DT&E conduct and control to the prime contractor
  - Ability to conduct government DT&E and independent analysis has significantly diminished
  - Increasing perception that Air Force test activities are migrating to "rent-a- range" support for system manufacturers
- Test personnel levels have decreased approximately 15% since 1992
  - Support contractor workforce increased from 20% to 50%
- Aeronautical systems engineering workforce declined 59% in last 15 years
  - Established Central Test Authorities (CTAs) in 2004
- Moved PEOs from the Pentagon to Product Command Centers
- DT reporting activity decreased from an average of 200 reports per year in the 1980's to approximately 50 reports per year since the mid 1990's

V. PRINCIPAL FINDINGS AND RECOMMENDATIONS

A. RELIABILITY, AVAILABILITY, AND MAINTAINABILITY (RAM)[15]

As a result of industry recommendations in the early 1970's, the Services began a concerted effort to implement reliability growth testing as an integral part of the development process. This implementation consisted of a reliability growth process wherein a system is continually tested from the beginning of development, reliability problems are uncovered, and corrective actions are taken as soon as possible. The Services captured this process in their reliability regulations (i.e., Army Regulation 702-3, the Air Force R&M 2000, and the Navy Willoughby Templates Best Practices, NAVSO P-6071). In 1980, DoD issued a new military standard on reliability, Mil Std 785 B *Reliability Program for Systems and Equipment Development and Production*, which included reliability growth and development testing as a best practice task. In 1981, DoD also issued a new military handbook on reliability growth, Mil Handbook 189, *Reliability Growth Management*, which addressed the concepts and principles of reliability growth, advantages of managing reliability growth, and guidelines and procedures to be used in managing reliability growth.

The goal of the reliability process from 1980 until the mid-1990's was to increase the reliability of the programs tested, mainly by focusing on reliability fundamentals during the design and manufacturing phases rather than merely focusing on numerical performance and test requirements during full scale DT and IOT&E. DoD data from 1980-1988 shows that programs adopting a reliability growth approach increased system MTBF by an average factor of four. In the 1990's and later, testing for reliability growth was not typically conducted. The logical conclusion, everything being equal, is that systems are being fielded with a fraction of the MTBF potential. This conclusion is consistent with the high percentage of systems not meeting requirements as well as the high support costs. On the other hand, those DoD systems that still employ reliability growth generally field systems with higher effectiveness and lower support costs than they would if they had not had a reliability growth program.

As stated above, the general practice of reliability growth was discontinued with the implementation of Acquisition Reform in the mid-to-late 1990's. An exception is aircraft development programs in the Navy and Air Force. Reliability growth testing during development remains a standard practice for most of these programs. However, as can be seen in Figures 1 and 2[16], RAM failure rates during IOT&E are high in many aircraft programs. Adequate attention is not given to initial requirements. Contributing factors include the elimination of a viable reliability organization in some acquisition organizations as well as a lack of up-to-date process guidance. Additionally, problems uncovered through the DR process and other reporting mechanisms are usually not corrected. With the current DoD practices, most other development contracts since the mid-1990's do not include a robust reliability growth program. These programs rely instead on an operational reliability test to demonstrate a numerical MTBF requirement. Experience over the past 35 years has shown that reliability demonstration tests not preceded by a strong reliability growth program during development

[15] See Appendix E for RAM definitions.
[16] See page 2.

have proven to be ineffective in providing acceptable levels of RAM. In contrast, reliability growth practices have proven effective on many major fielded systems (e.g., the Army's Black Hawk, M1 Abrams, Bradley, and Patriot, as well as many Navy and Air Force systems).

The successful achievement of system suitability is directly related to the management and engineering attention given to RAM-related failure prevention during the design phase. These management and engineering activities are primarily directed toward increasing the system's initial operational MTBF and establishing the system's growth potential MTBF above the minimum acceptable requirement. Reliability growth is successful if it is based on a fundamentally good design with a modest number of deficiencies.

The lack of continuous RAM improvement during design, and the resulting low initial MTBF and low Growth Potential are the most significant reasons that systems are failing to meet their operational suitability requirements. Good RAM is vital to many suitability factors, including training and logistics support. Dependable failure prevention starts with the reliability growth in the design phase. In order to attain needed and sustainable RAM in fielded systems, RAM generally needs to increase beyond the minimum acceptable requirements, toward the growth potential. This additional increase in reliability usually requires finding failure modes through continuous testing that could not otherwise be predicted and prevented in the design phase, such as due to interactions between components. Although failure modes are typically uncovered before the completion of DT, failure analysis and corrective actions generally do not occur in time for cost-effective, design-optimized solutions.

There are numerous adverse consequences that have been shown to result from suitability shortfalls. The first and most significant are under performance in the field and substantial increases in life cycle costs. These cost implications are discussed in more detail below. There are also additional near-term ramifications. Some programs are forced to extend the SDD phases and add unplanned resources for redesign, reengineer and retest. The V-22 program extended SDD by five years and spent over $1 billion additional funds to resolve its suitability issues. The Joint Air-to-Surface Standoff Missile (JASSM) was similarly delayed from proceeding to full production and deployment because of poor RAM. When poor performance in IOT&E delays the full production or fielding of a system, additional costs are incurred to operate and support the system in the near term. In some cases, the correction of RAM deficiencies are deferred until a major block upgrade or Service Life Extension Program (SLEP). The corrections, as well as remedial actions and the retrofits into already fielded systems, necessitate additional funding. The operation and maintenance of multiple configurations of fielded systems is another unplanned expense. Any system that is fielded with known RAM shortcomings increases the costs for maintenance and repair as well as for additional contractor logistics and spare parts support.

An OSD CAIG study shows that Operations and Sustainment (O&S) costs account for two-thirds or more of a system's total life-cycle cost[17]. According to Army studies[18], almost 90% of the in-service costs are directly correlated with the reliability of the system. Given the amount of resources consumed during sustainment, investments in reliability enhancements can provide a

[17] See Appendix I.
[18] Michael Cushing, et al, Improving Army Materiel Reliability: A Business Case Approach, 2007.

substantial return on that investment. A case study conducted by LMI[19], provides data that indicate an investment in total program reliability equal to twice the average production unit cost would yield an approximate 35% reduction in support costs. This conclusion was based on major combat systems; these systems are large enough to provide significant funding to implement a reliability program. The conclusions of the LMI case study suggest that a relatively modest investment in reliability programs would have a substantial positive impact on the lifecycle cost of the system. In addition, according to an Army study[20], increasing reliability had the largest positive effect on operational availability. Reliability issues are the most common root cause of 50% of systems being judged as not suitable in IOT&E. A recent GAO study[21] reinforced the need for increased attention to Weapons Systems quality issues, particularly reliability. A recent study by the Institute for Defense Analyses also suggests that investing in reliability is the most effectively accomplished during System Design and Development.[22]

FINDINGS

- Acquisition personnel reductions combined with acquisition system changes in the last 15 years had a detrimental impact on RAM practices
    - With some exceptions, the practice of reliability growth methodologies was discontinued during SDD
    - Relevant military specifications, standards and other guidance were not used
    - Suitability criteria, including RAM, were de-emphasized
    - The technical/managerial workforce was reduced in most government program offices and test organizations
        - Reliability staff was reduced from 50 to 5 in one major acquisition organization
- Improved RAM decreases life cycle costs and reduces demand on the logistics system
- RAM shortfalls are frequently identified during DT, but program constraints (schedule and funding) often preclude incorporating fixes and delaying IOT&E
- By the time reliability data are analyzed in IOT&E, it is generally too late to make significant design changes (i.e. improve the reliability within program resources)
- In some instances, programs had such serious RAM concerns that they were precluded from proceeding to production until the problems could be corrected. Improved RAM decreases life cycle costs and reduces demand on the logistics system
- The DR can be a valuable tool for early identification of RAM-related suitability problems when used in conjunction with an adequately resourced deficiency correction system

RECOMMENDATIONS

The single most important step necessary to correct high suitability failure rates is to ensure programs are formulated to execute a viable systems engineering strategy from the

[19] Jim Forbes, Presentation on Empirical Relationships Between Reliability Investments and Life-Cycle Cost, 2008.
[20] Michael Cushing, et al, Improving Army Materiel Reliability: A Business Case Approach, 2007.
[21] United States Government Accountability Office, Best Practices: Increased Focus on Requirements and Oversight Needed to Improve DoD's Acquisition Environment and Weapon System Quality, Washington, D.C., February 2008.
[22] K. Lo Tzee-Nan, Cost of Unsuitability: Assessment of Trade-offs Between the Cost of Operational Unsuitability and RDT&E Costs, IDA Draft Paper P-4330.

beginning, including a robust RAM program, as an integral part of design and development. No amount of testing will compensate for deficiencies in RAM program formulation. To this end, the following RAM-related actions are required as a minimum:

- Identify and define RAM requirements during the JCIDS process, and incorporate them in the RFP as a mandatory contractual requirement
- During source selection, evaluate the bidders' approaches to satisfying RAM requirements
  - Ensure flow-down of RAM requirements to subcontractors
  - Require development of leading indicators to ensure RAM requirements are met (all Technical Performance Measurements [TPMs])
- Make RAM, to include a robust reliability growth program, a mandatory contractual requirement and document progress as part of every major program review
- Ensure that a credible reliability assessment is conducted during the various stages of the technical review process (System Readiness Review [SRR], Preliminary Design Review [PDR], Critical Design Review [CDR], etc.) and that reliability criteria are achievable in an operational environment
- Strengthen program manager accountability for RAM-related achievements
- Develop a military standard for consistent RAM development and testing that can be readily referenced in future DoD contracts
- Modify the curriculum at the Defense Acquisition University (DAU) to stress the importance of a robust reliability design and test effort as part of the systems engineering process
- Ensure an adequate cadre of experienced RAM personnel are part of the Service acquisition and engineering office staffs
- **Action**
  - **USD(AT&L) charter an OSD/Service Task Force to implement RAM recommendations**

B. ROLES AND RESPONSIBILITIES OF GOVERNMENT TEST AND EVALUATION ORGANIZATIONS

The role of the government in the DT process has evolved over the past 50 years. Catalysts for change have historically included technological advances (e.g., software intensive systems, fly-by-wire, SoS), acquisition policy changes, government resource availability, and in recent years, the Global War on Terrorism. Perhaps the most significant acquisition policy changes in the past several decades were made as a part of Acquisition Reform in the mid-to-late 1990's. The cancellation of many military specifications and standards without replacing them with adequate commercial specifications and standards had a detrimental impact on system development. The promotion of performance-based contracting and TSPR contract vehicles impacted the government's role in the DT process. Additionally, the government acquisition and test workforce was reduced as a part of the reform process as well as by Congressionally mandated cuts. The Services test personnel drawdown specifics are discussed in detail in the preceding sections.[23]

DoD acquires a wide variety of commodities including surface ships, submarines, aircraft, missiles, ground combat vehicles, space systems, and numerous types of sensors and command & control systems. In the last two decades, systems became increasingly software intensive which, for most systems, enables and results in frequent block upgrades. This drives almost continuous test activity which may last for decades, in contrast to most programs in the 1950's-1970's which had discrete test program and workforce start and end points. There are some commonalities, but many differences, in the test practices among these systems. This is driven in part by the commodity, and in part by the Service acquisition and test organizational structures and culture. Some previous studies have focused on organizational change, with recommendations for more uniformity in the Service test organizational structures. However, the effectiveness of the government's involvement in DT is driven more by its role and related processes than by organizational construct.

