DEFENSE BUSINESS BOARD

For the Secretary of Defense

Best Practices for the Business of Test and Evaluation

DBB FY17-01

Recommendations on the test and evaluation enterprise to improve management and effect process improvements

PREFACE

This study, *Best Practices for the Business of Test and Evaluation*, is a product of the Defense Business Board (DBB). Recommendations by the DBB contained within are offered as advice to the Department of Defense (DoD) and do not represent DoD policy.

The DBB was established by the Secretary of Defense in 2002, as authorized by the Federal Advisory Committee Act of 1972 (5 U.S.C., Appendix, as amended), and governed by the Government in the Sunshine Act of 1976 (5 U.S.C. § 552b, as amended), 41 CFR 102-3.140, and other appropriate federal and DoD regulations. The DBB provides the Secretary and Deputy Secretary of Defense with independent advice and recommendations on how “best business practices” from the private sector’s corporate management perspective might be applied to overall management of DoD. The DBB’s members, appointed by the Secretary, are senior corporate leaders and managers with demonstrated executive-level management and governance expertise. They possess a proven record of sound judgment in leading or governing large, complex organizations and are experienced in creating reliable and actionable solutions to complex management issues guided by proven best business practices. All DBB members volunteer their time to this mission.

Best Business Practices for the Business of Test and Evaluation

TASK

In January of 2016, the Deputy Secretary of Defense directed the DBB to form a task group to evaluate DoD's management and use of the test and evaluation (T&E) enterprise. The Terms of Reference guiding this effort can be found at **Tab A**.

Mr. Bill Phillips chaired a task group that included Mr. Lon Levin and the Honorable Sandy Apgar. Lieutenant Colonel Anthony Cianciolo, Air National Guard, and Captain Thomas Koch, United States Marine Corps, served as the task group's staff representatives.

PROCESS

The task group interviewed officials and experts from within DoD, other government agencies, and the private sector. They reviewed applicable statutes, DoD policies, and DoD directives. The task group also reviewed strategic documents, reports, studies, and briefing papers from both the DoD and industry.

The task group defined the DoD T&E ecosystem as the collection of organizations that have some role and responsibility to provide T&E services to programs for the purpose of our study. This includes but wasn't limited to the Service T&E (developmental and operational) organizations, the Federally Funded Research and Development Centers (FFRDC), DoD level research and development (R&D) organizations, the Service run Major Range and Test Facility Bases (MRTFB), program management offices, contractors, and university participants.

The task group's draft findings and recommendations were presented to the full DBB membership for deliberation and vote at the October 20, 2016 quarterly public meeting wherein the DBB voted to approve all recommendations offered. See **Tab B** for the briefing presented to and approved by the DBB. **Tab C** has the public comments, **Tab D** has the DoD component comments, and **Tab E** has back-up slides and appendices.

BACKGROUND

Test and Evaluation Ecosystem

The task group looked at T&E in the context of the acquisition process. T&E is an important and critical part of a much larger process by which the Department procures major systems. 'Testing' in general, measures a particular subsystem, system, or integrated end item against a desired threshold, and the resulting data is evaluated. Major decisions within the acquisition process are based on data gained and evaluated through this process.

The cost of the T&E enterprise is difficult to quantify. In June of 2015, the Institute for Defense Analysis (IDA) attempted to "capture the cost of DoD testing DoD items, wherever performed and by whomever."[1] IDA aggregated several sources of data to come up with rough estimates for T&E's total cost because the T&E enterprise is so large, highly complex, and deeply embedded in the acquisition process. They estimated the total cost of T&E within the DoD at over $9 billion[2] and utilized a number of assumptions and provided caveats in determining this estimate. IDA also acknowledged that its reported total underrepresented actual T&E costs as many elements were not included because they could not be accurately identified. Those elements included, but were not limited to, contractor T&E costs, MRTFB maintenance, and non-MRTFB costs. These costs also do not include the derivative costs incurred by DoD and the Services when a test must be repeated. The costs are not included because the repeat tests are neither planned nor budgeted for, since the number is unknown.

The total full-time-equivalents involved in T&E across the Department is also difficult to quantify. However, the Department does track roughly 26,000 military, civilian, and contractors who work in the MRTFB infrastructure[3].

The MRTFB is the 'backbone' of the test ranges for the Department. This collection of 23 facilities, spread across the United States, enables the testing of a variety of systems capabilities required by the Department.

1 Madl, Dennis O., "Cost of Testing Analysis: Origin, Description, Data Sources, Assumptions and Limitations, and Results." *Institute for Defense Analysis*, June 2015.
2 Ibid.
3 Test Resource Management Center provided spreadsheet.

Additionally, contractors conduct an unspecified number of developmental tests on DoD systems at their respective corporate facilities. The task group did not attempt to quantify this.

There is other significant T&E-like activity that occurs at numerous FFRDCs, University Affiliated Research Centers, and DoD-owned labs. These facilities and their activities were not addressed by the task group and their costs were not included in IDA's estimate.

FINDINGS

The task group conducted interviews and reviewed materials, resulting in findings that were organized into three broad categories: process; infrastructure; and industry. Process findings dealt with the steps, actions, and decision-making associated with T&E management and execution. Infrastructure findings focused on the MRTFB and its role in facilitating tests. Industry findings addressed successful T&E business practices from various corporations. The task group met with corporations that conducted business with the Department across an 'engagement spectrum' from significant levels of engagement to very little engagement, as well as some that do not conduct business with DoD at all.

1. Process

A. The task group assessed that the overall process design integrating T&E into major acquisition programs seemed reasonable and logical. The process is well understood by process participants and time tested among many acquisition programs. All major programs include a T&E component, and the foundational document that guides the test plan design is the Test and Evaluation Master Plan (TEMP).

B. Programmatic test points, and the plan to validate them, are difficult to define in the early stages of a program. Fidelity is gained as the program and technology matures over time and, as a result, the applicable TEMP is updated and modified throughout the acquisition process. Many different organizations are involved with the development and approval of a TEMP. Large joint programs may require dozens of separate approvals, each of which effectively has

veto authority, adding time and complexity to a TEMP's approval process.

C. Accurate tracking of T&E costs is not generally viewed as a priority by the Department. Trends or benchmarks are not available to analyze because T&E's full costs are not thoroughly captured or tracked.

D. Philosophical differences exist regarding testing to the original requirements of an acquisition program versus testing to an evolved adversary's capability (or tactic, technique, and procedure) at the time the program is ready for fielding and operational testing. These two different sets of requirements often conflict, particularly due to the time it takes to mature a major acquisition program.

E. The authority to approve levels of risk is neither clear nor consistent within the T&E realm. Process participants cited the use of different risk decision authorities and overly vague consensus-based risk decision-making. In some cases, program managers felt they had the ultimate authority. In others, the test community felt they had the same authority. The task group concluded that with so many approval levels associated with a TEMP, no specific entity seemed to have ultimate authority to decide issues surrounding risk.

F. Many in the acquisition process believe the drive for zero risk is too strong and uncompromising. The DBB affirms that risk should always be minimized, but not by imposing an impossible standard of perfection, which often results in consuming excessive resources, lengthening the schedule, and hindering performance. With multiple officials acting as approvers associated with a TEMP, the most conservative perspective on acceptable risk too often becomes the controlling voice.

G. The Services seem to lack authority to make final determinations on the suitability and effectiveness of a program; they often feel 'left out' of risk decisions. The Services' voice in the risk conversation seems very low compared to their responsibility once a system is fielded.

H. Unplanned testing, especially when added toward the end of a program during operational testing, creates pressures on both budget and schedule.

I. The testing community faces challenges as it tries to execute a TEMP when testing is deferred due to budget or schedule pressures.

J. DoD and industry are engaged in processes intended to address emerging challenges associated with the testing infrastructure and the requirements for future testing capabilities. Progress is occurring, but cooperation is encumbered by DoD's overly bureaucratic and complex processes.

K. The Strategic Capabilities Office follows a testing philosophy intended to shorten the process. They focus on tight, rapid 'build, measure, learn' processes that find 'new' uses for existing platforms and systems. Testing to failure, versus testing to predetermined test points, allows them to explore the 'space' beyond the known capabilities of many of these platforms and programs. These practices are consistent with many commercial sector best practices, which account for tighter development cycles demanded by a competitive marketplace.

L. Overall, the task group found the acquisition process has many misaligned incentives that drive process participants to often act in divergent ways. Resource, requirement, and acquisition participants have their own culture and incentives that, when combined in a program management office, often influence testing and the subsequent evaluation of test data in negative or inconsistent ways.

2. **Infrastructure**

A. DoD's testing infrastructure is aging, and, in some cases, obsolete. Its current T&E infrastructure will not be able to support future testing of new technologies, such as hypersonic and autonomous systems without increased funding. Many legacy testing capabilities are maintained within the Department, creating a significant overhead burden. There are many reasons for retaining older and underutilized assets, ranging from the need to support older legacy programs still in use to legal restrictions imposed from outside the Department.

Many T&E facilities are seeking other government and private users to offset operations and maintenance costs. Success in this area varies, but the effort is critical given the dual realities of aging infrastructure and reduced funding.

B. Processes that DoD program managers use for scheduling tests on ranges are inconsistent among major test facilities. Administrative requirements differed from one test range to another, as well as within a single range. Generally, success often depends on having personal knowledge of specific employees or contractors on the base or range. A common answer given by interviewees is that the system works ‘as long as you know the right person.’ The need for common and consistent processes for scheduling and utilizing ranges was also a common refrain.

C. The Department is building a master database of government-owned testing assets to help both industry and the Department understand what capabilities can be tested and where. The database will help with improving management and utilization of these assets. Unfortunately, this database’s usefulness will be limited because the Department cannot include contractor assets due to government imposed legal restrictions, in addition to most contractor’s reluctance to participate for proprietary and intellectual property reasons.

D. Management metrics for infrastructure are limited and are applied inconsistently by location.

E. Testing costs can be reduced through relying less on physical testing at ranges and increasing reliance on modeling and simulation. Private sector organizations, with an understanding of DoD needs, have echoed this point and have called for a more aggressive approach to the application of modeling and simulation in DoD. There is, however, debate within the Department on the real impact of using modeling and simulation to reduce both cost and testing time. Consequently, dedicated funding necessary to move toward greater modeling and simulation has not been made a priority.

3. **Industry**

A. There is no consistent model for developmental and operational testing in private industry, but the independence of testing is valued and constructive friction between stakeholders is accepted and encouraged. In some companies, developmental and operational testing organizations are separate, as in DoD. In others, they have been combined. In some companies there are no dedicated development and operation organizations at all. In those instances, the testing engineers are embedded in the product development teams from the beginning; however, they are still expected to exercise independent judgement and have a voice with leadership as issues arise.

B. Private sector product development cycles are shrinking and there is pressure to be ‘first to market’ to achieve a competitive advantage. Thus, companies face greater urgency to streamline the entire development cycle, including the T&E function. Corporations are also thinking incrementally (consider the iPhone upgrade cycles) in order to have a steady stream of new products and capabilities in the market. This contrasts with DoD, where large programs take far longer to ‘get to market.’

C. Industry in general is much further along in the application of modeling and simulation for T&E. The same market pressures that are forcing companies to reduce development times and cost are also pushing them to rely on a greater application of modeling and simulation. While they continue to supplement T&E with physical tests to validate the models, they recognize that, in the long run, modeling and simulation provides significantly lower cost, greater accuracy and flexibility, and deeper sophistication for testing in general.

D. Corporate T&E costs are known and examined routinely. Different companies take different approaches to this, ranging from periodic zero-based analysis of T&E to the application of cost reduction targets. A corporation's approach to T&E cost management typically mirrors its approach and philosophy to overhead and cost management in general.

