CLEARED
For Open Publication

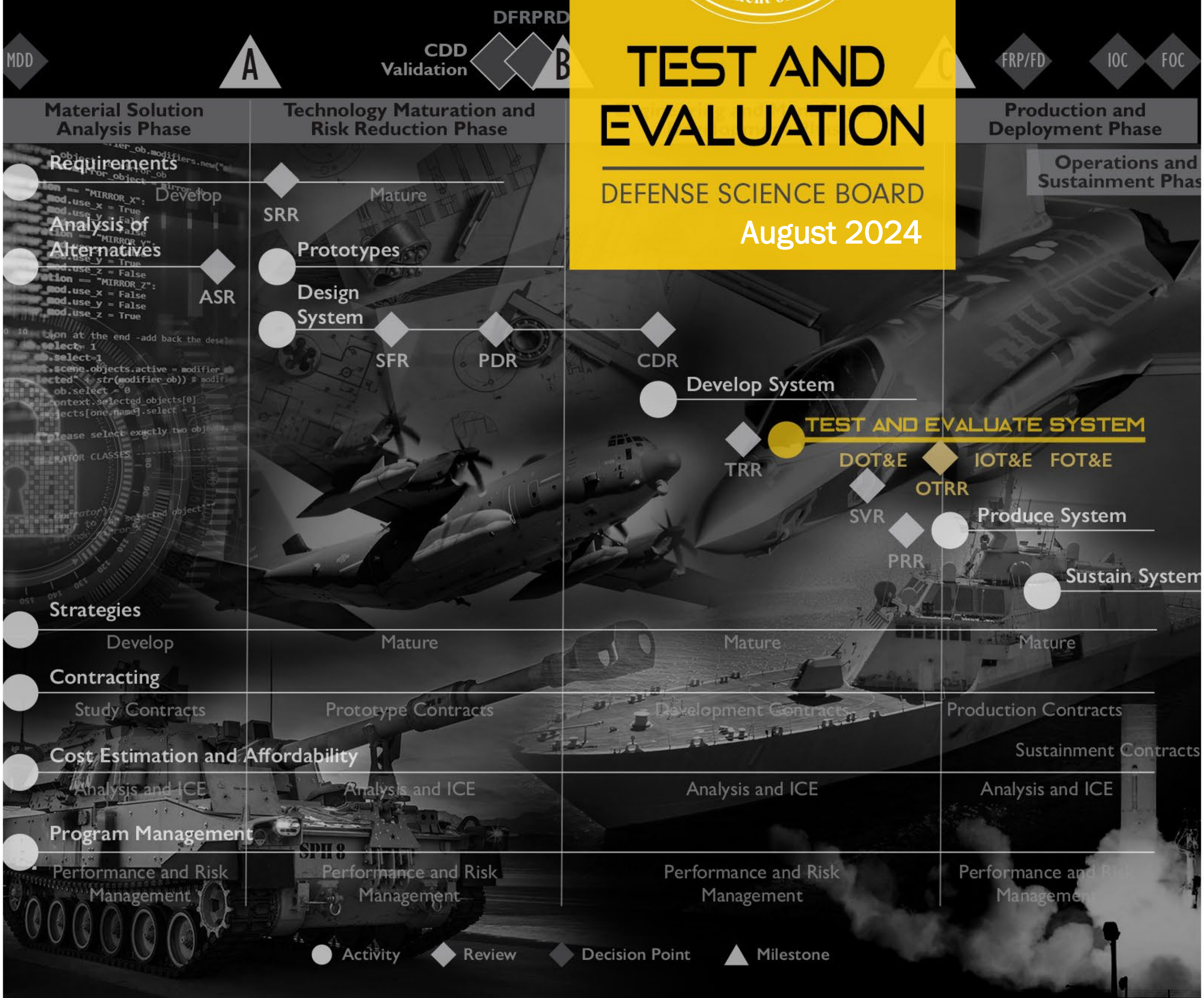
Aug 16, 2024

Department of Defense
OFFICE OF PREPUBLICATION AND SECURITY REVIEW

TEST AND EVALUATION

DEFENSE SCIENCE BOARD

August 2024



DEFENSE SCIENCE BOARD
UNDER SECRETARY OF DEFENSE FOR RESEARCH AND ENGINEERING

DEFENSE SCIENCE BOARD

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

MEMORANDUM FOR UNDER SECRETARY OF DEFENSE FOR RESEARCH AND ENGINEERING

SUBJECT: Defense Science Board (DSB) Report on Test and Evaluation

I am pleased to forward the final report of the DSB *Test and Evaluation (T&E)* study.