Commodity type has a major influence on the government's role in the DT process. An illustrative contrasting example is aircraft versus large surface ships. In the former case, dedicated test vehicles are flown at government-operated ranges. In the latter case, there are no dedicated test vehicles. The mission systems are sometimes installed at a location separate from the shipbuilder's facility and system alpha and beta testing are accomplished on the first operational ship by the ship's crew, in conjunction with personnel from the responsible test activity. At the other end of the spectrum, DT on many command and control systems is accomplished in part or totally by the manufacturer at the contractor facility.

What is the appropriate test and evaluation role for the government in today's environment?

As discussed previously, there is no single organizational model for the Services to follow. However, there are many common elements where knowledgeable government test expertise is needed. The first is participation of government DT during RFP preparation to help define test assets, define test resources, and scope the test program. Related critical issues include ensuring that the RFP gives the government rights to test data acquired by the contractor as well as access to contractor developed models and simulations. The test community should then participate in the source selection process to evaluate the test programs proposed by the offerors.

The traditional responsibilities of the government during the DT planning phase have included development of the test strategy and detailed test plans, as well as the identification of the test resource requirements and government test facilities. The degree of involvement was driven by the commodity type and the contractual approach. In the past, a limited amount of DT planning and execution was accomplished primarily or exclusively by the government. With rare exceptions, this is no longer the case. However, there should be some level of government involvement and oversight even when the contractor has the primary responsibility regarding

[23] See Background sections B, C, and D.

planning and execution of the DT program. A recent GAO study[24] corroborates the requirement for increased government oversight. In the past, when the program moved from the planning phase to the test execution phase, the government participated in test conduct and analysis, supplying operators and engineers with DT personnel who performed an evaluation of the test results. These evaluations took many forms. Some government organizations published a report summarizing their analysis of the information. This is still accomplished by some Service DT organizations but has been discontinued by others. “Quick Look” rapid feedback briefings are sometimes used as reporting mechanisms. While these fulfill a useful purpose, they tend to be perishable and are not archived for future reference. The most powerful reporting tool for RAM issues is the DR. This report is still used during most developmental programs; however, deficiency correction, except for major problem areas, gets far less attention. One Service audit agency recently issued a report that criticized the lack of emphasis on addressing problems identified through the DR process. This de-emphasis is caused in part by the lack of a robust reliability program that includes reliability growth as an integral part of the systems engineering effort.

The implementation of some Acquisition Reform initiatives has had detrimental impacts on T&E. The current OSD guidance is to use performance specifications and commercial standards in preference to design-specific specifications and standards. There had been some poor applications of military specifications and standards in the past, in that they were not sufficiently tailored to the system to be procured. However, these standards had evolved and were updated over decades and were excellent compendiums of best practices and lessons learned. Many were used to provide guidance for system development to ensure a systematic and disciplined approach, and to ensure that problems that occurred in any new complex high technology system would be uncovered early. Since the start of Acquisition Reform, these documents have not been kept up to date in many cases. One organization estimated that 80% of their technical handbooks have atrophied in the last ten years. This information on processes and procedures will be lost over time unless the Services take aggressive action to preserve it.

The congressionally mandated cuts mentioned earlier had a major detrimental impact on the government acquisition and test community. The Services’ DT workforce reductions are discussed in more detail earlier in this report.[25] Significant reductions were made by all Services to the government program management and supporting engineering workforce. In some cases, these cuts decimated the workforce in certain technical domains. For example, the reliability engineering workforce at one acquisition product center was reduced from 50 in 1997 to 5 in 2007, and further cuts are anticipated in 2008. These are the personnel who define reliability criteria for incorporation into RFPs, shape the reliability growth programs for the program office, participate in deficiency reporting and other reviews, and evaluate the status of the reliability test effort during development. It is not surprising that the RAM results from IOT&E are so dismal, given the loss of government expertise in developing initial requirements and evaluating results. A specific case from another product center involves a current major joint Service acquisition for which there were no suitability requirements specified in the initial contract. Not surprisingly,

[24] United States Government Accountability Office, Best Practices: Increased Focus on Requirements and Oversight Needed to Improve DoD’s Acquisition Environment and Weapon System Quality, February 2008.
[25] See Background sections B, C, and D.

the reliability of the system was extremely poor and the system's software crashed so often during IOT&E that testing was terminated.

The large congressionally mandated cuts, when combined with high retirement rates has a significant detrimental effect on the experience level of the government acquisition/test workforce. This trend is driven by the retirement of a high percentage of the experienced workforce and the absence of a steady pipeline of younger technical personnel to replace them. As an example, Appendix I is a chart depicting near-term retirement eligibility for MRTFB personnel. Two-thirds or more of the senior civil service personnel are eligible for retirement. The loss of government personnel will have an additional impact on industry in the future, as it is a major training source for people who transfer to industry.

A 2008 National Research Council study[26] contains an in-depth discussion of the six drivers of cost development time and performance risks. It characterizes inexperienced government and industry personnel in key leadership positions as the largest risk. Other risk areas to avoid or manage include external interface complexity, system complexity, incomplete or unstable requirements in Milestone B, reliance on immature technology, and reliance on large amounts of new software.

FINDINGS

The aggregate lack of process guidance due to the elimination of specifications and standards, massive workforce reductions in acquisition and test personnel, acquisition process changes, as well as the high retirement rate of the most experienced technical and managerial personnel in government and industry has a major negative impact on DoD's ability to successfully execute increasingly complex acquisition programs.

- Major personnel reductions have strained the pool of experienced government test personnel
    - Government test and evaluation capability has been severely impaired in some acquisition organizations
- A significant amount of developmental testing is currently performed without the needed degree of government involvement or oversight and in some cases, with limited government access to contractor data
- The current trend away from governmental involvement in DT makes operationally oriented testing in some areas during development even more difficult to accomplish.

RECOMMENDATIONS

- As a minimum, government test organizations should develop and retain a cadre of experienced T&E personnel to perform the following functions:
    - Participate in the translation of operational requirements into contract specifications, and in the source selection process, including RFP preparation
        - Scope test program and define test assets
        - Develop proposed test program

[26] Paul Kaminski, et al, Pre-Milestone A and Early Phase Systems Engineering: A Retrospective Review and Benefits for Future Air Force Acquisition, 2008.

  - Participate in DT&E planning including Test and Evaluation Master Plan (TEMP) preparation and approval
    - Develop and implement government DT strategy, integrate operationally oriented test conditions, determine adequacy of government test support resources
  - Participate in technical review processes
  - Participate in test conduct, data analysis, and evaluation and reporting; with emphasis on analysis and reporting
    - Assess and report on suitability periodically during DT to ensure that suitability problems, particularly RAM, are identified early in the developmental process
- Utilize red teams, where appropriate, to compensate for shortages in skilled, experienced T&E domain and process experts
- Develop programs to attract and retain government personnel in T&E career fields so that the government can properly perform its role as a contract administrator and as a "smart buyer"
- **Action**
  - **AT&L, Service Acquisition/Test Organizations**

C. INTEGRATED TEST AND EVALUATION

Integrated T&E is not a new concept within the Department of Defense but its importance in recent years has been highlighted due in part to the growth of asymmetric threats and the adoption of net centric warfare. A December 2007 OSD Test and Evaluation Policy Revisions memorandum reinforces the need for integrated testing.[27] Implementation of the integrated test concepts have been allowed to evolve on an ad-hoc basis. The time has come to pursue more consistency in integrated test planning and execution. A workable definition of integrated testing is:

> The collaborative planning and collaborative execution of test phases and events to provide shared data in support of independent analysis, evaluation and reporting by all stakeholders particularly the developmental (both contractor and government) and operational test and evaluation communities.

The above definition was developed by T&E offices within OSD and the Services and is slated to be inserted into the next Defense Acquisition Guidebook update. The Task Force fully supports this initiative and recommends also incorporating the above definition into the next edition of DoDI 5000.2.

The current DoDI 5000.2 states: "The PM shall coordinate DT&E, OT&E, LFT&E, family of systems interoperability testing, information assurance testing, and modeling and simulation activities into an efficient continuum, closely *integrated* with requirements definition and systems design and development." This definition is ambiguous and misleading because it is not focused on integrating developmental and operational testing. Instead, it clouds the meaning by

[27] See Appendix D.

adding other factors such as Modeling and Simulation (M&S) and requirements definition into the concept.

Collaboration between the developmental and operational testers to build a robust integrated test and evaluation program will increase the amount of operationally relevant data that can be used by both communities. Collaboration is especially important for tools that apply across DT and OT, such as modeling and simulation (M&S). While analytical or functional models are not particularly useful for predicting component reliability, they are useful in the decomposition of requirements to lower assemblies (allocation of unreliability or failure rates) and for predicting system results based on component tests. Most developmental and operational tests should be preceded by M&S to predict test outcomes, with corrections to models and data made as required following a block of testing. A continuous and integrated M&S process from design through joint systems of systems test and evaluation is necessary to support efficient acquisition. While there have been rising expectations on the benefits of M&S and many studies conducted on how M&S can be improved, significant improvements in T&E applications have been lacking.