E. The following characteristics were common among the successful corporate executives interviewed even though there are very different approaches to the application of T&E:

i. Successful corporations rely on small, empowered, trusted teams to execute product development and the operational testing that accompanies it.
ii. These core teams include business and engineering leaders who, together, make risk tradeoff decisions. Generally, the business leader has the 'final vote.'
iii. Many corporations use game theory and probability analytics to inform T&E decisions as well as internal risk calculations.
iv. Cost discipline is incorporated into the T&E process.
v. Iterative development is encouraged to enable speedy delivery to market. Competitive market gaps may be accepted with the intent to address them in subsequent iterations of the product. Corporations have a requirements 'lock,' which solidifies the basis for planning, budgeting and testing.
vi. Corporations establish agreements between stakeholders in a program, including the testing community, about what successful achievement of the requirements looks like and how success will be measured.

RECOMMENDATIONS

The DBB offers the following recommendations:

1. **Limit accountability for T&E to the few directly aligned with program management.** Other participants with a voice in the process should have a stake in meeting the budget, schedule, and performance of a program. Greater collaboration among testers and program managers throughout the program lifecycle is therefore required. Program managers and the T&E community should define success at the beginning of a program, and discuss how changes will be handled in the future. The number of stakeholders that have veto power over the TEMP, either stated or implied, should be strictly limited. It should be made clear who has veto power versus who fills an advisory role. The Department should consider successful commercial management models that better align the incentives of program managers, defense contractors, and the test community.

2. **Strengthen Service voice in critical T&E decisions.** The Milestone Decision Authority should consider delegating authorities for making certain risk acceptance decisions regarding T&E to the Service Chiefs or their designated subordinates. This would be consistent with the Services' Title 10 responsibilities to organize, man, train, and equip the force. A Service voice should provide a reasonable risk perspective, which will help counter the drive toward the impossible "zero risk" standard. This problem is particularly acute in the large ACAT I programs.

3. **Process participants should adopt a principled compromise between "zero risk" and the operating environment.** Tolerable risk should be stated, quantified, and disseminated on a program by program basis. Importantly, the analysis of test data should be standardized among process participants to lead to commonly understood and accepted outcomes. Generally, the Services are best positioned to determine "good enough" for fielding capabilities, and are equivalent to industry's business leads who have the final risk decision for programs.

4. **Use cost discipline as an essential management tool.** The Department should identify an organization that has direct responsibility for documenting all current T&E costs, including direct and indirect military, civilian, and contractor costs.

   A. The Department should consider designating the Assistant Secretary of Defense for Research and Engineering, supported by the Comptroller, as the lead agent for this effort.

   B. Management metrics should be developed based on the aggregate costs to identify benchmarks and improvement areas. These management techniques will create a better understanding of where the money is spent today and how to establish tradeoffs between modeling/simulation and physical testing in the future.

   C. It is strongly recommended that DoD not create a new organization, but instead urge each stakeholder to own their particular piece of the cost pie.

5. **Improve the effectiveness of the T&E infrastructure.** Common processes for range utilization can create efficiencies in cost and ease of access for customers. Common utilization metrics for ranges should be developed and applied in conjunction with the current efforts to build an infrastructure asset data base.

   A. Ranges should increase their ability to execute modeling and simulation to reduce physical testing costs.

   B. Existing industry tools and best practices should be adopted to recapitalize infrastructure and reduce the financial burden to the government.

   C. Department leaders should look to integrate testing philosophies, such as those used in the Strategic Capabilities Office, into modern test practices at facilities and ensure the facilities are postured to meet these new demands and approaches.

CONCLUSION

Today, the process of test and evaluation in the overall acquisition process operates as designed in the latest version of DoD Directive 5000.01 "*The Defense Acquisition System*" and the Director, Operational Test & Evaluation TEMP Guidebook. The process is well understood by the Services, the T&E organizations, and contractors. Unfortunately, the T&E process is so embedded in the broader acquisition process that any transformative change to T&E would require a similar transformative change to the acquisition process, which is beyond the scope of authorities provided to the Director, Operational Test & Evaluation.

When contrasting the DoD T&E process with analogous processes in successful businesses, however, it becomes clear that operating as designed is not good enough for the future of T&E in DoD. The ubiquitous nature of change, the advent of new technologies, and the rapid evolution of adversarial threats are putting greater pressure on DoD's T&E function. Reflecting on industry best practices, it is believed that the Services need a stronger voice in the T&E process, similar to the role of the CEO's voice in the private sector.

The above recommendations are intended to help move the Department in the direction of more agile, cost-effective testing, more effective risk assessment and acceptance decisions, and ultimately a T&E infrastructure that is better positioned to support future requirements.

On behalf of the Chairman and the Defense Business Board this study is respectfully submitted.

Bill Phillips
Task Group Chair

Defense Business Board

TAB A

TERMS OF REFERENCE

DEPUTY SECRETARY OF DEFENSE
1010 DEFENSE PENTAGON
WASHINGTON, DC 20301-1010

JAN 21 2016

MEMORANDUM FOR CHAIRMAN, DEFENSE BUSINESS BOARD

SUBJECT: Terms of Reference - "Best Practices for the Business of Test and Evaluation"

Efficient and effective Test and Evaluation (T&E) is critical to getting the right capabilities to the field in a timely and efficient manner. The Department of Defense (DoD) has many T&E organizations, and many more program offices and contractors that are responsible for developing and utilizing test resources. At a time when savings are an imperative to meeting tomorrow's security challenges, the Department must ensure it is applying the best business practices to its management of T&E resources and infrastructure.

Therefore, I am establishing a Task Group under the Defense Business Board (DBB) to develop actionable recommendations that the Department should consider in order to make the best use of DoD's T&E resources and infrastructure.

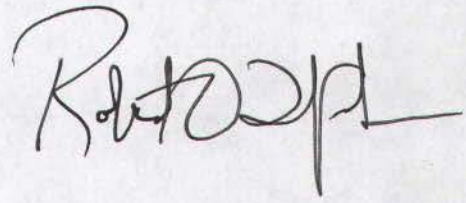
In particular, the DBB, through the Task Group, shall review the Department's T&E enterprise to assess its management and utilization of both organic and contract resources and infrastructure. This effort should be aimed at identifying improvements to ensure excellence in managing the enterprise, reduce duplication, and achieve savings. The review should include an assessment of the utilization of the Department's T&E resources and infrastructure to include an inventory of DoD facilities and the contractor-procured facilities available to the Department, and any other such matters as the DBB determines relevant.

Task Group recommendations will be presented to the DBB for thorough, open deliberation in a noticed, public meeting. The DBB will provide its findings and recommendations to the Secretary of Defense or the Deputy Secretary of Defense no later than October 20, 2016. The Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics will serve as the DoD liaison for this project with assistance from the Office of the Deputy Chief Management Officer.

In conducting its work, the DBB has my full support in all requests for data or information that may be relevant to this study. As such, DoD offices are directed to promptly facilitate the study by ensuring that the DBB staff has access to any and all personnel and information necessary to perform the study.

As a subcommittee of the DBB, this Task Group shall not work independently of the DBB's charter and shall report its recommendations to the full DBB for public deliberation and approval, pursuant to the Federal Advisory Committee Act of 1972, as amended, the

Government in the Sunshine Act of 1976, as amended, and other applicable federal statutes and regulations. The Task Group does not have the authority to make decisions on behalf of the DBB, nor can it report directly to any Federal representative. The members of the Task Group are subject to applicable ethics rules.



Defense Business Board

TAB B

PUBLIC MEETING BRIEFING

PROVIDED TO THE DEFENSE BUSINESS BOARD

DEFENSE BUSINESS BOARD

Best Practices for the Business of Test and Evaluation

October 20, 2016

Final – Approved by the DBB on 20 October 2016

Task Group

Members

- Mr. Bill Phillips (Lead)
- HON Sandy Apgar
- Mr. Lon Levin

Staff

- Capt Tom Koch, USMC
- Lt Col Tony Cianciolo, ANG

Task

"…Review the Department's T&E enterprise to assess its management and utilization of both organic and contract resources and infrastructure…to ensure excellence in managing the enterprise, reduce duplication, and achieve savings. The review should include…any other such matters as the DBB deems relevant."

- T&E TOR, signed 21 Jan 16

Bottom Line Up Front

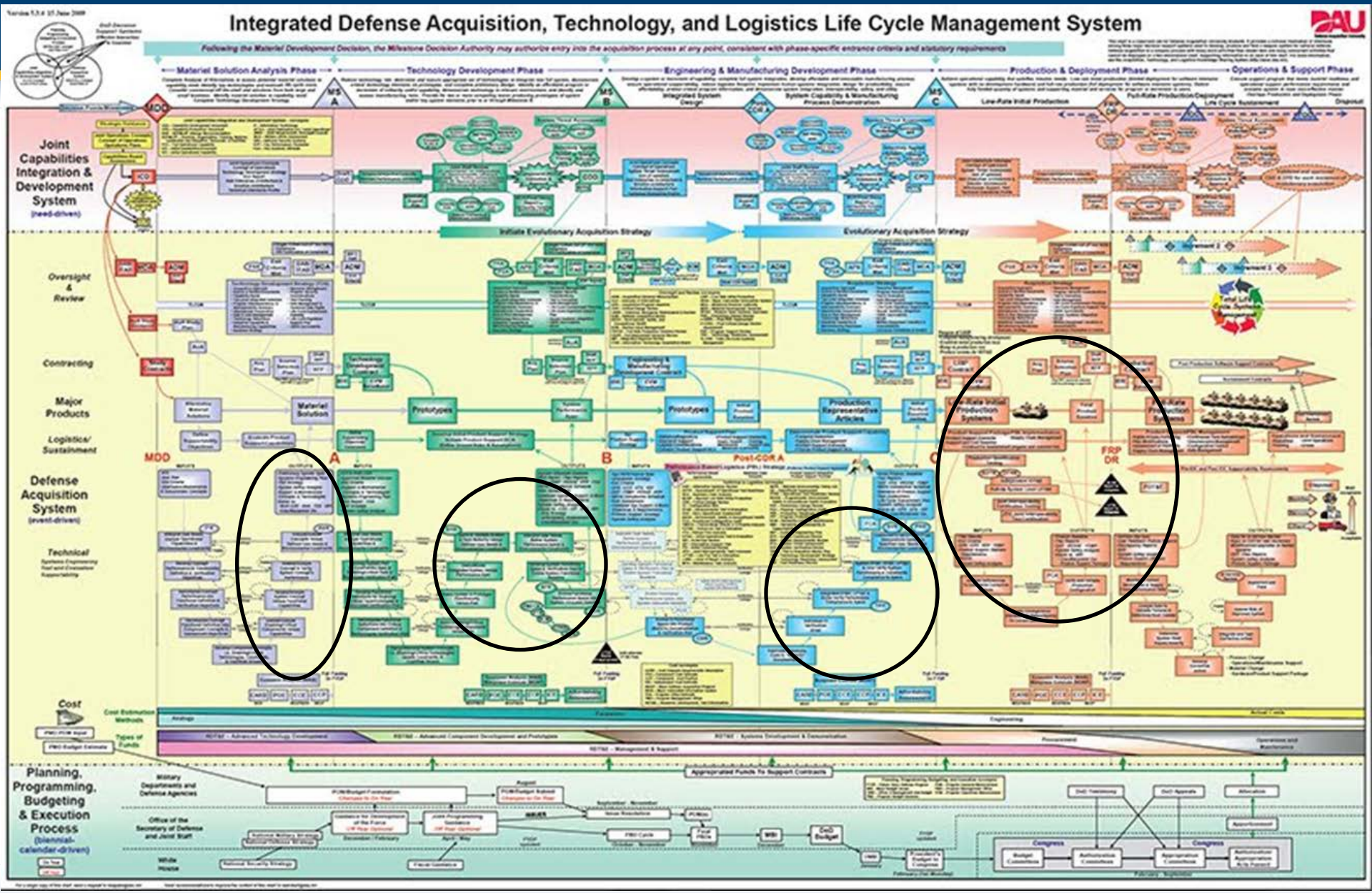
- "Test & Evaluation" (T&E) is a function of, and critical to, the acquisition process
  - True costs (dollars and people) are unknown
  - Infrastructure is not fully postured for the future
  - Different business and risk decisions made from the same test data pressure cost and schedule
  - Too many approvers, Services' feel unable to influence risk calculus
  - Diffuse risk accountability results in very low tolerance for risk
- Recommendations
  - Align T&E accountabilities with Program Management
  - Strengthen Service voice in risk decisions
  - Create cost discipline and apply it as a management tool
  - Improve effectiveness of T&E infrastructure