As requested in the Terms of Reference, the DSB was tasked to evaluate today's Department of Defense (DoD) T&E activities and explore opportunities for increasing T&E speed and efficiency and provide findings and recommendations. Starting with today's acquisition-based T&E and the Adaptive Acquisition Framework, the Task Force was also asked to explore statutory authorities and oversight organization responsibilities associated with Developmental Test & Evaluation (DT&E), Operational Test & Evaluation (OT&E) functions and needed T&E infrastructure.

The final report addresses the taskings in the Terms of Reference and provides five major findings and recommendations calling for a strategic shift in DoD that moves beyond today's acquisition-based framework to better: integrate science and technology; prepare for testing of emerging new technologies; move to continuous testing and threat assessment; and expand T&E to seamlessly flow into training and operations. Four additional findings and recommendations are provided that address key enablers for achieving this strategic shift, including increasing use of digital engineering for T&E, modernizing the T&E infrastructure, ensuring a trained T&E workforce, and refining the roles of DoD's oversight organizations.

I fully endorse all the study's recommendations and urge their careful consideration and adoption.

Dr. Eric D. Evans
Chair, Defense Science Board

THIS PAGE LEFT INTENTIONALLY BLANK

DEFENSE SCIENCE BOARD

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

MEMORANDUM FOR THE CHAIR, DEFENSE SCIENCE BOARD

SUBJECT: Report of the Defense Science Board (DSB) Task Force on Test and Evaluation

Attached is the final report of the congressionally directed DSB Task Force on *Test and Evaluation (T&E)*. Our task force was asked to evaluate today's Department of Defense (DoD) T&E activities in the context of the Adaptive Acquisition Framework. We were also asked to examine DoD and commercial approaches to T&E and make recommendations to improve test outcomes and improve efficiencies. Finally, we were asked to explore and make recommendations regarding T&E oversight organizations within the Office of the Secretary of Defense to improve effectiveness.

This report addresses the role of T&E as a component of a systems engineering process that produces military capabilities with sufficient confidence to be operated in complex environments. The report delves into the differences between developmental testing, which is aimed at ensuring a system can meet its design requirements, and operational testing, which is aimed at ensuring the system meets its operational intent. We explored challenges encountered with today's acquisition-based framework for T&E and made observations concerning opportunities for increased efficiencies. We also observed challenges with T&E in software-intensive weapons systems, emerging new technologies, and joint warfighting mission threads.

To adequately respond to today's challenges, we call for a strategic shift in the DoD's approach to T&E. We provide five findings and recommendations concerning the need to expand T&E beyond today's acquisition-based framework, moving to continuous development, testing, and assessment of adversary capabilities. We further provide recommendations regarding software-intensive system development and approaches to linking advanced technology development and testing strategies.

We also identify four enablers to achieving this strategic shift in T&E and we provide findings and recommendations for each. These enablers include improving the use of digital engineering for T&E, modernizing the T&E infrastructure, ensuring a trained T&E workforce, and refining the DoD organizational oversight of the T&E enterprise.

Dr. David Van Wie
Co-chair

Dr. Johney Green
Co-chair

THIS PAGE LEFT INTENTIONALLY BLANK

DSB Report on Test and Evaluation

Table of Contents

DSB Report on Test and Evaluation

Executive Summary

As requested in the National Defense Authorization Act (NDAA) for Fiscal Year (FY) 2022 (Public Law 117-81) and tasked by the Honorable Heidi Shyu, Under Secretary of Defense for Research and Engineering (USD(R&E)), the Defense Science Board (DSB) Task Force on *Test and Evaluation* conducted a study of the resources and capabilities of the test and evaluation organizations, facilities, and laboratories of the Department of Defense. The Task Force addressed the effectiveness of current developmental testing, operational testing, and integrated testing within DoD in meeting statutory objectives and the test and evaluation requirements of the Adaptive Acquisition Framework. The Task Force also examined industry and government best practices for conducting testing and evaluation (T&E).