The current policy and organizational structure effectively separates DT and OT planning, thereby inhibiting efforts by the Services to streamline test schedules. It also increases the acquisition timeline and the costs of the programs and tests. A recent National Research Council report on testing concluded that "the current distinction between operational and developmental testing will have to be reconsidered. A new paradigm, better coordinated between the two approaches, would allow system development to benefit from a more continuous, strategic, approach to testing…"[28] The report went on to recommend the USD (AT&L) develop policy for more developmental tests to "have an operational perspective in order to increase the likelihood of early identification of operational failure modes and system deficiencies…"[29] Title 10, Section 2399 requires DOT&E approval of test plans for IOT&E. However, by infusing operationally relevant mission profiles and an operationally representative environment during DT, many of the data requirements needed by the OTA could be obtained prior to IOT&E. OTAs should be encouraged to proactively review and use operationally relevant data (obtained during DT&E) to support their evaluation of a system's effectiveness and suitability.

A number of issues need to be addressed to better integrate DT and OT. First, current OSD policy (DoDI 5000.2, paragraph 3.7.6.) prohibits the OTAs from using data obtained during DT until DOT&E approves the OT&E portion of the combined DT/OT plan. Therefore, when an integrated plan is developed, it requires DOT&E approval before the OTAs can use any data. The bureaucratic process of an additional OSD approval is a major impediment. OTAs should be permitted to influence the scope of DT and use test data that they deem operationally relevant to reduce the scope of dedicated OT.

Additionally, there is a widely held belief by many, particularly in the OT community, that only data from independent OT is acceptable for operational evaluation purposes. Instead, the OTAs must be encouraged to capitalize on all data that is operationally relevant. While not all information from DT may be useable by the OTAs to support IOT&E, a significant amount of

[28] National Research Council, Testing of Defense Systems in an Evolutionary Acquisition Environment, Washington DC (The National Academies Press, 2006), 20.
[29] IBID.

developmental test data can be used to partially satisfy OT requirements. More importantly, operational influence and perspective earlier in the developmental process is a proven catalyst for early identification and correction of problems. Also, the OT community must have better access to DT data and selected models, including those developed by the prime contractor. This is an issue that needs to be addressed during program formulation. Finally, integrated DT and OT planning should be a requirement and the strategy to execute the integrated program should be outlined in the TEMP.

A related issue facing the OT&E community is the magnitude of the test and test support task. Integrated testing requires early involvement and a skill set that may not reside with the OTAs. On the other end of the spectrum, there is a need to test in a joint forces environment. These two requirements, when combined, strain an already limited operational test resource. Nevertheless, early influence and limited involvement has the potential of paying large dividends and can be done by a relatively small cadre at selected points in the acquisition cycle. One benefit would be a reduction in the amount of follow-on OT&E necessary after IOT&E.

To achieve maximum integration of DT and OT it is imperative for programs to establish a team made up of all relevant organizations (including contractors, developmental and operational test communities) to create and manage the process to incorporate integrated testing into the T&E Strategy and the TEMP. This team must be established as early as possible in the program, preferably during the concept refinement phase, to effectively identify test parameters, data, and resources required for the DT and OT plans and other required certifications (e.g., interoperability, system assurance, safety) to optimize test data collection while minimizing test resource requirements. The intent is to increase the overall efficiency of testing, improve product performance and decrease the acquisition timeline. The Milestone Decision Authority (MDA) should provide formal direction to establish the test team in the program's first Acquisition Decision Memorandum. Participation in the integrated test planning and execution by industry will need to be included in the RFP and subsequent contract.

FINDINGS

- Service acquisition programs are incorporating integrated testing to a limited degree through varying approaches
  - Army has integrated DT and OT organizations into one command
    - Integrated testing is possible without major organizational change
  - Navy utilizes a full-spectrum RDT&E approach to conducting T&E
    - Integrated test teams (comprised of contractor & government, DT, and OT personnel) participate in all phases of acquisition programs
  - Air Force employs Combined Test Force (CTF) concept which consolidates test execution
- Additional emphasis on integrated testing can result in greater T&E process efficiency and program cost reductions

RECOMMENDATIONS

- Implement OSD and Service policy[30] mandating integrated DT&E and OT&E planning and execution throughout the program

[30] See Appendix D.

    - Require sharing and access to all appropriate system-level and selected component-level test and model data by government DT and OT organizations as well as the prime contractor, where appropriate
    - Incorporate data access requirements in contract
    - Integrate test events, where practical, to satisfy OT and DT requirements
    - Define which testing will be accomplished by the prime contractor, government DT lead, and OT as the lead agency prior to award of contract
    - Require an operational evaluation framework as a part of the Milestone B TEMP
- Make available a cadre of operational personnel to support DT for Acquisition Category I (ACAT I) and special interest programs, as a minimum
- The MDA should provide formal direction to establish the test team in the initial Acquisition Decision Memorandum.
    - Industry participation in integrated test planning and execution should be included in the RFP and subsequent contract.
- Better integrate OTAs into the DR process to include participation on Joint Reliability Maintainability Evaluation Team (JRMET) or Corrective Action Review Board throughout DT
- **Action**
    - **DOT&E, OSD (AT&L), Service Acquisition Organizations, OTAs**

D. OPERATIONAL TEST READINESS REVIEW (OTRR)

I. ARMY OPERATIONAL TEST READINESS REVIEW (OTRR)

For programs on the OSD T&E oversight list, the Army Acquisition Executive (AAE) is required to evaluate and determine materiel readiness for IOT&E. This is to preclude systems from entering IOT&E prematurely by ensuring that they have demonstrated technical maturity under operational conditions. The materiel developer must provide OSD a DT&E report and a progress assessment that supports entry into IOT&E. This assessment can be a written document or a briefing to DOT&E and the USD(AT&L) (Systems Engineering/Assessments and Support).

In addition to the above process, ATEC conducts OTRRs. The OTRR is a forum where the designated operational tester and other members of the ATEC System Team (AST) bring together the representatives of all agencies associated with an event to determine overall readiness to begin the event.

An operational test may require as many as three OTRRs. At each OTRR, the evaluator will provide an assessment of the readiness of the tester to conduct the test and the readiness of the system to enter the test. The status of the TEMP will also be discussed at each OTRR.

OTRR 1 may be held in conjunction with a system T&E Working Integrated Product Team (WIPT) and should occur approximately 270 days before the scheduled test event. This is normally an action officer review. OTRR 2 is usually conducted 60 days before the scheduled test event, just before commitment of major resources for the event. The purpose of OTRR 2 is to confirm availability of required resources and review the planning status. This is the test resourcing OTRR and is conducted to identify any known problems that would delay test start. OTRR 3 is normally conducted after the pilot test, usually between one to five days before the scheduled test event. The purpose is to identify if the tested system, players, testers, instrumentation, and data collection and reduction procedures are ready for testing for record. All prerequisites for successful test execution are reviewed at this OTRR to confirm record testing is ready to begin.

II. NAVY OPERATIONAL TEST READINESS REVIEW (OTRR)

The formal review and certification process for all operational testing periods (including Operational Assessments (OA), Operational Evaluations (OPEVAL) and Follow-on Operational Test and Evaluations [FOT&E]) entails a pre-DT&E WIPT meeting, a post-DT&E WIPT meeting, a pre-OTRR, and an OTRR. The pre-DT&E WIPT provides for the initiation of the planning process and is used as a forum to discuss issues and DT test plans. The post-DT T&E WIPT allows for data review, preparation of the DT/OT transition report, the drafting of the certification message, use of the OTRR certification checklist, and identification and resolution of issues and deficiencies associated with the DT test period. The pre-OTRR evaluates compliance with all Secretary of the Navy (SECNAV) Instruction 5000.2C certification criteria and makes a recommendation regarding the system's readiness for an OTRR. The OTRR is a final review by the DA prior to the DA's decision to certify the system for OA, OPEVAL or FOT&E.

III. AIR FORCE OPERATIONAL TEST READINESS REVIEW (OTRR)

Air Force Manual 63-119, *Certification of System Readiness for Dedicated Operational Test and Evaluation*, describes the process the Air Force uses to monitor the progression of an acquisition program toward OT&E and to document the decision by the appropriate decision authority. In this case the decision authority is the PEO who decides if the weapon system under development is ready to proceed to dedicated OT&E. It does so by providing a structured mechanism for identifying and reducing risks associated with transitioning from DT&E to dedicated OT&E. In principle, it establishes a review and "certification process" beginning in the early stages of an acquisition program. It is a tool to help acquisition managers at all levels identify risks, reach negotiated agreements on issues, and render more accurate assessments of system readiness to begin dedicated OT&E. A standard framework or process is detailed in 33 templates, which contain information and advice about how to reduce risk. Air Force Manual 63-119 was last published in February 1995 and it would be prudent to complete a thorough review and update of the manual, with additional emphasis on the reliability and maintainability engineering process—currently addressed in the deficiency resolution template.