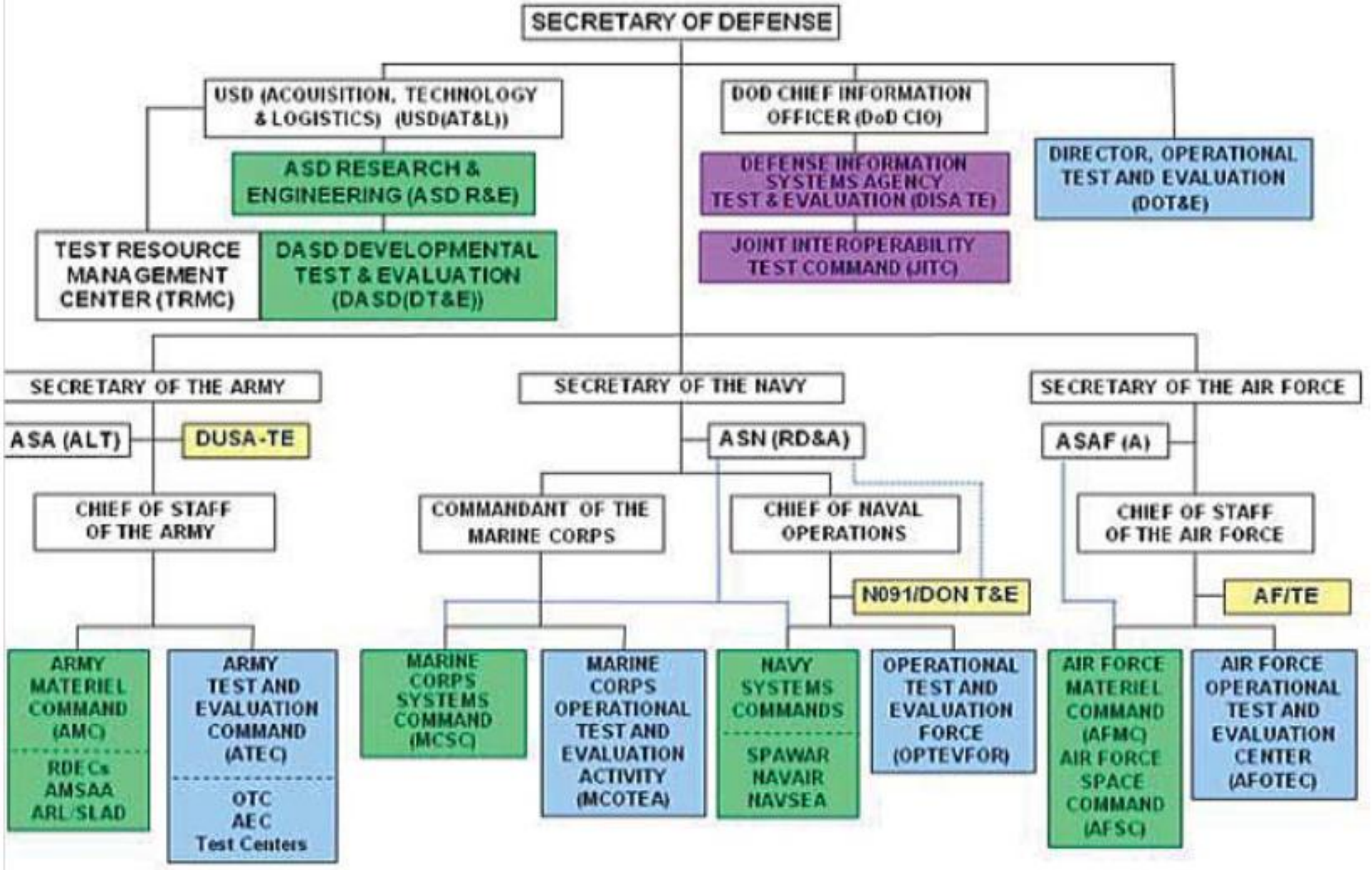
Test and Evaluation Ecosystem

- Purpose
  - Part of the overall DoD acquisition process
  - Assess fitness against requirements and performance
  - Key acquisition process decisions are informed by test results
- Function
  - Validate design parameters throughout:
    - Technology Maturation and Risk Reduction Phase (TMRR)(Between Milestone A&B)
    - Engineering and Manufacturing Development Phase (EMD)(Between Milestone B&C)
    - Production and Development Phase (P&D)(Post Milestone C)

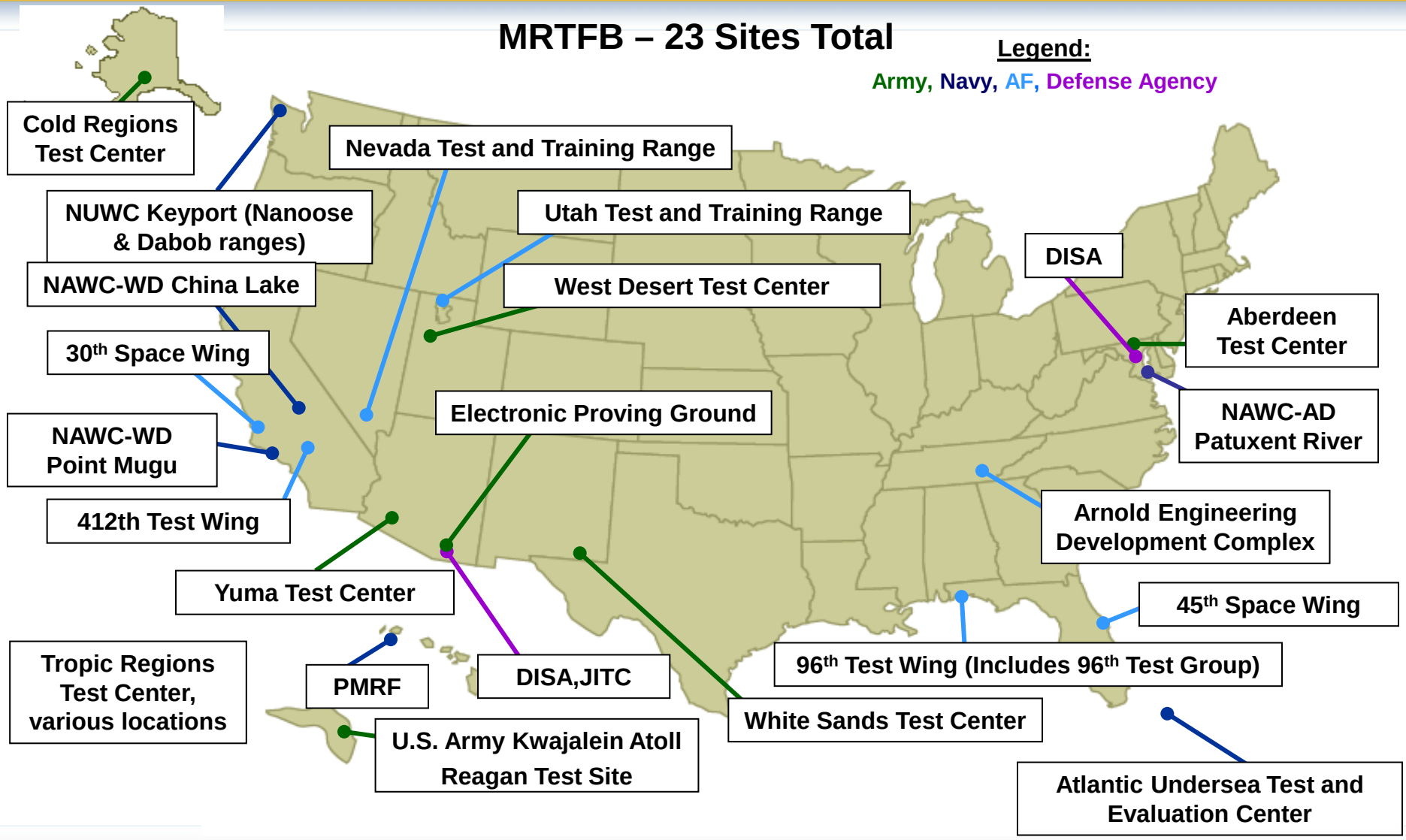
Test and Evaluation Ecosystem

- Cost
  - June 2015 IDA study (based on many caveats and assumptions)
    - FY14 - ~$8.23B
    - FY16 est. - $9.05B
  - Indirect costs unknown
- People
  - Many Organizations and Contractors – total FTEs unknown
  - ~26k Military / Civilian / Contractor in infrastructure
- Places
  - 23 MRTFB ranges all over country
  - Countless contractor facilities (primarily for DT that is conducted by a contractor as a part of an acquisition)


Integrated Defense Acquisition, Technology, and Logistics Life Cycle Management System
Following the Materiel Development Decision, the Milestone Decision Authority may authorize entry into the acquisition process at any point, consistent with phase-specific entrance criteria and statutory requirements
Materiel Solution Analysis Phase
Technology Development Phase
Engineering & Manufacturing Development Phase
Production & Deployment Phase
Operations & Support Phase
MS A
MS B
MS C
FRP DR
Joint Capabilities Integration & Development System (need-driven)
Initial Evolutionary Acquisition Strategy
Evolutionary Acquisition Strategy
Oversight & Review
Contracting
Major Products
Logistics/ Sustainment
Defense Acquisition System (event-driven)
Technical
Prototypes
Material Solution
Production Representative Articles
Low-Rate Initial Production
Full-Rate Production/Deployment
Post-CDR A
Cost
Planning, Programming, Budgeting & Execution Process (biennial-calendar-driven)
Appropriated Funds To Support Contracts
Military Departments and Defense Agencies
Office of the Secretary of Defense and Joint Staff
White House
Congress


SECRETARY OF DEFENSE
USD (ACQUISITION, TECHNOLOGY & LOGISTICS) (USD(AT&L))
ASD RESEARCH & ENGINEERING (ASD R&E)
TEST RESOURCE MANAGEMENT CENTER (TRMC)
DASD DEVELOPMENTAL TEST & EVALUATION (DASD(DT&E))
DOD CHIEF INFORMATION OFFICER (DoD CIO)
DEFENSE INFORMATION SYSTEMS AGENCY TEST & EVALUATION (DISA TE)
JOINT INTEROPERABILITY TEST COMMAND (JITC)
DIRECTOR, OPERATIONAL TEST AND EVALUATION (DOT&E)
SECRETARY OF THE ARMY
ASA (ALT)
DUSA-TE
CHIEF OF STAFF OF THE ARMY
SECRETARY OF THE NAVY
ASN (RD&A)
COMMANDANT OF THE MARINE CORPS
CHIEF OF NAVAL OPERATIONS
N091/DON T&E
SECRETARY OF THE AIR FORCE
ASAF (A)
CHIEF OF STAFF OF THE AIR FORCE
AF/TE
ARMY MATERIEL COMMAND (AMC)
RDECs AMSAA ARL/SLAD
ARMY TEST AND EVALUATION COMMAND (ATEC)
OTC AEC Test Centers
MARINE CORPS SYSTEMS COMMAND (MCSC)
MARINE CORPS OPERATIONAL TEST AND EVALUATION ACTIVITY (MCOTEA)
NAVY SYSTEMS COMMANDS
SPAWAR NAVAIR NAVSEA
OPERATIONAL TEST AND EVALUATION FORCE (OPTEVFOR)
AIR FORCE MATERIEL COMMAND (AFMC) AIR FORCE SPACE COMMAND (AFSC)
AIR FORCE OPERATIONAL TEST AND EVALUATION CENTER (AFOTEC)