This report addresses the role of T&E as an essential component of a structured systems engineering process that has proven to be capable of producing military systems with sufficient confidence to be operated in complex environments. Despite these past successes, today's acquisition-based approach to T&E is being challenged by rapidly increasing adversary capabilities, emerging new technologies, and significant commercial capabilities. These challenges are being amplified when dealing with the acquisition of software-intensive weapon systems and the development of joint warfighting mission threads that require a system-of-systems integration, especially when that integration crosses Military Service acquisition boundaries reliant on capabilities at different states of maturity.

To adequately respond to today's challenges, a strategic shift in DoD's approach to T&E is needed. Five major findings and recommendations are provided concerning this strategic shift in T&E to move beyond today's acquisition-based framework towards an approach based on continuous development, continuous testing, and continuous assessment of adversary capabilities.

Emerging new technologies provide unique challenges as these technologies stress today's approaches to T&E. The Task Force explored three emerging technologies (hypersonics, directed energy, and artificial intelligence) to investigate issues regarding future T&E needs and found that T&E approaches need to mature in parallel with the underlying technologies. Further efficiencies can be realized, and acquisition of new capabilities can be accelerated by closing gaps between science and technology and the beginning of acquisition programs.

The Task Force also found that additional efficiencies can be realized by closing gaps between operational testing and training. As significant insight into the performance of a system occurs after the completion of the T&E process, expanding the concept of T&E to encompass this knowledge will increase the speed and effectiveness of new capability development.

Four enablers were identified for achieving this needed strategic shift in T&E: improving the use of digital engineering for T&E, modernizing the T&E infrastructure, ensuring a trained T&E workforce, and refining the DoD organizational oversight of the T&E enterprise. Recommendations for addressing these enablers are provided.

THIS PAGE LEFT INTENTIONALLY BLANK

DSB Report on Test and Evaluation

1. Introduction

As an essential element of system development, effective test and evaluation (T&E) is critical to the delivery of new DoD capabilities that must operate in stressing environments often characterized by significant uncertainty. The DoD T&E enterprise has evolved over decades as a component of acquisition to produce systems and capabilities that are effective, reliable, and safe to operate.

The 2022 National Security Strategy[1] states that "we are now in the early years of a decisive decade for America and the world. The terms of geopolitical competition between the major powers will be set." The strategy further states:

> "We will: 1) invest in the underlying sources and tools of American power and influence; 2) build the strongest possible coalition of nations to enhance our collective influence to shape the global strategic environment and to solve shared challenges; and 3) modernize and strengthen our military so that it is equipped for the era of strategic competition with major powers, while maintaining the capability to disrupt the terrorist threat to the homeland."

Building from the National Security Strategy, the National Defense Strategy[2] identifies four top-level defense priorities as 1) defend the homeland; 2) deter strategic attacks against the United States, our allies, and our partners; 3) deter aggression and be prepared to prevail in conflict, when necessary; and 4) build a resilient joint force and defense ecosystem to ensure our future military advantage.

With this return to Great Power competition, challenges are being faced by DoD warfighting capabilities in all domains. The demand for effective solutions to address rapidly advancing adversary capabilities has resulted in a need to accelerate DoD system acquisition in a way that leverages both emerging technologies and new approaches for system integration. As an element of new capability and system acquisition, the DoD T&E enterprise is similarly being asked to increase speed and efficiency while working to adapt historical T&E approaches to emerging technologies and novel approaches to system integration.

As requested in the National Defense Authorization Act (NDAA) for Fiscal Year (FY) 2022 (Public Law 117-81), the Honorable Heidi Shyu, Under Secretary of Defense for Research and Engineering (USD(R&E)) tasked the Defense Science Board (DSB) with conducting a study on the resources and capabilities of the test and evaluation organizations, facilities, and laboratories of the Department of Defense. The study Terms of Reference (ToR), which is provided in Appendix A, includes the following taskings:

- Assess effectiveness of current developmental testing, operational testing, and integrated testing within DoD in meeting statutory objectives and test and evaluation requirements of the Adaptive Acquisition Framework (AAF).
- Identify industry and government best practices for conducting developmental testing, operational testing, and integrated testing and determine potential applicability of industry

---

[1] "National Security Strategy," The White House, October 2022.