OTRR FINDINGS

- DoDI 5000.2 requires that "the Service Acquisition Executive (SAE) shall evaluate and determine materiel system readiness for IOT&E"
  - For all Services, the OTRR process is an in-depth review of readiness
    - Actual process varies from Service to Service
    - Some Services do not currently use a detailed RAM template as a part of the review process
  - Decision authority is frequently delegated to the appropriate PEO
  - Materiel developer is also required to furnish a DT&E report to USD(AT&L) and DOT&E

- Shortcomings in system performance, suitability, and RAM are usually identified during the OTRR
  - In some cases, the developmental test schedule is so compressed that there is inadequate time to evaluate DT results prior to the review
- In most cases, the operational test readiness certifying authority is aware of the risk of not meeting OT criteria when major shortcomings exist
- Because of funding constraints, the low priority given to sustainment, as well as the urgency in recent years to get new capabilities to the Warfighter, major suitability shortcomings have rarely delayed the commencement of dedicated IOT&E

OTRR RECOMMENDATIONS

- Conduct periodic operational assessments to evaluate progress potential for achieving pre-determined entrance criteria for operational test events
  - **Action: PEOS/Program Management Offices (PMOs)**
- Conduct an independent AOTR prior to the OTRR (included in latest draft DODI 5000.2) for oversight programs
  - Results of the AOTR should be briefed at the OTRR
  - **Action: OSD (AT&L)**
- Include a detailed RAM template in preparation for the OTRR
  - **Action: Service Acquisition Organizations**
- Require the CAE to submit a report to OSD that provides the rationale for the readiness decision
  - Include an evaluation of weapon system's capabilities against Critical Operational Issues (COIs)
  - Certify that the DT evaluation is complete and indicates acceptable risk of passing OT
  - Include an explanation for recommending go-ahead when there is a reasonable expectation that the system is not effective and/or suitable
  - **Action: Service PEOs**

E. OSD Test and Evaluation Organization

The Task Force was asked to assess OSD roles and responsibilities for T&E oversight. T&E has been a visible part of the Office of the Secretary of Defense since the early 1970's, reporting to the Research and Engineering command section when it was in charge of acquisition oversight and subsequently to the Under Secretary of Defense for Acquisition (now AT&L). For a short period, OT&E was part of Program Analysis and Evaluation (PA&E). The early T&E office was responsible for all T&E, ranges, resources oversight, and policy. In 1983, Congress established an independent DOT&E organization, reporting directly to the Secretary of Defense, responsible for operational test and evaluation policy, budget review and assessments of operational effectiveness and suitability. The LFT oversight function was created by Congress and added to the DT&E office responsibilities in the mid 1980's. Later, the LFT oversight function was moved to the DOT&E organization. Appendices F and G depict the acquisition organization in

1992 and 2007 respectively. In the mid 1990's, the systems engineering function was established and incorporated in the DT&E office.

In 1999, the DT&E organization was dismantled by DoD. Many functions were moved to DOT&E, including test ranges and resources, and joint T&E oversight. Some of the remaining personnel billets were eliminated to comply with a congressionally mandated (AT&L) acquisition staff reduction. The residual DT&E policy and oversight functions were separated and moved lower in the AT&L organization.

Currently, there is no single OSD organization with comprehensive DT&E oversight responsibility and authority. The existing residual organizations are fragmented, with DT&E as one of many responsibilities, and lack experience and clout to provide useful and consistent guidance to individual acquisition programs. There is also a disproportionate emphasis on facility oversight and lack of interaction between facilities and program oversight organizations. Another result of the dismantling of the DT&E function was that DT&E was no longer viewed as a key element in AT&L system acquisition oversight. One impact has been a gradual loss of access to T&E data and analysis results for programs on the OSD T&E oversight list. Another impact has been the loss of thorough assessments of system maturity and risk mitigation, based on the status of DT, during major programmatic reviews.

Test facilities and resources have been a part of the OSD T&E oversight responsibilities since the 1970's. A 1999 DSB Task Force Study of Test and Evaluation Capabilities[31] recommended that DoD create a test and evaluation resource enterprise within the office of the DOT&E to provide more centralized management of T&E facilities. This recommendation ultimately led to many significant changes including the removal of test ranges and resources oversight from DOT&E, the abandonment of the notion of centralized management and the establishment of the Test Resource Management Center (TRMC) in AT&L (as directed by the National Defense Authorization Act for FY2003.) TRMC responsibilities include the oversight of T&E facilities and resources, which was originally in the OSD DT office, and later in the DOT&E office. The TRMC developed and bi-annually updates a strategic plan to reflect the T&E resource needs of DoD, and to guide the Services and Defense Agencies in planning and budgeting for future capabilities. The TRMC also reviews component budgets to certify their compliance with this plan. The TRMC strategic plan for T&E capability investments is driven by current test requirements as well as emerging weapon system technologies. It covers the upcoming Program Objective Memorandum (POM) years plus five. Because the TRMC is not involved in reviewing TEMPs, it has no systematic mechanism for determining the adequacy of near term programmatic resource requirements. The TRMC should review the resources section of the TEMP and incorporate the results in their strategic planning process. Separation of the TRMC from the other elements of OSD T&E oversight contributes to a fragmented approach to the

[31] Defense Science Board, Report of the Defense Science Board Task Force on Test and Evaluation Capabilities, Washington, D.C., December 2000.

performance of the oversight function. The Task Force considered including the TRMC in the recommendation to consolidate the T&E-related functions in AT&L; the majority of the task force government advisors voted in favor of the inclusion, however, the majority of the task force members did not.

When Congress directed DoD to establish OSD live fire oversight in the mid 1980's (as a result of survivability problems with the Bradley fighting vehicle and other platforms), the function was established initially in the DT office. It was moved to DOT&E in the mid 1990's. Live Fire Test and Evaluation (LFT&E) identifies potential areas of vulnerability and lethality early in the developmental process when programs can address and incorporate improvements. Later, it contributes to operational assessments that include survivability and lethality. LFT consists of technical, physics-based testing. It typically employs a building-block approach, progressing from early component-level LFT&E, to subsystem/system level testing, and sometimes followed by full-scale LFT. The building-block approach provides the earliest possible understanding of munitions/target interaction phenomena during development and enables early fixes to identified problems. LFT also addresses the physics of various damage mechanisms, including cascading damage, and critical system vulnerability and lethality issues. It provides the program manager with a step-wise approach to acquire test information to enhance the system design during the development process. Programs conduct LFT&E at specialized government live fire test facilities, supplemented at times by testing at contractor facilities. The LFT directorate is also responsible to provide warfighters with information on the lethality of systems by maintaining and updating the Joint Munitions Effectiveness Manuals (JMEMs). There appears to be little, if any, involvement in LFT&E by the OTA's, except for limited use of LFT vulnerability data to augment susceptibility test data for the purpose of developing an indicator of system survivability. The Army is the exception in that it has had a fully integrated Test and Evaluation Command (ATEC) since 1999. The LFT process has matured in the past decade and no specific problems have arisen as a result of the move of LFT to DOT&E. The Task Force was not briefed by the LFT office, and at the present time its effectiveness seems to benefit from it being in the relatively more visible and powerful DOT&E organization. However, because of the developmental nature of LFT&E and the close connectivity to the responsibilities of the Service acquisition and DT organizations, consideration should be given to moving it after a robust DT organization is reestablished in AT&L.

FINDINGS

Current policy, as of December 2007, mandates that developmental and operational test activities be integrated and seamless throughout the system life cycle. There must be experts in OSD who are able to understand and articulate the lessons learned in early testing and execute the new T&E policy. That policy is to "take into account all available and relevant data and information from contractors and government sources" in order to "maximize the efficiency of the T&E process and effectively integrate developmental and operational T&E."

- Currently there is not an OSD organization with comprehensive DT oversight

responsibility, authority or staff to coordinate with the operational test office.
  - The historic DT organization has been broken up and residual DT functions were moved lower in organization in 1999, and lower yet in 2002
  - Programmatic DT oversight is limited by staff size and often performed by generalists vice T&E experts
  - Recruitment of senior field test personnel is hampered by DT's organizational status
  - Existing residual organizations are fragmented and lack clout to provide DT guidance
  - System performance information and DT lessons learned across DoD has been lost
  - DT is not viewed as a key element in AT&L system acquisition oversight
  - Documentation of DT results by OSD is minimal
- Access to models, data, and analysis results is restricted by current practice in acquisition contracting, and the lack of expertise in the DT organization
- The TRMC has minimal input to program-specific questions or interaction with oversight organizations on specific programs
  - Organizational separation is an impediment

RECOMMENDATIONS

- Implementation of integrated and seamless DT and OT will require, at a minimum, greater coordination and cooperation between all testing organizations.
- Consolidate DT-related functions in AT&L to help reestablish a focused, integrated, and robust organization[32]
  - Program oversight and policy, and FCT
  - Have Director, DT&E directly report to DUSD(A&T)
  - Restore TEMP approval authority to Director, DT&E
- Integrate TRMC activities early into DT program planning
  - Make TRMC responsible for reviewing the resources portion of the TEMP
- If such an organization is established and proves itself effective, consider as part of a future consolidation moving LFT back to its original DT location (this would require congressional action and DOT&E concurrence)

CHANGES TO STATUTORY AND REGULATORY AUTHORITY

RECOMMENDATIONS

- In addition to the organizational changes previously recommended, USD(AT&L) should add the following to DoDI 5000.2, Enclosure 5:

  An assessment of the results of a system's developmental test and evaluation program is required before the system enters the design and development phase and Low Rate Initial Production (LRIP). The assessment will be accomplished by the office in AT&L responsible for developmental testing. This assessment will be briefed to the Milestone

[32] Three Task Force members out of fourteen voted against consolidation

Decision Authority for Major Defense Acquisition Program (MDAP) programs in conjunction with each Milestone review.

Addition to Table E3.T2.:

Information Required	Source	When Required
OSD (AT&L) Assessment of Developmental Test Results	This instruction	MS B/C

VI. OTHER ISSUES

A. PROGRAM STRUCTURE

Inadequate funding and unrealistic schedules are common reasons for the lack of proper attention to setting and achieving appropriate operational suitability performance requirements, and in some cases, even program delays. In many cases, program funding and schedules are not sufficient to accommodate technical complexity, identified program risks, and critical path planning and execution.

Sufficient resources (time, money, and personnel) must be allocated and available for a program to successfully develop it within the program baseline. It is important that a program obtain and sustain funding to support its core program of work, including all dimensions of required operational effectiveness and operational suitability. This flow of funding needs to be stable and steady. The funding (amount and profile) and schedule duration to perform all the planned activities should be determined by systematic estimating methods which may include past completed program cost and schedule 'actuals' (history), independent cost estimates, etc., and be consistent with the program's Integrated Master Plan.

FINDINGS

- Operational suitability shortcomings are affected by common occurrences of programs being started without adequate funding and schedule reserves to address known and unknown technical issues. Funding is unstable, and timelines are success-oriented and schedule-driven vice event-driven
  - In some cases, funding types (RDT&E, Procurement, Operations and Management [O&M]) are not phased properly
  - Schedules are not based on an analysis of the technical efforts that are required for the program to be successful

RECOMMENDATIONS

- DoD policy should require the program's system engineering strategy be accommodated when establishing funding and schedule requirements for SDD
- Require that the technical strategy be documented and approved prior to release of an RFP to ensure the contractual requirements accurately reflect the needed technical effort
- **Action**
  - **OSD(AT&L), Service Acquisition Organizations**

B. REQUIREMENTS DEFINITION

Operational requirements set the foundation for the system acquisition process by stating the operational and support-related performance attributes of a system(s) that provide the capabilities required by the warfighter – attributes so significant they must be verified by testing or analysis. If operational and support-related performance requirements are not consistently stated in terms that are testable, quantifiable, measurable, and reasonable, it becomes very difficult or impossible to verify that the system design and development has been successful in providing a desired operational capability.