Infrastructure



Findings: Process

- T&E is a small but essential function within the large and complex DoD acquisition process
- The overall process design integrating T&E into major programs seems reasonable and logical
  - Methodology is well understood and time tested
  - All programs include a T&E component
  - The Test and Evaluation Master Plan (TEMP) is the key T&E document for a program
- Test points and the plan to validate them (via the TEMP) are difficult to fully define early in a program
  - Fidelity is gained as the program and technology matures
  - Large programs may have up to 30 separate entities that must approve the “plan” along the way

Findings: Process (continued)

- Overall T&E costs are unknown and not viewed as a priority
  - Lack of incentive to reduce cost because cannot track them directly
- Philosophical differences between testing to original requirements vs. new adversary capabilities
- Accountability for risk trade-off decisions is not clear or consistent
- Perception that drive for zero risk is too strong and uncompromising
- Services can feel left out of final decisions and lack authorities to make final “calls” on suitability and effectiveness
- Lack of resources when additional testing is required creates pressure on budget and schedule
- DoD and industry are engaged in addressing the emerging challenges and generally work well together; but many view DoD process as bureaucratic and overly complex

Findings: Process (continued)

- Strategic Capabilities Office (SCO) offers alternatives to the current test approach
  - New take on the “development / test / deployment” method
- Misaligned incentives drive process participants to act in certain ways
  - Budget (PPBE), Requirements (JCIDS), Program Management, T&E

Findings: Infrastructure

- Infrastructure is aging and unlikely to be feasible or cost effective testing new or future technologies (e.g. cyber, hypersonics)
  - Legacy programs cited as reason for holding on to old and expensive assets – inadequate analysis of actual or opportunity costs
  - Causes overhead burden and need to find other private and public users to offset operations and maintenance costs
- Processes used for scheduling and testing are inconsistent among major test facilities
  - "Works fine as long as you know the right person"
- DoD building master database of government-owned testing assets
- Management metrics for T&E infrastructure are limited and applied inconsistently by services and locations
- Recognize need to move toward reducing physical testing through software and simulation, but leadership disagrees on potential impact and progress is slow

Findings: Industry

- No consistent model for DT and OT in industry
  - But, independence of T&E is valued, friction accepted/encouraged
- Product development cycles are shrinking
  - Urgency to be “first to market” for competitive advantage
- Industry in general is further along in the application of software and simulation for T&E
  - Important to reducing cost and time to market
  - Models are validated and supplemented by historical live testing data and select physical tests
- T&E costs are known and examined routinely

Findings: Industry (continued)

- Successful companies have common characteristics:
  - Rely on small, empowered and trusted teams to execute product development and appropriate DT and OT
  - Teams include leadership from business and engineering organizations who together make risk tradeoff decisions, with business leader generally holding the “final vote”
  - Apply extensive use of game theory and probability analytics to inform T&E decisions including risk calculations
  - Accept and incorporate cost disciplines throughout T&E process
  - To enable speed to market, they may accept competitive market gaps, relying on next generation to address them
    - Requirements “lock” and subsequent iterative development, testing, and fielding
  - Agreement going into a program among stakeholders about what successful achievement of requirements looks like and how success is measured

Recommendations: Process

- Create appropriate and direct accountabilities for T&E community, aligned with Program Management
  - Should have a stake in meeting budget, schedule and performance elements
  - Will require a greater involvement and commitment of T&E community at the beginning and throughout program life-cycle
  - Definitions of “success” should be agreed between PM and T&E community from the beginning and with changes that occur over time
  - Sharply limit the number and weight of stakeholders that have veto power (stated or implied) over TEMP development, implementation and modification
    - Clarify those with approval and veto power vs those who have advisory role

Recommendations: Process (continued)

- Strengthen Service voice in critical program decisions
  - Authorities should come from the Service Secretary to the Service Chief and designated subordinates
  - Supported by stronger more reasonable risk assessment determination as programs progress
    - Today's drive toward "zero risk" is an impossible standard and should not be used by the T&E community
    - Adopt "principled compromise" between absolute zero-defect test doctrine and situational operating requirements
    - "Tolerable" risk should be stated and quantified
    - Apply more consistent analysis of common data in the T&E process
  - Services are best positioned to determine "good enough" when program length results in new threats that didn't exist when requirements set
  - In effect the equivalent of the private sector "business" leaders in terms of influence on final fit for Service

Recommendations: Process (continued)

- Create Cost Discipline and Utilize as Critical Management Tool
  - Identify organization to take responsibility for documenting current costs – direct and indirect, military, civilian and contractor
    - Consider ASD Research & Engineering (ASD R&E), supported by OSD (Comptroller)
  - Establish management metrics based on these costs to identify improvement areas, including where resources are inadequate as well as where resources are redundant or inefficiently used
  - Creates better understanding of where money is spent today, facilitates better decisions regarding obsolete assets and facilities and provides a path forward for greater use of simulation and software tools
  - Do not create new organization, but instead rely on existing organizational elements to “own” their piece of the cost

Recommendations: Infrastructure

- Improve the effectiveness of T&E infrastructure through:
  - Common processes for enterprise-wide facility utilization
  - Develop and apply common utilization metrics
  - Continue efforts to build T&E asset data base
  - Aggressively invest in and utilize software based simulation to reduce percentage of T&E relying on physical testing
  - Adopt existing privatization tools to help recapitalize T&E and reduce the financial burden and risk of future needs of the government
  - Seek ways to integrate the methodological approach to agile development and testing being led today by the Strategic Capabilities Office

Defense Business Board

TAB C

PUBLIC COMMENTS

As a Federal advisory committee, the DBB is statutorily bound to make publically available comments received in response to its studies. The DBB additionally offers those DoD entities wherein a study focuses, the opportunity to respond to the study's recommendations. During the course of a study, DBB task group members seek DoD feedback to the findings in order to ensure the data collected is as accurate as possible.

PUBLIC COMMENTS

As of the date of this study being published no public comments were received by the Defense Business Board for inclusion.

Defense Business Board

TAB D

DoD COMPONENT RESPONSES

SUBMITTED TO THE DEFENSE BUSINESS BOARD

DEPARTMENT OF DEFENSE COMPONENT RESPONSES

One Department of Defense component response was received for inclusion as of the completion of this study. On January 10, 2017, then-Director, Office of Test and Evaluation, Dr. J. Michael Gilmore, provided a letter in response to the public briefing of the Best Practices for the Business of Test and Evaluation. Dr. Gilmore's letter is provided in its original form following the task group chair's response.

TASK GROUP CHAIR RESPONSE

Dr. Gilmore's reply reads somewhat like a reply to a GAO report and misses several key points completely, although he also makes a number of good points. It should be pointed out that the response was to the briefing slides presented at the DBB's public meeting; the full study had not been completed and formally presented to the Secretary of Defense. The completed study contains reference material not contained in that public briefing. I offer the following observations about Dr. Gilmore's feedback.

First, Dr. Gilmore's seemed to focus on defending his specific organization, the Office of Test and Evaluation (OT&E), against DBB findings and recommendations. Yet the fact is that the DBB looked at the far larger and much broader DoD T&E ecosystem, of which OT&E constitutes only a small, though very important, part. Dr. Gilmore apparently forgot that in that single meeting we had with him, he started the conversation by asking why the DBB task group wanted to even talk to him because the study's ToR (which he said HE wrote) was not about OT&E, but instead about the rest of the T&E process across DoD. Nonetheless, the task group solicited Dr. Gilmore's perspective before tackling the rest of T&E landscape. He agreed to talk to us initially, however, requests for follow-on discussions were rebuffed.

In fact, even though the specter of OT&E loomed over many task group discussions, the task group did not focus on that organization at all. The task group recognized the unique and independent nature of what OT&E is charged to do and agreed with the need for that independent voice. The focus of this study was on the larger $9B+ T&E enterprise -

Services, test ranges, developmental testing, contractors, etc. It is certainly understood that the task group looked at things through the lens of where it sits, but as a result, Dr. Gilmore seemed to miss or misunderstand several DBB observations and recommendations.

For example, in his letter Dr. Gilmore immediately disagreed with the DBB comment about the Department not knowing the cost of T&E. He quickly jumps to the fact that this is not true because he "knows" the cost of DOTE. The DBB agrees that he does and the GAO confirmed it in their report of 2015. However, the fact is, OT&E is not the entirety of DoD's T&E ecosystem and time and time again the DBB task group found the Department does not actually know what T&E costs. This isn't unique to T&E, because the Department doesn't really know what most anything truly costs (hence the inability to yet achieve a clean audit). The DBB task group met with several elements of the broader T&E ecosystem, did the same with a sampling of successful private sector companies, and based on that made four recommendations for the Department (not OT&E) to consider.

Second, Dr. Gilmore disagrees with the T&E community headcount of 26,000. In his letter, he marks it at 2,000; the number of personnel in the Service operational test agencies. Again, Dr. Gilmore seemed to lose sight of the study's broader view of the DoD-wide T&E community.

Third, in a number of areas Dr. Gilmore seemed to disagree with a DBB observation or recommendation, yet then he supported those findings with his comments. For example,

- The idea that things would work better with operational testing involved earlier in the process. The task group agrees and made that point about it as a best practice.

- Believing that the Service acquisition teams are not equipped to make the right risk trade-off decisions. The DBB point was that the operational side of the Services (Chief of Staff, etc.) are better equipped, not the acquisition side.

It is certainly a fair observation that in a couple instances Dr. Gilmore's comments reflect areas that the DBB study could have stated something clearer than it did, or enumerated a point further or better. The DBB is confident the points and recommendations we made remain valid. They are here presented for the Department to do with as they choose.

OFFICE OF THE SECRETARY OF DEFENSE
1700 DEFENSE PENTAGON
WASHINGTON, DC 20301-1700

OPERATIONAL TEST
AND EVALUATION

JAN 10 2017

MEMORANDUM FOR CHAIRMAN, DEFENSE BUSINESS BOARD

SUBJECT: Briefing on "Best Practices for the Business of Test and Evaluation" is Fatally Flawed and Unsuitable for Decision-Making

The purpose of this memorandum is to provide my feedback on the Defense Business Board (DBB) briefing slides dated October 20, 2016 and titled, "Best Practices for the Business of Test and Evaluation." While Operational Test and Evaluation (OT&E) and Live Fire Test and Evaluation (LFT&E) are only parts of the Department's Test and Evaluation activities, I felt it necessary to provide my feedback as Director, Operational Test and Evaluation to the report's content as it pertains to activities under my purview. The briefing slides make numerous claims of fact absent any supporting data. Perhaps this is because these "facts" are, in fact, false. For example, the first finding on slide 3 states "true costs are unknown." This is not true. As we have documented, we do know the cost of operational testing, and it is a relatively small percentage (typically less than 1 percent) of a program's acquisition costs. A more important question, left unasked by the study, is "what is the cost of losing in combat when our weapons don't work as required or when required because testing was inadequate?" As discussed in detail in the remainder of this memorandum, there are so many instances of false findings that the briefing's recommendations, such as we can understand them, are not useful at all for decision-making or for action by either the Department's leadership or the Congress. This unfortunate situation could have been avoided if the study group had accepted our invitation to engage in more than the single interaction with DOT&E that occurred.