[2] "2022 National Defense Strategy of the United States of America," Department of Defense, October 27, 2022.

and government best practices for testing within the Department to improve test and evaluation outcomes.

- Identify duplication of efforts and other non- or low-value added activities that reduce speed and effectiveness of test and evaluation activities.
- Assess the research, development, test, and evaluation infrastructure master plan required under Section 252 of the NDAA for FY 2020.
- Assess test and evaluation oversight organizations within the Office of the Secretary of Defense (OSD), including their authorities, responsibilities, activities, resources, and effectiveness, including with respect to acquisition programs of the military departments and Defense Agencies.
- Develop and assess potential courses of action to improve the effectiveness of oversight of developmental testing, operational testing, integrated testing activities, and test and evaluation resources within OSD, including as one such course of action establishing a single integrated office with such responsibilities.
- Develop such recommendations as the Defense Science Board may have for legislative changes, authorities, organizational realignments, and administrative actions to improve test and evaluation oversight and capabilities, and facilitate better test and evaluation outcomes.

In response to this tasking, the DSB established the *Task Force on Test and Evaluation* (referred to as the Task Force in this document) to conduct this study and draft findings and recommendations. The membership of the Task Force, which is provided in Appendix B, drew on expertise from government, universities, industry, national laboratories, and Federally Funded Research and Development Centers (FFRDCs). Following completion of the activities of the Task Force, the complete set of findings and recommendations were formally adopted by the full Defense Science Board operating under the terms of the Federal Advisory Committees Act.

During the conduct of this study, the Task Force held eleven meetings collecting input and perspectives on today's approach for DoD test and evaluation and suggestions for areas of needed improvement. A wide range of inputs were solicited, covering the perspectives of DoD and Military Services T&E organizations, laboratories, FFRDCs, and industry. A complete listing of the input received by the Task Force is provided in Appendix C.

This report is organized in four main chapters following this introduction. Chapter 2 provides an introduction of today's T&E, discussing the role of T&E in system development, the approach to T&E currently being used by DoD, and a brief summary of industry best practices for T&E.

Chapter 3 addresses the T&E challenges being faced by DoD today. These challenges are discussed in the context of the AAF and the legislated acquisition authorities used by DoD in exercising its T&E roles. The challenges associated with both emerging new technologies and mission-focused capabilities are also addressed.

Chapter 4 provides the major findings and recommendations calling for a needed strategic shift in DoD T&E to both address noted challenges and to enable DoD to develop new capabilities at the pace needed in Great Power competition. Five findings and recommendations identify the needs and solutions paths for addressing the strategic shift in T&E. The Task Force also identified a set of key enablers for realizing the needed strategic shift in DoD T&E and provided four additional findings and recommendations regarding these enablers.

2. Today's Test & Evaluation

2.1 Test & Evaluation (T&E) as a Critical Component of Systems Engineering

Test and evaluation is an essential element of system development, where it is closely tied to the concept of verification and validation (V&V) of the developed system. Used in this context, verification and validation focus on different aspects of the system under development, as follows:

- Verification – Focuses on ensuring that the system requirements are met.
- Validation – Focuses on ensuring that the system performs as intended.

As illustrated in the classic systems engineering "Vee" shown in **Figure 1**, formal approaches to the development of complicated systems have evolved to ensure that the system under development meets well-defined system requirements. The left side of the "Vee" produces high-level system requirements, beginning with feasibility studies and exploration of concepts that are then matured into a Concept of Operations (ConOps) for the new system to be developed. This ConOps is decomposed through a design process that results in a detailed design capable of meeting the high-level requirements. Through this left side of the "Vee", system V&V plans are developed. The verification plans are developed for the overall system and its subsystems, units, and devices. After the integrated system is verified to meet its design requirements, system validation occurs to demonstrate that it meets its operational intent. Following validation of the developed system, systems engineering functions continue through operations, maintenance, and upgrades.