If operational requirements are not properly defined at the onset of the program, the design baseline may not reflect the users' intended requirement. The program office may have incorrectly interpreted the operational requirements and carried forward that misinterpretation into the design specification. This can lead to disputes at the time of test planning and test analysis concerning the intent of the operational requirement and the capability of the weapon system.

FINDINGS

- Operational requirements are not consistently stated in terms that are testable, quantifiable, measurable, and reasonable
  - If operational requirements are not properly defined at the onset of a program, the design baseline may not reflect the user's intended requirement
  - The program office sometimes misinterprets the operational requirements and carries that misinterpretation into the design specifications

RECOMMENDATIONS

- Establish procedures to help ensure that operational requirements are testable or evaluable, quantifiable, measurable, and reasonable throughout the JCIDS process
  - Involve DT&E and OT&E early in the requirements definition process
  - Make user representatives available to the program office to help resolve any poorly defined requirements
- **Action**
  - **Joint Staff, OSD(AT&L), DOT&E, Service Acquisition Organizations**

C. CONTRACTUAL PERFORMANCE REQUIREMENTS

It is imperative that the Services effectively and accurately translate operational requirements into contractual requirements. Testing and deployment environments in mission profiles and in CONOPs should be taken into account in the contractual requirements. The lack of accuracy and completeness in translating an operational requirement, or the misrepresentation of that requirement, in the contractual specification that governs the SDD effort can cause many problems. These include shortfalls in achieving the requisite systems operational effectiveness and operational suitability, and the possible delay or limitations in the fielding of a needed military capability, which could ultimately lead to a waste of resources.

The SRR is a multi-disciplined, technical review to ensure the contractor's understanding of user requirements are adequately reflected in contractual documents. The SRR ensures that all system requirements and performance requirements derived from the Initial Capabilities Document or draft Capability Development Document are defined in adequate detail and are consistent with cost (program budget), schedule (program schedule), risk, and other system constraints. Generally, this review assesses the system requirements as captured in the system specification, and ensures that the system requirements are consistent with the preferred system solution as well as available technologies resulting from the technology development phase. During the SRR, the systems requirements are evaluated to determine whether they are fully defined and consistent with the mature technology solution, and whether traceability of systems requirements to the Initial Capabilities Document or draft Capability Development Document is maintained. A successful review is predicated on the determination that the system requirements, preferred system solution, available technology, and program resources (funding, schedule, staffing, and processes) form a satisfactory basis for proceeding into the System Development and Demonstration (SDD) phase.

Compliance teams are often used to review commercial telecommunications and aircraft designs to assure that best practices and procedures are used. Compliance teams in the DoD context could be formulated with appropriate Federally Funded Research and Development Centers (FFRDCs) and government employees to review basic requirements as well as participating in early design reviews.

FINDINGS

- Program offices sometime lack the ability to effectively translate operational requirements into contractual requirements, by not taking into account the differences in testing and deployment environments, failure definitions, etc.
  - Any errors in this translation impact the design and development efforts and generally become evident in OT, especially in the case of RAM requirements
  - Application of an effective systems engineering process can minimize the likelihood of disconnects in operational and contractual requirements

RECOMMENDATIONS

- Systems engineering tools and techniques (e.g., internal government system requirements review) should be applied by Service acquisition and engineering organizations prior to contract execution to ensure operational requirements are accurately reflected in design-related contractual requirements. The training and tools must be readily available to all parties.
- Implement the commercial practice of separate compliance teams for basic requirements and early design reviews
- **Action**
  - **OSD(AT&L), Service Acquisition Organizations**

D. ALIGNMENT OF DOD TERMINOLOGY WITH SYSTEMS ENGINEERING PROCESS

Lack of accuracy and completeness in translating an operational requirement, or misrepresentation of that requirement in the contractual specification that governs the system design and development effort will result in shortfalls in achieving the requisite systems operational effectiveness and suitability. This misrepresentation often manifests itself as poor performance during the IOT&E. To improve the operational performance in the operational environment, it is necessary for the Services to effectively and accurately translate operational requirements into contractual requirements, taking into account variations in testing/deployment environments, in mission profiles and in CONOPs. It is then imperative that these requirements be defined in a manner that they are testable (across the various environments) and are traceable to mission success. The systems engineering process provides a structured methodology to decompose operational requirements into technical specifications which can become the basis for systems development contracts. Unlike the systems engineering process, the decomposition of operational requirements into evaluation metrics is not well structured and is usually based on some level of military judgment.

The most significant disconnect between technical requirements definition and operational performance of the final product can be traced to the terminology used to define the level of performance required, under what conditions and how it is measured. The systems engineering process allows the use of test, demonstration, analysis and inspection as valid methods to collect relevant data to determine actual or predicted performance in relationship to the specification requirement. The metric is often reflected through terms like qualification of a design, verification of performance, validation of results and, compliance to the specification requirement. The operational measures, for the same level of performance are effective, suitable, and survivable which imply a level of subjectivity. A level of performance resulting in meeting the technical requirements of the specification may result in a system that is not effective, suitable or survivable when the operational environment is factored in. By aligning the systems engineering and the test and evaluation planning processes through a common terminology definition, the results of developmental testing can become a better predictor of operational performance during the IOT&E.

FINDINGS

- Inconsistent terminology used to define the level of performance, conditions and measurements, can contribute to differences between technical requirements definition and operational performance of the final product
  - The systems engineering process provides a structured methodology to decompose operational requirements into technical specifications which can become the basis for systems development contracts.
  - The decomposition of operational requirements into operational test related evaluation metrics is not well structured and is usually based on some level of military judgment.

RECOMMENDATIONS

- Align the systems engineering and the test and evaluation planning processes through a common terminology definition and ensure the common terminology is reflected in DoD and Service level policy and guidance.
- During the System Requirements Review, ensure through the participation of the engineering and test communities that operational requirements are accurately reflected in design-related contractual requirements and are consistent with planned operational test criteria for assessing operational effectiveness and suitability.
- **Action**
  - **OSD(AT&L)**

E. COMMERCIAL OFF-THE-SHELF (COTS)

Commercial-Off-The-Shelf (COTS) components, systems, and subsystems have become inherent in the defense industry. Often, commercial equipment that will effectively perform in military applications is not readily available in today's marketplace. Commonly, COTS will not stand up to the full spectrum of thermal, vibrational or other environmental demands needed in the military application. In these situations, protection is required to make certain that COTS systems are adequately isolated and protected so they can properly perform.

Too often COTS which do not meet the application requirements are used. This leads to higher than anticipated failure rates and poor system mission reliability. Program managers must ensure that COTS components are able to operate satisfactorily in military mission environments. Two excellent detailed guidance manuals for the use of COTS items are:

- "Selection of Equipment to Leverage Commercial Technology (SELECT) User Manual", by David Nicholls, David Clark(Reliability Analysis Center (now the Reliability Information Analysis Center [RIAC], June 1998)
- "Evaluating the Reliability of Commercial Off-The-Shelf (COTS) Items" by Ned H. Crisimagna (Reliability Information Analysis Center, , August 1999).

Problems have also arisen on programs because of the use of Government Furnished Equipment (GFE) in a different environment or application. All subsystem components being considered must be evaluated based on the intended application and not be excluded from program criteria because of a prior less stressful application.

FINDINGS

- COTS components are used in all defense systems today at a variety of system and subsystem levels
  - COTS components are selected to reduce cost in systems, often without proper planning to accommodate challenges relating to reliable mission performance
- COTS capabilities may not be properly matched with the full spectrum of system mission requirements and the related operating environments.
  - COTS, regardless of Work Breakdown Structure (WBS) level, must either meet system requirements (e.g., g forces) or the system must be designed in such a way as to protect COTS from the more demanding military environment
- Integrating COTS as a subsystem frequently requires no less design integration and test substantiation than a newly designed subsystem
- Programs have been delayed or terminated because of a disconnect between requirements and actual capabilities of COTS subsystems or systems for military applications

RECOMMENDATIONS

- Require that all components (including COTS and GFE) be analyzed and selected based on the full spectrum of operational needs and expected environments
  - Entire system must meet all key mission performance requirements, including RAM, in an operational military environment
  - Structure COTS-related program acquisition strategies based on a low-risk approach to ensure adequate resources
  - Conduct gap analysis to address resource requirements for COTS program execution (e.g., modification, integration, qualification, testing in an operational environment)
- Ensure quick-reaction COTS shortcuts do not allow unqualified or unproven devices with lesser capabilities to be incorporated without a detailed risk assessment
- **Action**
  - **Service Acquisition Organizations**

F. SYSTEMS OF SYSTEMS

The Department of Defense is developing and fielding capabilities through both single systems and Systems-of-Systems (SoS). However, the current acquisition process and methodologies are, for the most part, optimized to single system acquisitions. In general, requirements are organized, budgets are programmed, program charters are established, and contracts are awarded to acquire single systems. SoS represent a larger scale integration of systems across multiple organizations to deliver complex interdependent capabilities. Deficiencies in individual systems RAM can substantially restrict SoS capabilities. The challenge for the T&E community is to test performance which involves multiple major systems, with varying engineered interrelationships, across multiple organizational boundaries. Shortcomings in SoS requirements and CONOPs hamper rigorous and conclusive T&E. No single program manager "owns" the performance or the verification responsibility across the multiple constituent systems, and there is no widely used adjudication process to readily assign responsibility for SoS capabilities, with the exception of command and control systems. In addition, arranging the operational resources to support SoS test events is particularly difficult. This compounds the testing challenge, because of the need to assemble a joint T&E team at the outset of testing.