Background

In 1983, Congress created the position of Director, Operational Test and Evaluation (D,OT&E) within the Office of the Secretary of Defense, and the Director was given specific authorities in Title 10 U.S. Code. The Congressional concerns that led to the establishment of this office were many, but included: poor performance of weapon systems, inaccurate reports from the Services, shortcuts in testing because of budget pressure, and a lack of realistic combat conditions and threats in testing. The unique independence of this office, free from conflicts of interest or pressure from Service senior leadership allows us to:

- Highlight problems to DOD and Congressional Leadership to inform their decisions before production or deployment
- Tell the unvarnished truth
- Ensure operational tests are adequately designed and executed

As Director, OT&E, I do not make acquisition decisions but inform those who make them about weapon system performance under combat conditions. My staff of 60 action officers includes 17 active duty military officers from all Services in addition to civilians with advanced

engineering and science degrees. Our mission is to inform acquisition officials before the systems are purchased en masse, and combatant leadership before the systems are deployed operationally, about how weapons will work in combat, including live fire survivability and lethality, before the systems are deployed.

Furthermore, the workforce of the Services' operational test communities is relatively small compared to the approximately 26,000 personnel mentioned on slide 5. In fact the Army, Air Force, Marine Corps, and Navy operational test agencies in total employ less than 2,000 people, and, as shown in Figure 1, that total has declined more than 11 percent in the past 10 years, primarily in the number of military personnel. This is in contrast with trends reported for the broader acquisition workforce, which has grown by 24 percent since FY08.[1]

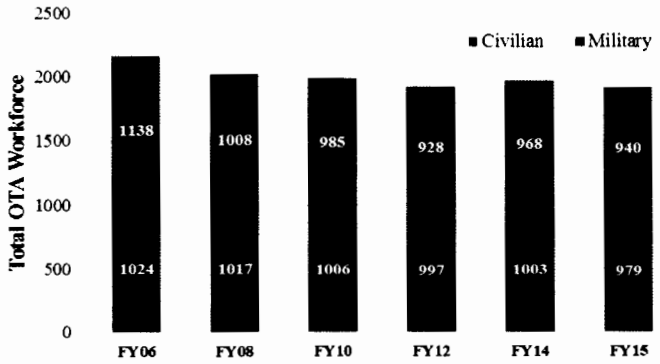


Figure 1. Proportion of Civilian vs. Military Personnel in the OTA Workforce Growth FY06-FY15

Below, I discuss the three principal areas covered by the report: the process of defense test and evaluation, its supporting infrastructure, and possible lessons to adapt from industry.

[1] Schwartz, Francis, and O'Connor (2016). The Department of Defense Acquisition Workforce: Background, Analysis, and Questions for Congress, https://www.fas.org/sgp/crs/natsec/R44578.pdf

Process of Defense Test and Evaluation

Process Findings (slide 10-11)

"Overall T&E costs are unknown and not viewed as a priority: Lack of incentive to reduce cost because cannot track them directly"

This is not true; we do know the costs of operational and live fire test and evaluation. The Test and Evaluation Master Plan (TEMP), required for every acquisition program under OT&E or LFT&E oversight, has a section devoted entirely to resource allocation for both developmental and operational testing, and I have provided clear guidance and examples for this section of the TEMP in my TEMP guide, which has been published online since 2011.[2] The program manager, in coordination with all the T&E stakeholders, must identify and plan for all T&E resources needed to adequately support Developmental Test and Evaluation (DT&E), OT&E, and LFT&E. The first step is to develop data requirements in the evaluation framework and from those data requirements determine the resources needed at each stage of the program. The Resource Summary section of the TEMP should flow directly from these analyses and identify the test resources to conduct the tests described earlier in the TEMP.

In my Fiscal Year 2011 Annual Report, I examined cost data from recent acquisition programs. Existing data show that T&E cost is only a fraction of the acquisition costs of a program (typically less than 1 percent). This small relative cost stands in stark contrast with the potential savings from problems identified that can be corrected before full-rate production and the likely result that the system will work when called upon in combat. We evaluated marginal cost of operational test and evaluation to programs as a percentage of total acquisition cost. A review of 78 programs in the Army, Air Force, and Navy showed that the average marginal cost of OT&E is approximately 0.65 percent of the total acquisition cost. Few programs that we reviewed (7 out of 78) required more than 1.5 percent of program acquisition costs for OT&E. For those programs with above average OT&E costs, a relatively low program acquisition cost was the dominant cause of larger proportional OT&E cost (e.g., AIM-120C Electronic Protection Improvement Program with $87M acquisition costs). Expense of test articles and their expendability was another major driver. Figure 2 shows the distribution of the marginal cost of OT for the 78 programs we examined.

T&E costs can be inflated by adding in development costs for programs which were eventually abandoned, though investments in T&E early in system development can create savings there by identifying problems with systems before good money is sent after bad. The Decker-Wagner Report, commissioned in 2010 by the Secretary of the Army, gives concrete evidence for the need for early operational testing to illuminate problems by examining the Army's failure rate of initiating and then canceling new development programs.[3] The study found that between 1990 and 2010, the Army terminated 22 Major Defense Acquisition Programs (MDAPs), and that 15 of those terminations occurred since 2001. Further, excluding the Future Combat System (FCS), the Army spent more than one billion dollars *per year* since 1996 on programs that were eventually canceled before completion. The study cited many

[2] http://www.dote.osd.mil/tempguide/

[3] Army Strong: Equipped, Trained and Ready, Final Report of the 2010 Army Acquisition Review, January 2011.

reasons for the failed programs including unconstrained requirements, weak trade studies, and erosion of the requirements and acquisition workforce. However, none of the reasons cited included T&E. Earlier and more robust T&E would have revealed problems and solutions sooner, when they would have been less costly to fix or allowed decision makers to cancel or restructure programs and avoid wasting billions of dollars.

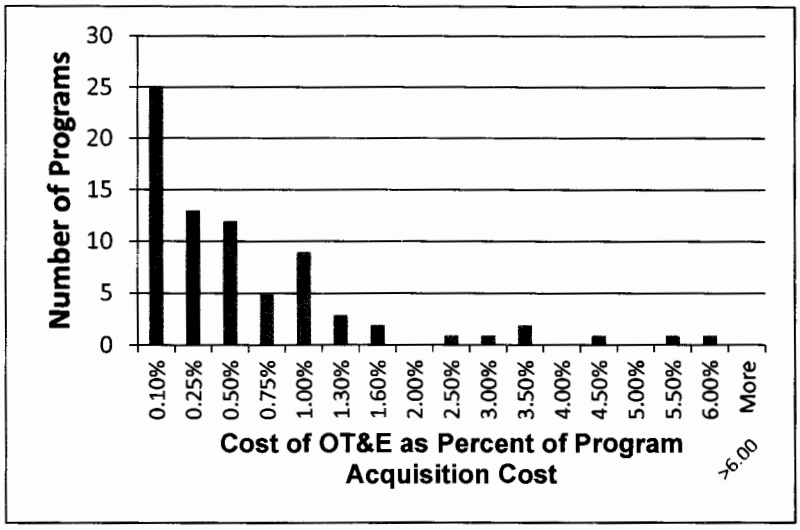


Figure 2. Marginal Cost of OT&E relative to program acquisition cost

"Philosophical difference between testing to original requirements vs. new adversary capabilities"

Programs often complain that DOT&E requires testing beyond threshold requirements, or even threshold key performance parameters (KPPs). However, our observation is that KPPs continue to provide an evaluation metric describing a level of performance that, while necessary, fails to sufficiently meet anticipated operational needs. KPPs often lack the context of the complex operational environment, including current threats. For example, the P-8A Poseidon is a maritime patrol aircraft that will replace the P-3C Orion and conduct Anti-Submarine Warfare (ASW) and other missions. However, the KPPs required only that the P-8A be reliable, be equipped with self-protection features and radios, and carry a requisite number of sonobuoys and torpedoes, but not actually demonstrate an ability to find and prosecute submarines. DOT&E, working with the Navy's Operational Test Agency, focused the testing on examining quantitative mission-oriented measures, beyond the limited KPPs, in order to characterize the aircraft's ASW capabilities. The Department of Defense Inspector General specifically criticized the Navy and Joint Staff for not promulgating requirements for the P-8 capturing the key aspects of the combat missions it is meant to conduct.[4]

Similarly, the F-35 Joint Strike Fighter, the department's largest ever acquisition program, is a fifth-generation fighter expected to operate in highly contested peer or near-peer threat environments. Between the F-35 program origins, which began from the Joint Advanced

[4] Naval Air Systems Command Needs to Improve Management of Waiver Requests, DODIG-2015-122, May 2015.

Strike Technology (JAST) program in 1993, and today, our adversaries have developed, fielded, and exported advanced integrated air defense systems – systems the F-35 was not originally designed to defeat. While the program office has a "philosophical" difference with DOT&E about testing against these threats that have in fact been fielded for a decade, it is DOT&E's statutory responsibility, and is simply common sense, to test systems in realistic operational conditions expected in actual combat. F-35 capability is critical to the U.S. prevailing in combat against existing and emerging threats worldwide. Performance of the F-35, and all combat systems, must be characterized across its operational environment so that military leaders can make informed decisions regarding their employment.

If programs were tested solely to their KPPs and ambiguous requirements documents, we often would not be able to evaluate whether systems can accomplish their primary missions. While we must always pay attention to requirements documents, we also have to interact with the operators and understand the concepts of operation and the war plans. Additionally, the Services update a program's threats on a regular basis from the intelligence information on the latest threats. All of these things will tell us how to do an operational test under the circumstances the system will actually be used in combat and enable to characterize the performance of systems across their operational envelope – not just at one key parameter. Operational testing is not a bureaucratic game to be played to obtain a "good grade." Rather it is a deadly serious activity whose purpose is to discover whether systems will perform as needed, consistent with our war plans, before they are used in combat when failure cannot be tolerated.

"Accountability for risk trade-off decisions is not clear or consistent"

A large portion of this briefing is focused on the risk calculus, which should be considered in T&E to adequately scope the T&E program. However, this study is focused on the wrong aspect of it --- it mainly targets the Services' perceived diminished role in the process of designing adequate combat testing – causing pressure on cost and schedule. However, the risk of weapons not working as required or when required in combat is not addressed. If future studies wish to truly investigate how to adequately balance resources for T&E activities, they must look at both sides of the scale. Moreover, DOT&E's reports, including its most recent Annual Report, document many significant and substantial shortfalls in operational effectiveness that never would have been known if Service proposals to "balance risk," which almost always amount to false reasons for minimizing testing, had been accepted.[5]

"Perception that drive for zero risk is too strong and uncompromising"

Testing never strives for nor achieves zero risk. DOT&E reports annually on problem discovery during operational testing – the tests conducted at the end of development before fielding. This testing reveals that nearly two-thirds of programs find significant problems during operational testing (significant enough to impact operational effectiveness, suitability, or survivability). Finding and addressing these problems before production and deployment is critical; otherwise they would be discovered in combat.

DOT&E has advocated for rigorous and scientific approach to test planning and resourcing; this scientific approach uses design of experiments (DOE) to cover the operational

[5] http://www.dote.osd.mil/annual-report/index.html

environment as efficiently as possible, allowing for us to balance risk and information. In fact, many of the tests conducted allow a high level of risk of incorrectly concluding that a system is performing adequately---which we use rigorous analytical techniques to quantify--- due to limited data. Often resources are limited up-front (when testers are not included in early resource decisions) and testers use DOE to get the most out of what they are given. DOT&E recently has been trying to define test scope/resources well-ahead of the MS B request for proposals from industry.

Furthermore, the claim that testers strive for zero risk before fielding new systems is false because DOT&E is not responsible for fielding decisions. In fact, fielding decisions are authorized by the Services, not DOT&E. It is common practice for Services to field systems with major deficiencies, transferring risk from the acquisition community to the warfighter and the nation.