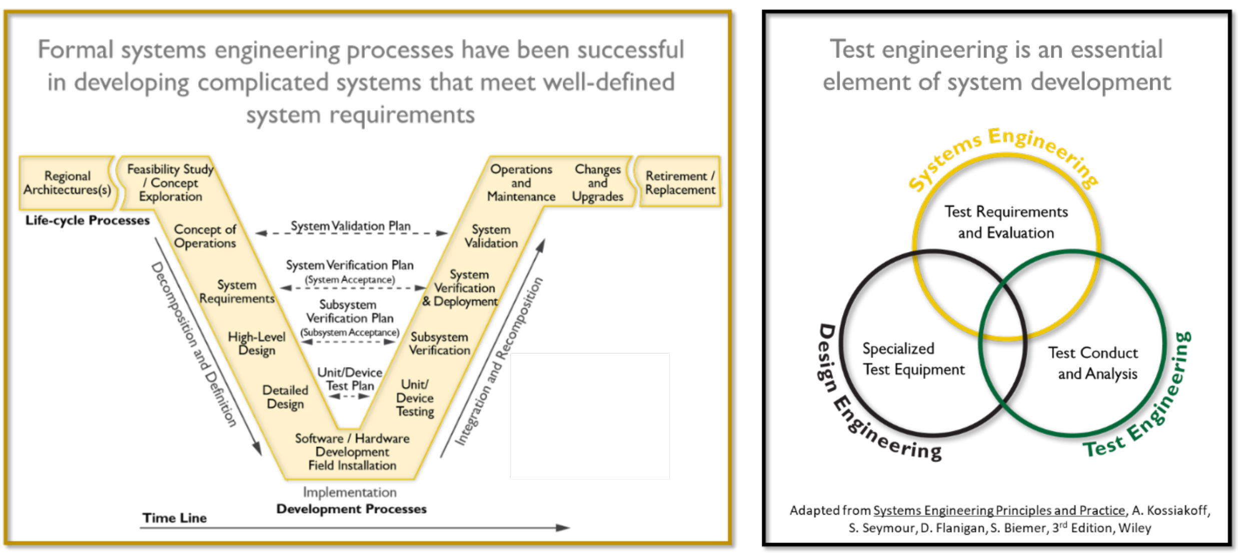


Figure 1. Test & Evaluation are Essential Elements of System Development

[Adapted from Systems Engineering Principles and Practices, A. Kossiakoff, S. Seymour, D. Flannigan, and S. Biemer, 3rd Edition Wiley]

Figure 1 also describes T&E from the differing perspectives of the systems engineer, design engineer, and test engineer. Setting the overarching T&E requirements for the system is a role for the systems engineer, who must work with both the design engineer and the test engineer in establishing realistic and achievable T&E requirements. The design engineer must produce a design that meets the system objectives while also enabling its testability, which often results in the need for development of specialized test equipment. Finally, the test engineer must conduct the tests and analyses to verify that the system and its components both meet their design specifications and are suitable for their intended use. Through the intersecting activities of system engineers, design engineers, and test engineers, the capabilities of a new system capability are documented and demonstrated, providing confidence in these capabilities.

2.2 DoD Test & Evaluation

Within DoD, T&E is separated between developmental test and evaluation (DT&E) and operational test and evaluation (OT&E). DT&E is the process by which a system or components are compared against requirements and specifications through testing, which corresponds to the verification of the system capability and its components. As defined under Title 10, U.S. Code, Section 139, OT&E "means the field test, under realistic operational conditions, or any item of (or key components of) weapons, equipment, or munitions for the purpose of determining the effectiveness and suitability of the weapons, equipment, or munitions for use in combat by typical military users." OT&E effectively corresponds to the validation of the system.

The Task Force was asked to review the T&E oversight organizations within OSD. A simplified view of this T&E oversight is shown in in **Figure 2**. Today's T&E processes are built around new capability acquisitions, with the capability development occurring through Service Acquisition Programs under the oversight of the Service Acquisition Executives and the Under Secretary of Defense for Acquisition and Sustainment (USD(A&S)).