The testing challenge is compounded by resource constraints. Ideally, all of the individual resource sponsors would provide funding and the resources themselves would be available when needed to execute a SoS test. In reality, scheduling limitations, asset non-availability, and insufficient funding impede the execution of a test program to adequately assess the effectiveness and suitability of SoS capabilities. Additionally, the management challenge of coordinating test strategies and plans, and resolving responsibility or accountability concerns between the participating acquisition programs, has often proven to be impossible.

FINDINGS

- The fact that no single manager owns overall performance and verification across multiple constituent systems is an impediment to achieving required operational suitability and effectiveness in SoS.
- Acquisition process generally is tailored to the development of individual systems, not SoS.
- Some mission critical testing of SoS capabilities areas often accomplished for the first time in OT or when deployed
- Testing all the SoS requirements of all the systems is impossible
- SoS requires better approaches to requirements definition, resources, and acquisition processes, including judicious T&E.

RECOMMENDATIONS

- The Portfolio Capability manager should be assigned to structure developmental programmatic responsibilities and to authoritatively and expeditiously adjudicate jurisdictional questions as issues arise.
- Initiate a study of current SoS programs to document lessons learned and best practices.
- Formulate alternative strategies to adapt current requirement, acquisition, and funding processes to enable timely, efficient, and effective T&E of SoS capabilities.
- **Action**
    - **OSD(AT&L), DOT&E, Service Acquisition Organizations**

APPENDIX A: TERMS OF REFERENCE

THE UNDER SECRETARY OF DEFENSE
3010 DEFENSE PENTAGON
WASHINGTON, DC 20301-3010

ACQUISITION, TECHNOLOGY AND LOGISTICS

APR 30 2007

MEMORANDUM FOR CHAIRMAN, DEFENSE SCIENCE BOARD

SUBJECT: Terms of Reference -- Defense Science Board (DSB) Task Force on Developmental Test and Evaluation (DT&E)

You are requested to establish a DSB Task Force on DT&E.

Approximately 50 percent of programs entering Initial Operational Test and Evaluation (IOT&E) in recent years have not been evaluated as Operationally Effective and Operationally Suitable. Although numerous DoD initiatives were developed to drive better systems engineering practices and life cycle management principles into acquisition programs, the IOT&E failures suggest deficiencies in DT&E processes. The Quadrennial Defense Review and associated studies precipitated changes to DoD acquisition approaches, to include time-certain acquisition. However, within the context of emerging approaches, DT&E processes remain mainly unchanged. Accordingly, I request the DSB establish a task force to examine T&E roles and responsibilities, policy and practices, and recommend changes that may contribute to improved success in IOT&E along with quicker delivery of improved capability and sustainability to Warfighters.

The Task Force should examine, from a T&E perspective, OSD organizational roles and responsibilities, policy and practices in oversight of acquisition programs. The Task Force should assess:

- OSD organization, roles, and responsibilities for T&E oversight. Compare organization, roles, and responsibilities in both DT&E and OT&E. Recommend changes that may contribute to improved DT&E oversight, and facilitate integrated T&E.

- Changes required to establish statutory authority for OSD DT&E oversight. Title 10 USC has an OT&E focus and does not address OSD authority in oversight of DT&E. Recommend changes to title 10 or other U.S. statutes that may improve OSD authority in DT&E oversight.

- Many IOT&E failures have been due to lack of Operational Suitability. Specific problems have been in the materiel readiness sustainment areas of reliability, maintainability, and availability. Recommend improvements in

DT&E process to discover suitability problems earlier, and thus improve likelihood of operational suitability in IOT&E.

The Under Secretary of Defense for Acquisition, Technology and Logistics, Deputy Under Secretary of Defense for Acquisition and Technology, and Director, Operational Test and Evaluation, will sponsor this study. Mr. Charles "Pete" Adolph will Chair the Task Force. Mr. Chris DiPetto, ODUSD(A&T), and Dr. Ernest Seglie, ODOT&E, will serve as Executive Secretaries. Major Chad Lominac, USAF, will serve as 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 title 18, U.S. Code, section 208, nor will it cause any member to be placed in the position of acting as a procurement official.

Kenneth J. Krieg

2

APPENDIX B: TASK FORCE MEMBERS & GOVERNMENT ADVISORS

Mr. Charles (Pete) Adolph, Chairman

Dr. Ernest Seglie, Executive Secretary

Mr. Chris DiPetto, Executive Secretary

Dr. Larry Crow, Member

RADM (Ret) Bert Johnston, Member

Mr. Richard Johnston, Member

VADM (Ret) John Lockard, Member

Mr. Raymond Lytle, Member

MG (Ret) Wilbert (Doug) Pearson, Member

Mr. Matt Reynolds, Member

Dr. George Schneiter, Member

Mr. Gregory Shelton, Member

Mr. Michael Toole, Member

Mr. George Williams, Member

Dr. Marion Williams, Member

Mr. Mike Zsak, Member

Dr. John Claxton, Government Advisor

Mr. Edward Greer, Government Advisor

Dr. David Jerome, Government Advisor

Ms. Darlene Mosser-Kerner, Government Advisor

Mr. Tom Simcik, Government Advisor

COL Richard Stuckey, Government Advisor

Dr. James Streilein, Government Advisor

Support Staff

Ms. Michelle Ashley, SAIC
Mr. M. Brett Sterlacci, SAIC
Ms. Lauren York, SAIC

APPENDIX C: BRIEFINGS RECEIVED

31 May 2007

OSD/DT Perspective	Mr. Chris DiPetto	OUSD (AT&L)/DS
Marine OTA Perspective	Col Michael Bohn	USMC
DOT&E Presentation	Dr. Ernest Seglie	OSD

1 June 2007

Navy OTA Perspective	RADM William McCarthy	Navy
Army OTA Perspective	MG James Myles	ATEC

27 June 2007

PM FCS Reliability Improvement Plan	Mr. Tom Hartigan	PM FCS (BCT)
Past Performance: Historical Perspective across Commodities	Mr. James Johnson	US Army Developmental Test Command
Present Direction: Systems of Systems Testing and Software Blocking	Mr. Neil Brown	ATEC/AEC
Future Challenges: Net-Enabled Command Capability	Mr. Mike Dabney	Army CPMO
AFOTEC Brief	Mr. Jerry Kitchen	AFOTEC Technical Advisor
The Cost of Unsuitability	Dr. Jim Forbes	LMI
OPEVAL is No Longer Enough	Mr. Brian Willoughby	BMP Center of Excellence
Reliability Overview	Dr. Larry Crow	Crow Reliability Resources, Inc

28 June 2007

T&E in Air ASW, Assault and Special Mission Programs	RDML(SEL) David Dunaway	Deputy PEO (A)
T&E in Carrier Programs	RADM David Architzel	PEO Carriers
T&E in Naval Aviation	Mr. Steve Cricchi	NAVAIR Integrated Systems Evaluation, Experimentation and Test Department
T&E in Submarine Programs	RDML William H Hilarides	PEO Subs
Task Force Sponsor Perspective	Dr. Charles McQueary	DOT&E

Air Force Developmental Test and Evaluation Perspective	Brig Gen David J. Eichhorn	Director of Air, Space and Information Operations, Air Force Materiel Command

25 July 2007

FOA Overview, Benefits, and Lessons Learned	COL Mark Mills	AEC
Systems Engineering Systemic Analysis Brief	Mr. Dave Castellano	OSD
Navy OTA Perspective	Mr. Steve Whitehead	COMOPTEVFOR
Navy Operational Test Readiness Review (OTRR)	Mr. Ed Greer	NAVAIR
Army Operational Test Readiness Review (OTRR)	Dr. James Streilein	ATEC

26 July 2007

Development and Acquisition of Command and Control Systems	Ms. Frances A. Duntz	Electronic Systems Center
Development and Acquisition of Aeronautical Systems	Mr. Gerry L. Freisthler	Aeronautical Systems Center
Development and Acquisition of Air Armament Systems	Ms. Judy A. Stokley	Deputy Air Force PEO for Weapons, and Executive Director, Air Armament Center
Improving Army Materiel Reliability: A Business Case Approach	Dr. Michael Cushing Mr. Steve Yuhas Dr. David Mortin	AEC AEC AMSAA

22 August 2007

DT Needs to Be More Operationally Relevant and That Means Joint	Mr. John Bolino	Joint Warfighting Center
White Sands Missile Range Perspectives	Mr. Tom Berard	White Sands Missile Range
FCS Data Retrieval and Storage	Mr. Mike Toole	Boeing
Testing in a Joint Environment	Col Eileen Bjorkman	JTEM
Contracting for Reliability	Mr. Shay Assad	OSD (AT&L)
Systems of Systems Software Testing	Mr. Scott Lucero	OSD
Logistics and Material Readiness Perspective on Reliability	Mr. Chuck Silva	OSD
OSD DT&E Organizational Evolution	Ms. Darlene Mosser-Kerner	OSD (AT&L)

23 August 2007

Independent Review Team for AFPEO Weapons Programs	Dr. Steve Butler	USAF
Revitalizing Developmental T&E	Dr. James Finley	DUSD (AT&L)
Overlap between DT and OT – Findings and Recommendations from National Research Council Studies	Mr. Art Fries Mr. Mike Cohen Mr. John Christie	IDA NAS LMI
Nunn McCurdy Triage	Mr. Dave Castellano	OSD

19 September 2007

Requirements Management Certification Training	Mr. Pat Willis	Defense Management Certification Training

20 September 2007

Test Resources Management Center Overview and Priority Test Resource Needs	Dr. John Foulkes	Test Resource Management Center
Major Range and Test Facility Base Metrics Study	Dr. Ed Kraft	AEDC/CZ
Acquisition Program Performance	RDML Steve Eastburg	Naval Air Systems Command
MGS Reliability History	LTC David Rohall	Product Manager Stryker BCT Development

APPENDIX D: TEST AND EVALUATION POLICY REVISIONS MEMO

OFFICE OF THE SECRETARY OF DEFENSE
WASHINGTON, DC 20301-1000

DEC 22 2007

MEMORANDUM FOR: SEE DISTRIBUTION

SUBJECT: Test and Evaluation Policy Revisions

The fundamental purpose of test and evaluation is to provide knowledge to assist in managing the risks involved in developing, producing, operating, and sustaining systems and capabilities.