"Services can feel left out of final decisions and lack authorities to make final "calls" on suitability and effectiveness"

Title 10 U.S. Code defines the responsibilities of the Director, OT&E. It is the Director's independent authority to report the operational effectiveness, operational suitability, survivability and lethality of weapons systems. The Services each have their own independent operational test agency that uses the same data but may report different overall conclusions. It is up to the Services to decide whether to do anything in response to DOT&E's findings, and as previously stated they often field the program despite the significant problems that DOT&E reports. Moreover, DOT&E works closely with the Services' operational test staffs and leadership to design realistic test. In the vast majority of cases, these interactions result in agreement between DOT&E and the Service's on the operational testing that needs to be done.

"Lack of resources when additional testing is required creates pressure on budget and schedule"

A primary purpose of operational testing is to identify critical problems that can be seen only when systems are examined under the stresses of realistic combat conditions, prior to the full-rate production decision and fielding to combat units. This identification permits corrective action to be taken before large quantities of a system are procured and avoids expensive retrofit of system modifications. The assertion that testing causes delays misses the essential point: fixing the deficiencies causes delays, not identifying them. DOT&E conducted a study of over 100 programs on DOT&E oversight that experienced delays and found that less than 25 percent of cases did a problem with test conduct delay the program; the most common reason contributing to delays was performance problems discovered during testing that had to be addressed before continuing development and fielding.[6] Furthermore, taking the time to correct serious performance problems is exactly what we desire in a properly functioning acquisition system.

"DoD and industry are engaged in addressing the emerging challenges and generally work well together; but many view DoD process as bureaucratic and overly complex"

[6] http://www.dote.osd.mil/pub/presentations.html, "Causes of program delays"

Some may view DOD process as bureaucratic and overly complex compared to industry; however, the systems DOD is developing, procuring, and fielding are the most complex and expensive in the world. Acquiring capabilities critical to the national interest – and doing so with limited public funds – necessitates a level of careful oversight above and beyond what might be required in industry. If future reports focus on identifying inefficiencies or bureaucratic processes which can be eliminated without diminishing the effectiveness of the T&E process to find and fix problems, investigating the problem instead of merely complaining about its scale, we would welcome such a study.

"Strategic Capabilities Office (SCO) offers alternatives to the current test approach: New take on the "development/test/deployment" method"

The strategic capabilities office, by definition, does not produce integrated weapon systems to be operated and sustained by a warfighter. Our experience with programs developed by the SCO is that they undergo rigorous test and evaluation just as other major combat systems do, but their deployment or early fielding is stymied by the lack of trained operators, maintainers, spare parts, and concepts of operation. Moreover, we have engaged in numerous examples of rapid testing to support rapid fielding, with the Mine Resistant Ambush-Protected Vehicle program being a notable---but by no means the only---notable example.

"Misaligned incentives drive process participants to act in certain ways: Budget (PPBE), Requirements (JCIDS), Program Management, T&E"

Programs clearly have an incentive to denounce testing as unfair when it reveals performance problems. Cost and schedule overruns, especially those that are the direct result of poor program management, reflect poorly on program managers and program executive officers. There's a terrible fear that exists that a negative DOT&E report will kill a program; however, it is much more likely that performance problems reported by DOT&E lead to a greater allocation of resources and time to fix them. However, by engaging in bureaucratic games, rationalizing problems, and minimizing testing, acquisition officials do a great disservice for the people for whom we work – the men and women in combat whose lives depend on the systems we give them to achieve their mission and the nation whose security depends on that mission's success.

<u>Process Recommendations (slides 15-17)</u>

"Create appropriate and direct accountabilities for T&E community, aligned with Program Management"

Independence is the long pole holding up DOT&E's tent – every other virtue (i.e., Accuracy, Integrity, Objectivity) flows from it. None of these would be possible if DOT&E were required to subvert its expertise and scientific investigations to bow to budgets and schedule considerations. DOT&E already takes budgets into account by using Design of Experiments to most efficiently cover the operational envelope and by using modeling and simulation to complement physical tests when doing so is scientifically justified. As stated earlier, tests are routinely planned with risk levels higher than most industry standards with substantial probability of concluding that a system is performing adequately, when it is not. By decreasing the power and thereby increasing the risk of accepting poorly performing systems, eventually the test will be a coin-toss. Is a test with only 50 percent power worth conducting?

The ultimate outcome of making T&E accountable to program management is to turn T&E into a simple box-checking exercise. Consider the contractor-run KC-46A Aerial Refueling Tanker test program. This system was purchased under a firm fixed-price contract where the contractor, Boeing, ran the integrated test program. Boeing consistently produced unrealistic and overly-optimistic schedules as analyzed in DOT&E annual reports since 2011. Reality drove the schedules past planned calendar dates leading Boeing program management to cut testing. As noted in DOT&E's KC-46A OA-2 report, Boeing's original plan was to have 75 percent of testing completed by Milestone C, but in the end, a mere 25 percent of testing was completed even though the milestone was delayed by over 12 months. Boeing only accomplished the bare minimum of events needed to check off the items written in the Milestone B Acquisition Decision Memorandum. While demonstrating the last of these items, refueling using the boom with three different receivers, a serious problem emerged. Loads within the boom were approaching structural limits while pushing and pulling on the receiver. Boeing drove to continue the demonstrations and proceed to Milestone C and the Low-rate Initial Production (LRIP) portion of the contract since they did successfully offloaded fuel. After pressure from the 412th Test Wing, Air Force Operational Test and Evaluation Center (AFOTEC), the KC-46A Government Program Office, and the A-10 program office which withdrew their clearance to test letter, Boeing postponed testing and redesigned the boom. In the process, Boeing had a safety stand-down day to internally discuss how they missed early indications of the excessive loads in collected data while they sped ahead to complete testing.

"Definitions of "success" should be agreed between PM and T&E community from the beginning and with changes that occur over time"

During the development of requirements and the concept of operations, the PM and warfighter communities agree on the definition of success. T&E tests to this definition.

"Sharply limit the number and weight of stakeholders that have veto power over TEMP development, implementation, and modification"

This recommendation is not specific enough to be useful – but it should be noted that the Services generally have 15-20 signatures on a TEMP while OSD has only one or two. For example, the recent TEMP for the Global Broadcast Service (GBS) required 22 signatures from the Services, including PMs, Program Executive Officers, requirements authors, and Operational Test Agency (OTA) commanders, but only one signature from OSD – mine.[7]

"Authorities should come from the Service Secretary to the Service Chief and designated subordinates"

This is true today. Service Acquisition Executives work for the Service Secretaries. They make decisions to buy systems. Service Chiefs and subordinates make decisions to field systems – with or without serious deficiencies.

[7] Test and Evaluation Master Plan for Global Broadcast Service, 21 March 2016

"Today's drive toward "zero risk" is an impossible standard and should not be used by the T&E community"

This claim is silly and false. Zero-risk is not used by the T&E community and never has been. However, if the DBB briefing is confusing T&E with requirements, I completely agree that we should move towards stating and quantifying "tolerable risks" in the requirements documents, both in terms of allowable performance degradations and our ability to know them. For many programs, I have engaged early with requirements writers to determine whether 99.99 percent probability-based requirements are necessary, because both contracting for such systems and testing them would be extremely expensive. For example, 99.99percent reliability requirement is extremely expensive to test if we want to ensure we are not actually 99.98 percent. Whereas requiring 90 percent within +/- 5percent is achievable in a relatively short test.

"Adopt "principled compromise" between absolute zero-defect test doctrine and situational operating requirements"

The claim that there is a zero-defect test doctrine is demonstrably false. Operational testing characterizes system performance across its intended operating environment. "Zero-defect" is not a test planning goal, nor could it ever be possible to test to "zero-defects." The test doctrine does not determine the number of allowable defects. The requirements do. The test doctrine then determines how many tests we need to do to prove out that requirement with acceptable statistical risk. Operational testing uses sound test science principles to design efficient and adequate testing, meant to determine the actual performance of systems in as close to operational environments as possible. Implementation of rigorous test design and analysis provides defensible, factual information to scale test designs and to provide credible test results on system performance to the Congress and civilian and military leaders so that they could make informed decisions regarding acquisition and employment of those systems.

"Tolerable" risk should be stated and quantified"

It is. Design of Experiments (DOE) provides a defensible and efficient methodology for not only determining test adequacy but also ensuring that we obtain the maximum value from scarce test resources. DOE has been proven to elicit the maximum information possible from constrained resources, provide the ability to combine information across multiple independent test events, and produce defensible rationale for test adequacy and quantification of risk as a function of test size. One clear advantage of statistical approaches to evaluating test adequacy is that they provide a means to quantify how much information can be derived from each test point.

"Apply more consistent analysis of common data in the T&E process"

Our reports demonstrate clearly that we use to the fullest all the data that are available on operational performance under realistic conditions, to do otherwise would be stupid. The Services and DOT&E provide independent analyses and evaluation. My office is working towards developing analysis tools that can be used by all Services, but ultimately the evaluations and supporting analyses between oversight and the Services should remain independent.

"Services are best positioned to determine "good enough" when program length results in new threats that didn't exist when requirements set"

Actually, Service acquisition professionals are not best positioned to determine "good enough," especially considering that the current acquisition system incentivizes them to stay on cost and schedule regardless of the performance of their system. Before Congress created the office of DOT&E, the Services provided inaccurate reports of system performance, took shortcuts in testing because of budget pressures, and conducted inadequate testing without realistic combat conditions and operationally realistic threats. We would do well not to forget the lessons we learned in that time. Whenever a program's requirements were established, it must operate in combat conditions present today and in the future. Operational testing against new threats will inform Commanders in which environments the system can operate. Moreover, in instances when the threats against which we should test are contentious, as in the case of the Joint Strike Fighter, DOT&E has taken those issues to the Secretary of Defense for decisions. In all cases the current Director has done this, the Secretary has decided to adopt the Director's recommendation.

Early Testing

The final process recommendation deals specifically with costs, so it seems appropriate to mention here that the report ignores recent "shift left" strides in T&E involvement. T&E needs to be involved earlier in process to ensure appropriate resources and test conditions are part of the Requests for Proposal (RFPs) sent to industry.

DOT&E has long advocated for earlier realistic testing and problem discovery so that acquisition decision makers can make timely decisions. To help avoid expensive programs continuing in development while not delivering military utility, DOT&E now requires operational assessments (OAs) for all programs prior to the Milestone C production decision, when problem discoveries may highlight significant mission shortfalls and problems are cheaper to fix.

Early testing (both developmental test events and OAs) should inform the development process and enable the early identification of major problems. More than just providing an early opportunity for problem detection, OAs provide a chance to build knowledge on how the system will perform once placed in an operational environment. The use of design of experiments, even in early testing, allows efficient test designs that cover the operational envelope. Knowledge gained from OAs can help refine the resources necessary for the Initial Operational Test and Evaluation (IOT&E), such as the most significant factors affecting operational performance, potentially reducing the scope for the IOT&E. In ideal cases, the use of sequential test design from early testing including OAs through IOT&E can provide even more efficient use of test budgets by combining information across test phases.

Industry (Slides 13-14)

The briefing slides appear to show a fundamental misunderstanding of the different incentives between commercial industry and defense when it comes to development and acquisition of new products. In the commercial sector, companies are motivated by the threat of losing customers and failing to make profits; therefore, they are motivated to find and fix any problems before selling their products. The Defense Department, however, is the sole customer of highly complex systems, often bought in low quantities (compared to the commercial sector). The government program managers of these systems are incentivized to hide problems in order to keep their program within the resourced cost and schedule and to avoid informing leadership in the Department and the Congress of problems that might lead to funding reductions.