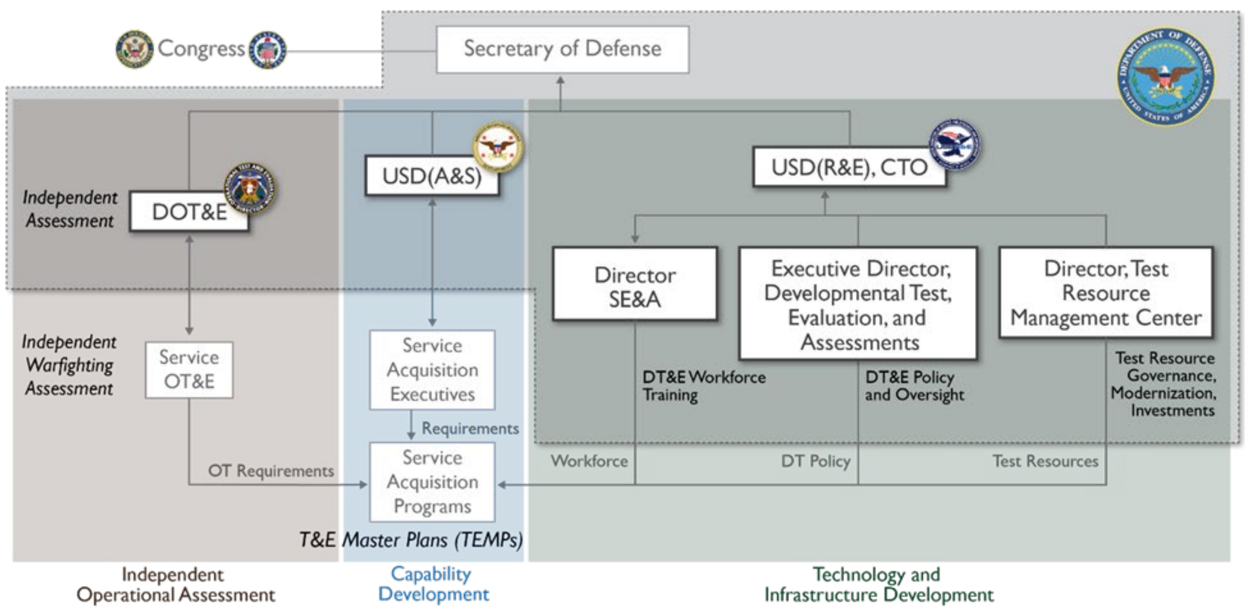


Figure 2. A Simplified View of T&E Oversight within DoD

Under Title 10, Section 133(b), USD(A&S):

> "...serves as the chief acquisition and sustainment officer of the Department of Defense with the mission of delivering and sustaining timely, cost-effective capabilities for the armed forces (and the Department); and establishing policies on, and supervising, all elements of the Department related to acquisition (including system design, development, production, and procurement of

> goods and services) and sustainment (including logistics, maintenance, and materiel readiness)."

The function of verifying the system requirements falls to Military Services test centers working closely with the Service Acquisition Programs. Under Title 10, Section 233(a), the USD(R&E) is responsible for:

> "…establishing policies on, and supervising, all defense research and engineering, technology development, technology transition, appropriate prototyping activities, experimentation, and developmental testing activities and programs and unifying defense research and engineering efforts across the Department."

Oversight of developmental testing (DT) is provided by the Executive Director for Developmental Test, Evaluation, and Assessments (ED,DTE&A) operating within the OUSD(R&E). The DTE&A mission is to provide systems engineering and T&E rigor to the DoD AAF program development pathways and USD(R&E) modernization priorities to ensure delivery of relevant and timely warfighting capabilities.

Oversight of two additional T&E elements operate under OUSD(R&E): the Test Resource Management Center (TRMC), and the Principal Deputy Executive Director for Systems Engineering and Architecture (PDED,SE&A). The Director of TRMC, operating under Title 10, Section 4173, is responsible for: reviewing and providing oversight of DoD budgets and expenditures for the T&E facilities and resources of the DoD Major Range and Test Facility Base (MRTFB); reviewing proposed significant changes to the T&E facilities and resources of the MRTFB with respect to expansion, divestment, consolidation, or curtailment of activities; completing and maintaining a quadrennial strategic plan regarding T&E facilities, and; administering the Central Test and Evaluation Investment Program (CTEIP) and the DoD Test and Evaluation Science and Technology (TEST) program. The PDED,SE&A, leads the policy, guidance, and workforce development for DoD engineering and technical workforce, to include test engineering.