T&E measures progress in both system and capability development. T&E provides knowledge of system capabilities and limitations to the acquisition community for use in improving the system performance, and the user community for optimizing system use in operations. T&E expertise must be brought to bear at the beginning of the system life cycle to provide earlier learning about the strengths and weaknesses of the system under development. The goal is early identification of technical, operational, and system deficiencies, so that appropriate and timely corrective actions can be developed prior to fielding the system. Consequently, to achieve this goal we have decided to immediately implement the following policies:

- Developmental and operational test activities shall be integrated and seamless throughout the system life cycle. As technology, software, and threats change, follow-on T&E should be used to assess current mission performance and inform operational users' during the development of new capability requirements.

- Evaluations shall include a comparison with current mission capabilities using existing data, so that measurable improvements can be determined. If such evaluation is considered cost prohibitive the Service Component shall propose an alternative evaluation strategy.

- T&E should assess improvements to mission capability and operational support based on user needs and should be reported in terms of operational significance to the user. Consequently, evaluations shall be conducted in the mission context expected at time of fielding, as described in the user's capability document, and consider any new validated threat environments that will alter operational effectiveness.

- To maximize the efficiency of the T&E process and more effectively integrate developmental and operational T&E, evaluations shall take into account all available and relevant data and information from contractor and government sources.

- Operational evaluators will continue to fulfill their statutory roles in providing assessments of operational effectiveness, operational suitability, and survivability to the MDA. In addition, program managers shall report the results of completed developmental testing to the milestone decision authority at milestones B and C. The report shall identify strengths and weaknesses in meeting the warfighters' documented needs based on developmental evaluations. The operational evaluators assessment will be provided to the MDA at the full rate production review.

- To realize the benefits of modeling and simulation, T&E will be conducted in a continuum of live, virtual, and constructive system and operational environments.

These policies will be incorporated in the next revision to DoDI 5000.2.

Dr. Charles E. McQueary
Director, Operational Test & Evaluation

John J. Young, Jr.
Under Secretary of Defense
for Acquisition, Technology & Logistics

APPENDIX E: RAM DEFINITIONS[33]

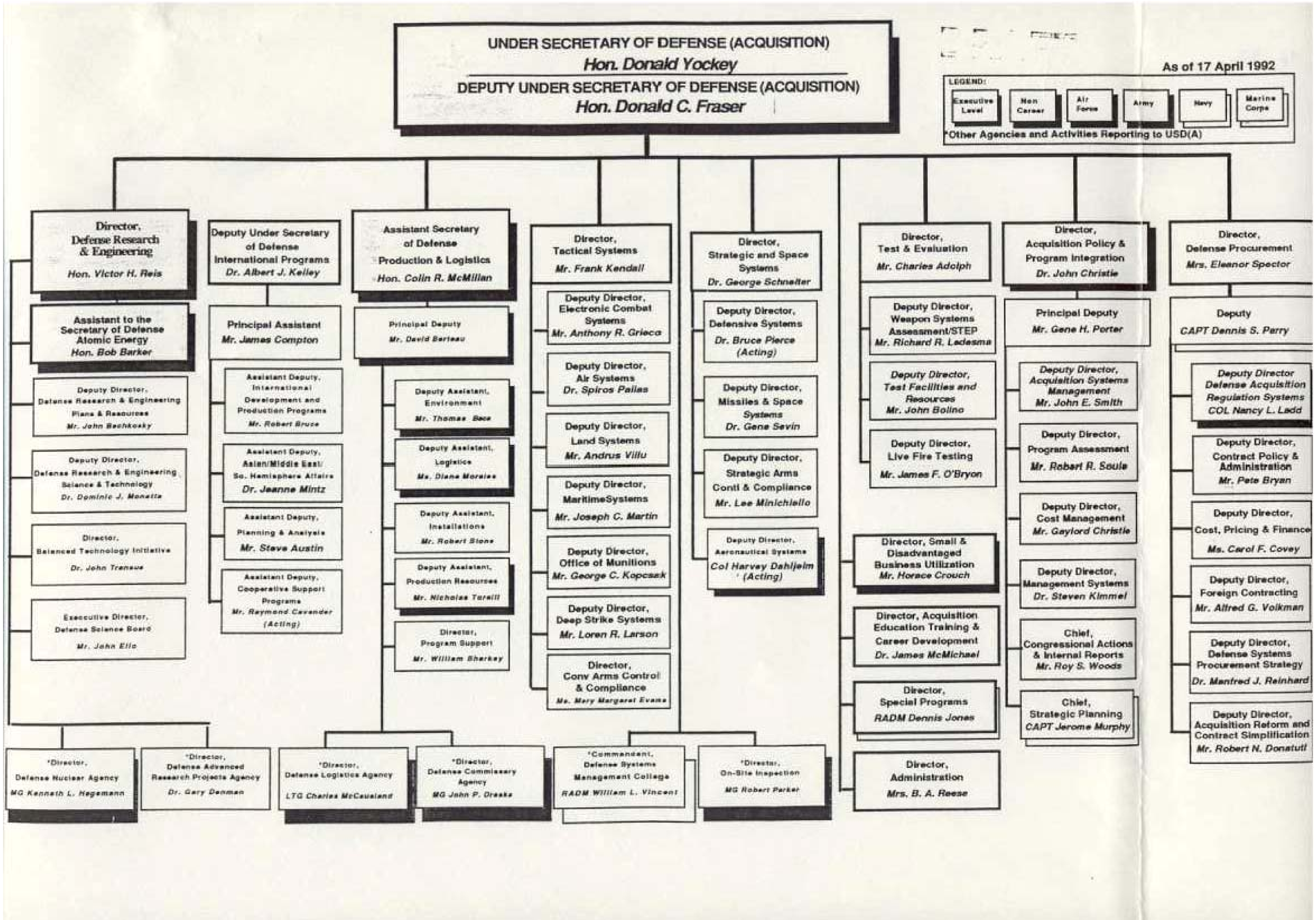
Reliability: The probability of carrying out a mission without a mission-critical failure

Availability: the readiness of the system

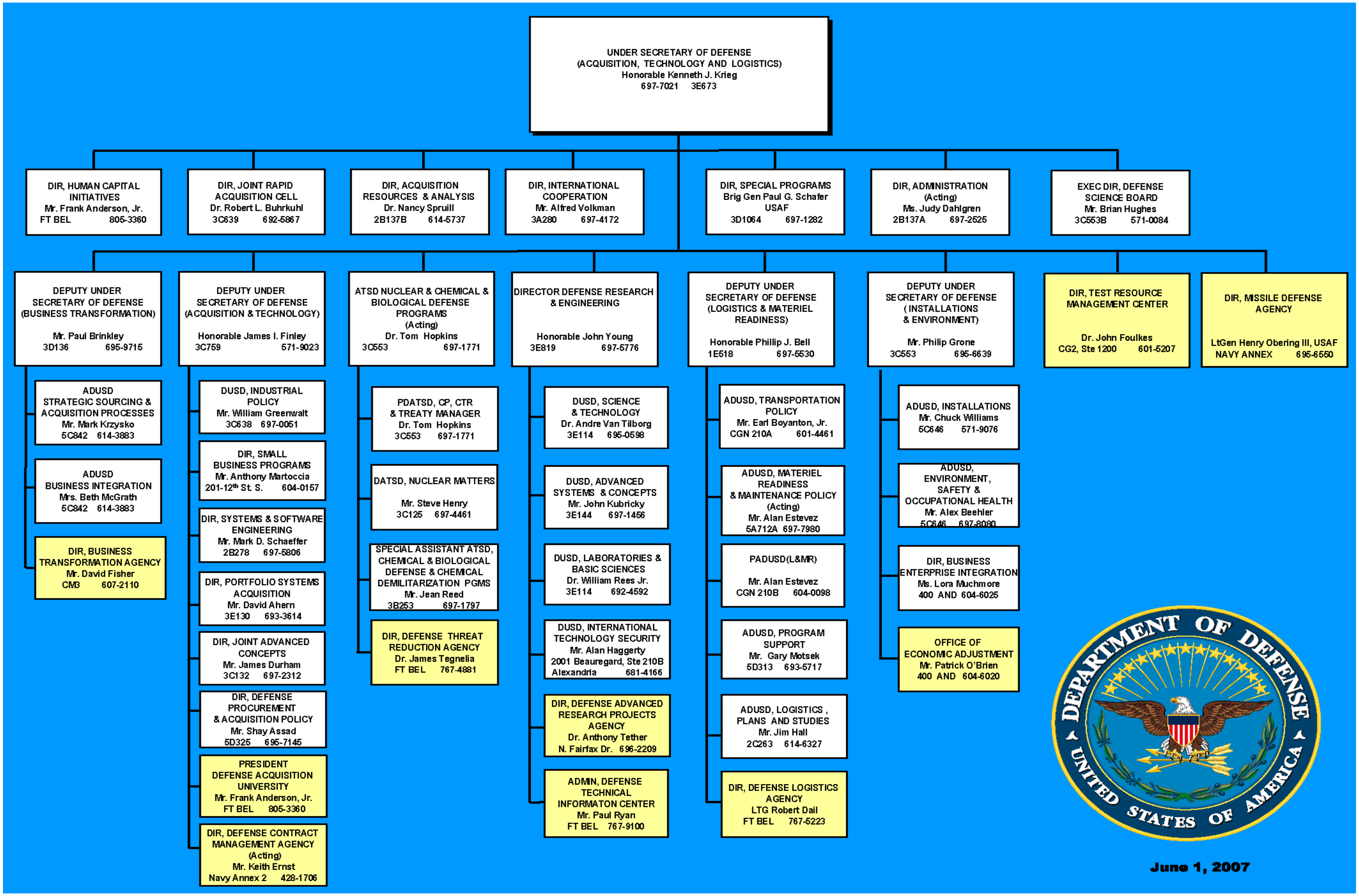
Maintainability: the ease and efficiency with which servicing and preventive and corrective maintenance can be conducted

[33] Defense Acquisition Guidebook (4.4.8)

APPENDIX F: OUSD (A) ORGANIZATIONAL CHART (APRIL 1992)



APPENDIX G: OUSD (AT&L) ORGANIZATIONAL CHART (JUNE 2007)



APPENDIX H: RETIREMENT ELIGIBLE MAJOR RANGE AND TEST FACILITY BASE CIVILIANS[34]