"Industry in general is further along in the application of software and simulation for T&E"

The claim is that "industry in general" is ahead of government in using modeling and simulation (M&S) for T&E – models reduce cost and time to market and are validated and supplemented by historical live testing data and select physical tests. However, commercial and national defense products face very different challenges in applying historical data to new systems. In defense products, historical data may not apply to new systems. Some phenomena, including especially new advances in threat systems, cannot be simulated without having collected live data first.

"Successful companies ... include leadership from business and engineering organizations who together make risk tradeoff decisions, with business leader generally holding the "final vote""

As I have previously stated, in no phase of defense T&E is a tester required or asked to make acquisition decisions. DOT&E finds and presents information to the Services, which then make decisions regarding fielding. Weakening the independence of OT&E and LFT&E would not make this process any quicker or less contentious, but it would risk decreasing the quality of the performance information taken into consideration at the time of a fielding decision.

In the KC-46A example described above, the independent action of the Air Force test agencies led the contractor program management to halt unsafe testing and redesign the system. Fortunately, testing discovered the system deficiency early, and the only casualty was an already-unrealistic schedule. Boeing, a premier defense contractor (with a similarly reputable commercial industry) is reported to be losing billions of dollars over their mismanagement of the KC-46 test program.[8]

Finally, as noted above, incentives in industry---and for its leadership--- to find and fix problems are essentially the opposite of the incentives facing the Department's leadership.

[8] http://www.defensenews.com/story/defense/air-space/air-force/2016/07/21/boeing-kc46-cost-overrun-penalty-air-force-boom/87409004/

Infrastructure (Slide 12)

The independence of DOT&E allows us to require adequate and realistic operational testing and to advocate for resources to improve our T&E capabilities. I have observed that some of the most important capabilities or tests that we have prescribed have been met with substantial resistance from the Services, sometimes requiring adjudication by the Deputy Secretary of Defense. In light of the remarkable resistance from the Services to prioritize adequate testing and test assets in their acquisition programs, it is even more apparent that the independence of this office is critical to the success of finding problems before systems are used in combat.

An alarming trend I have seen during my tenure is that our threats are increasing their capabilities faster than our test infrastructure can adapt to them. Through the yearly budget review process, I have advocated for resources to improve test range infrastructure to support rigorous testing of modern combat systems. Most notably, in 2012, I convinced the Department to invest nearly $500M in the Electronic Warfare Infrastructure Improvement Program (EWIIP) to upgrade open-air test ranges, anechoic chambers, and reprogramming laboratories in order to understand the performance of the F-35 Joint Strike Fighter and other advanced air platforms against near-peer threat integrated air defense systems. The open-air test and training ranges owned and operated by both the Air Force and Navy lack advanced threat systems that are being used in combat by our adversaries today, are proliferating, or are undergoing significant upgrades; yet both services strongly resisted incorporating these modern threats that we proposed until directed to do so by the Deputy Secretary.

Other significant T&E shortfalls that I have highlighted include: Fifth Generation Aerial Target, Self Defense Test Ship, Multi-stage Supersonic Target and the Warrior Injury Assessment Mannequin for assessing force protection of ground combat vehicles to underbody blast events.

The conclusion that the infrastructure is not fully postured for the future is perhaps the sole, valuable contribution of this briefing. A follow-on study should focus on assessing the state of the art of our T&E capabilities (to include M&S) in the context of existing and anticipated theaters of operations: what are the T&E shortfalls and what does the Department need to do to close those capabilities gaps --- what would it take, how long would it take, how much would it cost and what would it buy us. The Air Force Science Advisory Board is conducting such a study now.[9] A business plan for that would also likely highlight any T&E management shortfalls that could then be more specifically addressed.

J. M. Gilmore

J. Michael Gilmore
Director

[9] Air Force Science Advisory Board, "Adapting Air Force Test and Evaluation to Emerging System Needs," Proposed 2017 study.

Defense Business Board

TAB E

BACK-UP SLIDES / APPENDICIES

APPENDIX

Key Terms

- T&E – Test and Evaluation
- TEMP – Test and Evaluation Master Plan
- DT – Developmental Test
- OT – Operational Test
- DOT&E – Director, Operational Test and Evaluation, OSD
- DASD DT&E – Deputy Assistant Secretary of Defense for Developmental Test and Evaluation
- TRMC – Test Resource Management Center
- MRTFB – Major Range and Test Facility Base
- PM – Program Manager
- PEO – Program Executive Officer

Interviews

- Service Chiefs
- Director, OT&E, OSD
- DASD DT&E / Director, TRMC
- Service Acquisition Executives
- Service T&E Leads
- Service Operational Test Activities
- Major Research and Test Facilities
- Director, Strategic Capabilities Office
- Government Accountability Office
- Defense Acquisition University
- Select PEO/PMs from all services
- NDIA T&E Working Group
- Industry
  - Pepsi
  - Raytheon
  - Boeing
  - Northrop Grumman
  - General Motors
  - General Electric

Documents and Briefings

Prior DSB Studies

- DT&E - MAY 08
- Test and Evaluation - SEP 99

GAO Studies

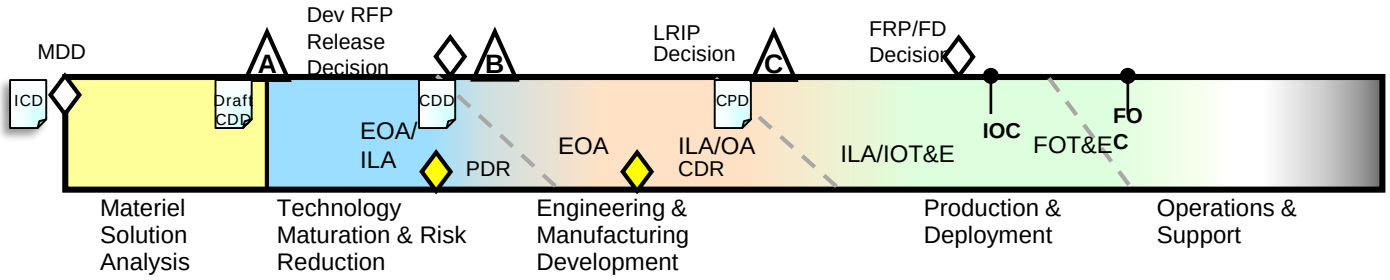
- Several around the early 1990s
  - Focus on TEMP, processes within OT&E, DT&E
  - Resulting from NDAA or Defense Appropriation statute or report language directing a study into an aspect of T&E
- Service Chief Involvement in Acquisition – MAY 14
- DOD Operational Testing - JUN 15

DoD Reports, Papers, Policy and Instruction

- DOT&E Annual Reports
- DAU – Test and Evaluation Management Guide
- Army Report to Congress on Chief of Staff of the Army Acquisition Authorities - MAR 16; Section on T&E responsibilities of Service Chief
- Incorporating Test and Evaluation into DoD Acquisition Contracts
  - OCT 11
- Performance of the Defense Acquisition System (Annual Report)
  - SEP 15
- DOT&E "Reasons for Program Delay" Briefing - 2014
- DASD DT&E Test and Evaluation "Serving the Warfighter in a Cost Constrained Environment" - MAR 11
- "Briefing to the DBB T&E Task Group" – Dr Brown, DASD DT&E - MAR 16
- MRTFB Map and Data - JUN 16
- IDA Report on T&E Cost – 2014
- DoDD 5000.02
- DoDD 5141.02
- DoDD 5105.71
- DoDD 3200.11

Statute

- Joint Explanatory Statement in FY15 NDAA
  - Directs GAO to review DOT&E oversight activity
  - Looks at disputes between DOT&E and OT
  - Report Produced
- P.L. 98-94, Sep 1983 – DOT&E establishment
- 10 USC 139
- 10 USC 2399
- 10 USC 2366



T&E Activities

• Designate CDT, LDTO, & LOTA • Review draft CDD for testability • Determine T&E Strategy • ID test resources • Designate test events for each MS and develop exit criteria • Describe evaluation methodology • Assess T&E impacts of CONOPS/OMS/ MP/ use cases	• Develop TEMP • Identify EMD AS T&E risks/ opportunities • Conduct prototype testing • Conduct EOA or LUT • Conduct early Reliability testing • Provide projected and observed software maturity metrics • Conduct T&E to support modeling & simulation • Conduct BCA on use of MRTFB facilities	• Update TEMP • Identify Production AS T&E risks/opportunities • Conduct DT&E & OA on developmental articles • Conduct coupon, subsystem, and major system LFT&E • Conduct cybersecurity testing • Plan/conduct QT • Plan/conduct PQT • Plan/conduct FAT • Plan/conduct RGT • Update models & simulations • Evaluate the system using HWIL • Conduct interoperability testing	• Conduct IOT&E on production representative system • Complete FUSL LFT&E • Plan and execute FOT&E (if required) • Plan T&E for PIP or next increment • Conduct FAT • Conduct PQT	• OA of modifications testing • Safety testing • Reliability improvement testing • Conduct FAT • Conduct QT • Conduct FOT&E

Documentation and Reviews

• Acquisition Strategy • Draft CDD/CONOPS • MS A TEMP • Initial PPP, LCSP • Cybersecurity T&E strategy • Initial Reliability Growth T&E strategy	• System (Performance) specification • CDD/OMS/MP/Use Cases • MS B TEMP, PPP, LCSP • Request FUSL LFT&E Waiver (if applicable • DSAD(DT&E) MS B Program Assessment • **PDR**	• CPD • System (Performance) and product specifications • Interface Control Documents • DSAD(DT&E) MS C Program Assessment • MS C TEMP, PPP, LCSP • OAR & LFT&E Report • **CDR, TRR**	• CPD • DOT&E IOT&E Report • DOT&E LFT&E Report • FOT&E Report (if required) • OA of modifications testing reports • Safety testing reports • Reliability improvement testing reports • **PCA, OTRR**	• OA of modifications test reports • Safety test reports • Reliability improvement test reports

Source: DAU

Acronyms

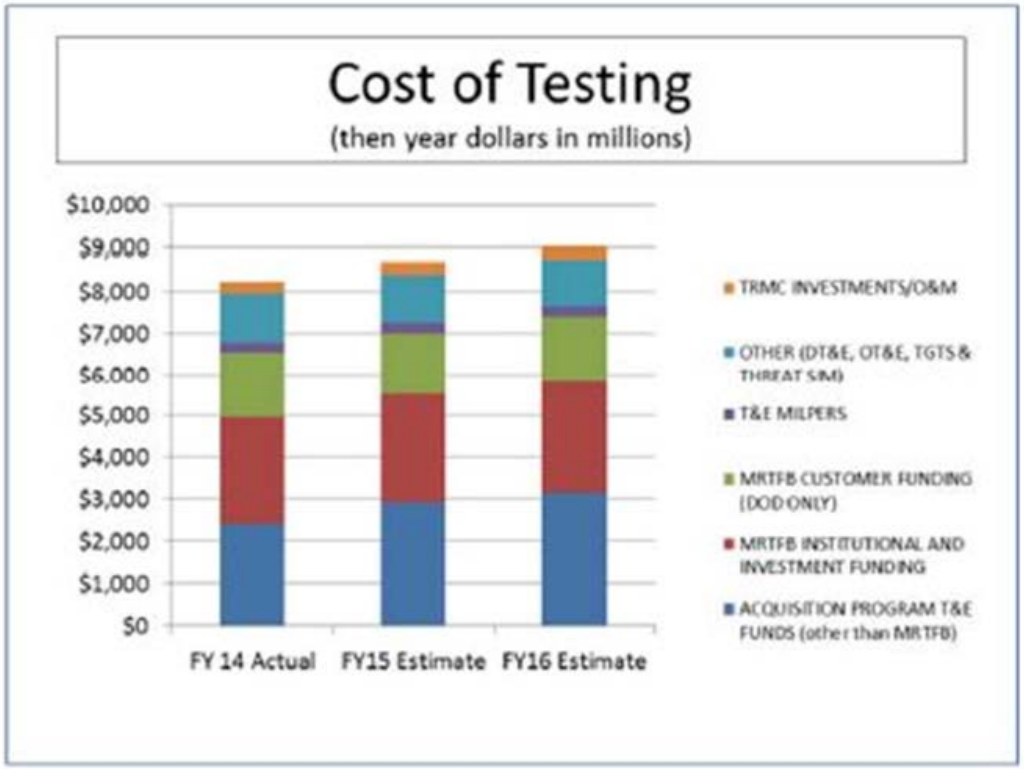
AS: Acquisition Strategy
ASR: Alternative Systems Review
CDD: Capability development document
CDR: Critical Design Review
CDT: Chief Developmental Tester
CPD: Capability Production Document
CONOPs: Concept of Operations
DEF: Defense exportability features
EOA: Early Operational Assessment
FAT: First Article Testing
FEMCA: Failure Modes Effects Criticality Analysis
FOC: Final Operational Capability
FOT&E: Follow-on Test & Evaluation
FUSL: Full up System Life-fire
HWIL: Hardware –in-the-Loop
ICD: Initial Capabilities Document
ILA: Independent Logistics Assessment
IOC: Initial Operational Capability
IOT&E: Initial operational test & evaluation
LCSP: Life-Cycle Sustainment Plan
LDTO: Lead Developmental Test Organization
LOTA: Lead Operational Test Agency
LRIP: Low-Rate Initial Production
LUT: Limited User Tests
MDD: Materiel Development Decision
MRTFB: Major Range Test Facilities Board
M&S: Modeling and Simulation
OA: Operational assessment
OMS/MP: Operational Mode Summary/Mission Profile
OTRR: Operational Test Readiness Review
PCA: Physical Configuration Audit
PDR: Preliminary Design Review
PIR: Post-implementation Review
PPP: Program Protection Plan
PQT: Production Qualification Testing
QT: Qualification Testing
R&D: Research & development
RGT: Reliability Growth Testing
SEP: System Engineering Plan
TEMP: Test and Evaluation

Infrastructure Headcount FY15

		Mil	Civ	Contr
ARMY				
	Aberdeen Test Center	14	657	691
	Electronic Proving Ground	39	166	366
	Kwajalein Atoll Reagan Test Site	4	35	357
	West Desert Test Center	3	351	168
	White Sands Test Center	8	793	708
	Yuma Test Center	21	414	988
	Cold Regions Test Center	4	34	10
	Tropics Regions Test Center	0	4	15
NAVY				
	Atlantic Undersea Test & Evaluation Center	54	53	592
	Naval Air Warfare Center (NAWC) - Aircraft Division, Patuxent River	338	761	1518
	NAWC - Weapons Division, China Lake and Point Mugu	121	832	810
	Naval Undersea Warfare Center, Keyport	0	78	5
	Pacific Missile Range Facility	0	7	54
AIR FORCE				
	Arnold Engineering Development Complex	40	211	1781
	Nevada Test & Training Range	57	98	450
	Utah Test & Training Range	11	143	149
	30th Space Wing	410	181	676
	45th Space Wing	87	218	1255
	96th Test Wing	575	1074	1683
	96th Test Group	77	256	213
	412th Test Wing	1555	2879	578
DISA				
	DISA/Joint Interoperability Test Center	70	333	598
TOTALS		**3488**	**9578**	**13665**

Note: All data reflect FY15 year end actuals

Source: TRMC



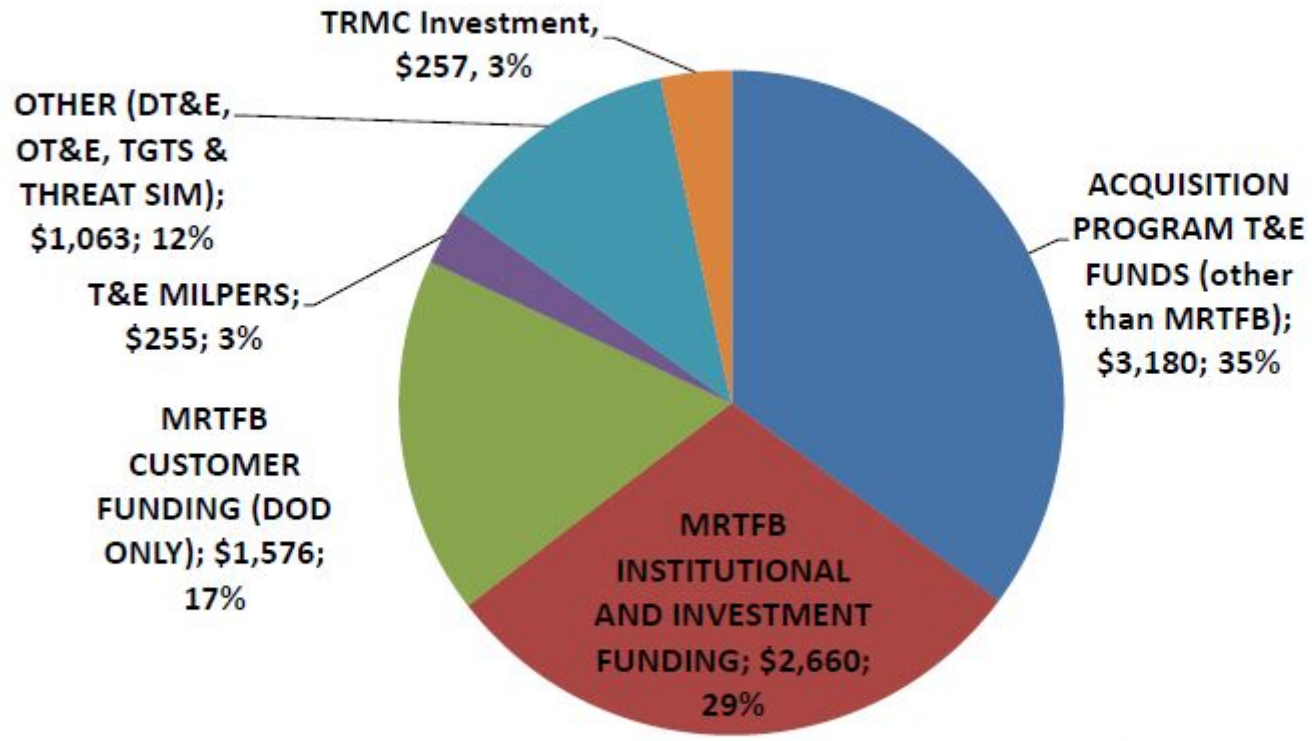
Source: Institute for Defense Analysis

The data are:

	FY 14 Actual	FY15 Estimate	FY16 Estimate
ACQUISITION PROGRAM T&E FUNDS (other than MRTFB)	$2,430	$2,932	3180.34
MRTFB INSTITUTIONAL AND INVESTMENT FUNDING	$2,542	$2,583	2659.689
MRTFB CUSTOMER FUNDING (DOD ONLY)	$1,569	$1,505	1575.621
T&E MILPERS	$250	$255	255.202
OTHER (DT&E, OT&E, TGTS & THREAT SIM)	$1,182	$1,110	1062.691
TRMC INVESTMENTS/O&M	$257	$320	311.714
Total	$8,231	$8,704	$9,045

Cost of Testing

FY16 Estimate, $ in millions

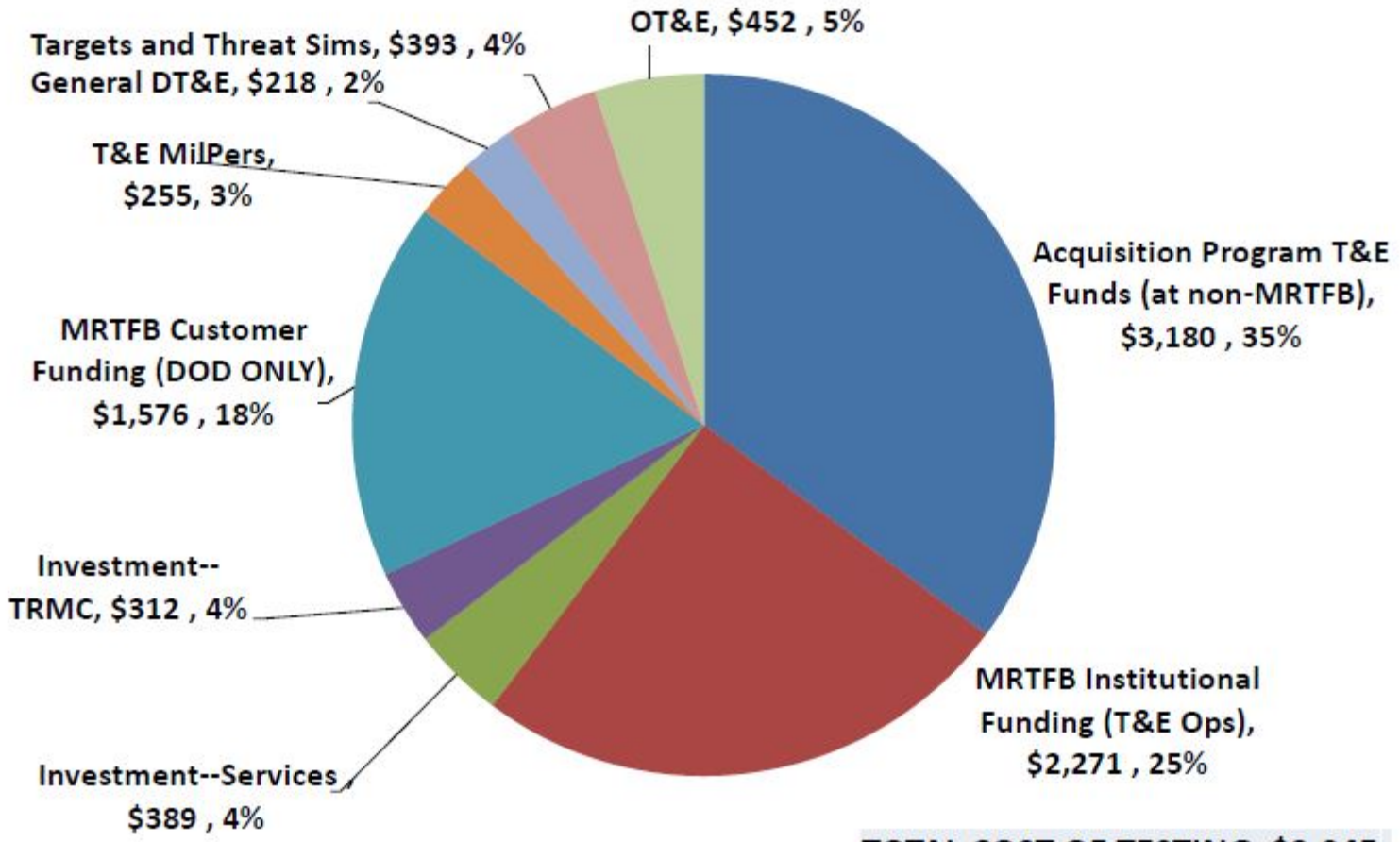


TOTAL COST OF TESTING	$9,045

Source: Institute for Defense Analysis

Cost of Testing

FY16 Estimate, $ in millions, alternate structure



TOTAL COST OF TESTING $9,045

Source: Institute for Defense Analysis

Defense Business Board
1155 Defense Pentagon
Room 5B1088A
Washington, DC 20301-1155
571-256-0835

http://dbb.defense.gov/

Roma K. Laster, Executive Director
Webster E. Bridges III, Deputy Director
Steven M. Cruddas, Office Manager