Retirement Eligible MRTFB Civilians

RETIREMENT PROFILE FY 2005 – FY 2010

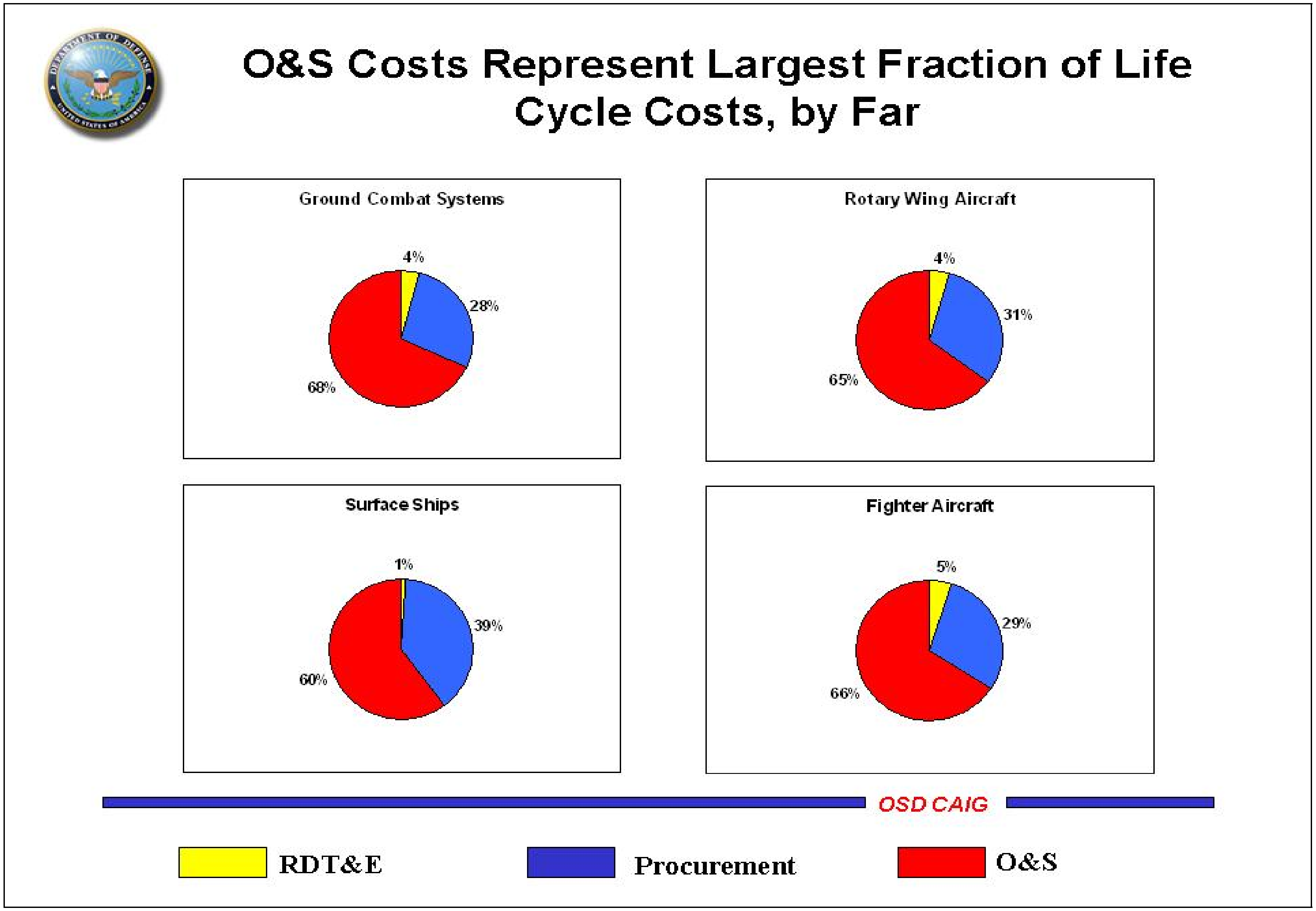
Total Civilians (FY05)		FERS Retirement Eligible					
		FY05		FY 06-10		6-Year Total	
Category	No.	FERS	% in grade	FERS	% in grade	FERS	% in grade
GS-7/GS-12	3897	103	23.9%	441	35.3%	544	32.4%
GS-13/GS-15	4117	61	15.0%	319	28.7%	380	25.0%
GS-6 & Under	703	12	30.0%	57	42.2%	69	39.4%
SES	50	0	0.0%	2	10.0%	2	6.7%
Total	8767	176	19.8%	819	32.6%	995	29.2%

Total Civilians (FY05)		CSRS Retirement Eligible					
		FY05		FY 06-10		6-Year Total	
Category	No.	CSRS	% in grade	CSRS	% in grade	CSRS	% in grade
GS-7/GS-12	3897	328	76.1%	808	64.7%	1136	67.6%
GS-13/GS-15	4117	345	85.0%	793	71.3%	1138	75.0%
GS-6 & Under	703	28	70.0%	78	57.8%	106	60.6%
SES	50	10	100.0%	18	90.0%	28	93.3%
Total	8767	711	80.2%	1697	67.4%	2408	70.8%

24

[34] Institute for Defense Analyses, Demographic Analyses of Elements of the Test and Evaluation Workforce: Path to Workforce Shaping, Prepared for Direct, TRMC and President, DAU, Washington, D.C., January 2008.

APPENDIX I: OSD CAIG LIFE CYCLE COSTS[35]



[35] Walt Cooper, O&S Trends and Current Issues, OSD PA&E/CAIG, Washington, D.C., May 2007.

APPENDIX J: ACRONYM LIST

A	
AAE	Army Acquisition Executive
ACAT	Acquisition Category
AETF	Army Evaluation Task Force
AOTR	Assessment of Operational Test Readiness
ASB	Army Science Board
ASDS	Advanced SEAL Delivery System
AST	ATEC System Team
ATEC	Army Test and Evaluation Command
AT&L	Acquisition Technology and Logistics
B	
BCT	Brigade Combat Team
BRAC	Base Realignment and Closure
C	
C4I	Command, Control, Communication, Computers and Intelligence
CAE	Command Acquisition Executive
CAIG	Cost Analysis Improvement Group
CDR	Critical Design Review
CEC	Cooperative Engagement Capability
CEP	Cooperative Engagement Capability
CMWS	Common Missile Warning System
COI	Critical Operational Issues
COMOPTEVFOR	Commander, Operational Test and Evaluation Force
CONOPS	Concept of Operations
COTF	Commander, Operational Test and Evaluation Force
COTS	Commercial-Off-The-Shelf
CTA	Central Test Authorities
CTF	Combined Test Force
D	
DA	Decision Authority
DAU	Defense Acquisition University
DJC2	Deployable Joint Command and Control
DoD	Department of Defense
DoDI	Department of Defense Instruction
DOTE	Director, Operational Test & Evaluation
DR	Deficiency Report

DSB	Defense Science Board
DT	Developmental Testing
DT&E	Developmental Test and Evaluation
DUSD(A&T)	Deputy Under Secretary of Defense (Acquisition and Technology)

F

FCS	Future Combat Systems
FCT	Foreign Comparative Test
FFRDC	Federally Funded Research and Development Center
FOT&E	Follow-on Operational Test and Evaluations
FY	Fiscal Year

G

GFE	Government Furnished Property
GWOT	Global War on Terrorism

H

HIMARS	High Mobility Attack Rocket System

I

IOT&E	Initial Operational Test & Evaluation
IT	Information Technology
ITT	Integrated Test Team

J

JASSM	Joint Air-to-Surface Standoff Missile
JCIDS	Joint Capabilities Integration Development System
JDAM	Joint Direct Attack Munitions
JMEM	Joint Munitions Effectiveness Manual
JPATS	Joint Primary Aircraft Training System
JRMET	Joint Reliability Maintainability Evaluation Team

L

LFT	Live Fire Test
LFT&E	Live Fire Test and Evaluation
LMI	Logistics Management Institute
LRIP	Low Rate Initial Production

M

MDA	Milestone Decision Authority
MLRS	Multiple Launch Rocket System

MRTFB	Major Range and Test Facility Base
M&S	Modeling and Simulation
MTBF	Mean Time Between Failure

N

NAVAIR	Naval Air Systems Command
NAVSEA	Naval Sea Systems Command

O

OA	Operational Assessment
OEM	Original Equipment Manufacturer
OEF	Operation Enduring Freedom
OIF	Operation Iraqi Freedom
OSD	Office of the Secretary of Defense
OPEVAL	Operational Evaluations
OPTEC	Operational Test and Evaluation Command
OPTEMPO	Operational Tempo
O&M	Operations and Management
O&S	Operations and Sustainment
OT	Operational Test
OTA	Operational Test Agency
OTRR	Operational Test Readiness Review
OT&E	Operational Test and Evaluation

P

PA&E	Program Analysis and Evaluation
PEO	Program Executive Officer
PDR	Preliminary Design Review
PMO	Program Management Office
POM	Program Objective Memorandum

R

RAM	Reliability, Availability and Maintainability
RDT&E	Research, Development, Test & Evaluation
RFI	Rapid Fielding Initiative
RFP	Request for Proposal

S

SAE	Service Acquisition Executive
SDB	Small Diameter Bomb
SDD	System Design and Development

SDD	System Development and Demonstration
SECDEF	Secretary of Defense
SECNAV	Secretary of the Navy
SEWIP	Surface Electronic Warfare Improvement Program
SLEP	Service Life Extension Program
SOMTE	Soldier, Operator, Maintainer, Test and Evaluation
SoS	Systems of Systems
SPAWAR	Space & Naval Warfare Systems Command
SRR	System Readiness Review

T

TECOM	Test and Evaluation Command
TEMP	Test and Evaluation Master Plan
TEWS	Tactical Electronic Warfare System
TOR	Terms of Reference
TPM	Technical Performance Measurement
TRMC	Test Resource Management Center
TSPR	Total System Performance Responsibility
T&E	Test and Evaluation
T&E WIPT	Test and Evaluation Working Integrated Product Team

U

USAF	United States Air Force
USC	United States Code
USD (AT&L)	Under Secretary of Defense for Acquisition, Technology and Logistics

W

WBS	Work Breakdown Structure
WIPT	Working Integrated Process Team