The function of validating new system capabilities is executed through independent OT&E functions operated by each Military Service under the guidance and oversight of the Director of OT&E (DOT&E). The authorities of DOT&E, which are defined under Title 10, Section 139, include prescribing policy and procedures for the conduct of DoD OT&E, monitoring and reviewing all OT&E within DoD, coordinating operational testing (OT) conducted jointly by more than one military department or defense agency, and making recommendations to the Secretary of Defense (SecDef) on all budgetary and financial matters related to OT&E. DOT&E also executes the Live Fire Test and Evaluation (LFT&E) program, which may combine both DT&E and OT&E to assess the vulnerability and/or lethality of a system before it is approved for full-rate production. DOT&E is charged with developing "an annual report summarizing operation test and evaluation activities (including live fire testing activities) of the Department of Defense during the preceding fiscal year."

DOT&E and USD(R&E) jointly publish the OSD T&E Oversight List, which identifies programs designated for DT oversight or engagement, and OT and LFT&E oversight. In 2023, approximately 270 acquisition programs were under OSD oversight.[3]

To assist in understanding OSD T&E oversight, the Task Force explored the history that led to today's organizational structure. **Figure** 3 shows an abbreviated history for the OSD oversight of developmental testing dating to the formalization of T&E oversight within the Department beginning in approximately 1972. At that time, T&E fell under the Deputy Director of Defense Research and Engineering (DD,DR&E) with responsibility for policy oversight and resources. This organization remained in place until approximately 1983, at which time DOT&E was established and DT oversight was provided by the Director of Test and Evaluation (D,T&E) under the Under Secretary of Defense for Acquisition (USD(A)) with responsibility for DT policy and oversight, ranges, and resources.

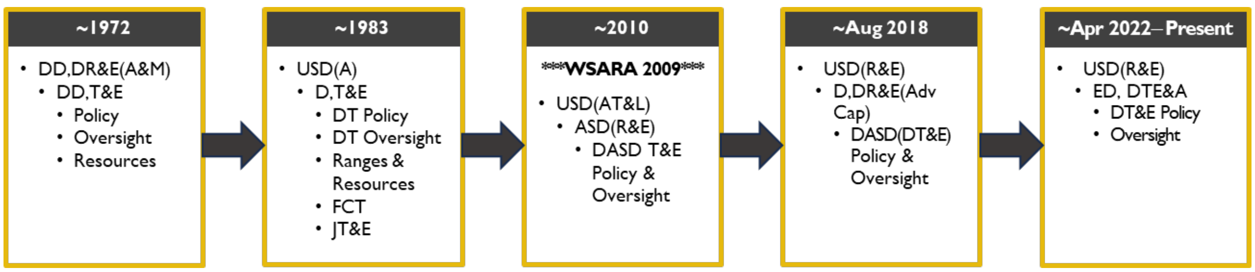


Figure 3. Abbreviated History of OSD Oversight of Developmental Testing

In 2009, the Weapon System Acquisition Reform Act (WSARA) led to the creation of the position of Deputy Assistant Secretary of Defense for T&E under the Under Secretary of Defense for Acquisition, Technology, and Logistics (USD(AT&L)) with statutory responsibility for DT policy and oversight. Following congressional desire to streamline OSD functions, USD(AT&L) was split into USD(A&S) and USD(R&E) in 2018, with DT&E statutory authority withdrawn and DT&E policy and oversight functions falling under the Director of Defense Research & Engineering for Advanced Capabilities (D,DR&E(Adv Cap)) within the OUSD(R&E). Finally, in April 2022, the role of DT policy and oversight was elevated to a direct report to USD(R&E) as the Executive Director, which is the organizational structure in place today.

A brief history of the OSD OT&E oversight organization in shown in **Figure 4**. The office of the Director of Operational Test and Evaluation (DOT&E) reporting to the SecDef was established in 1983 to provide OT policy and oversight and to provide an independent assessment of the readiness of new acquisition systems for production and transition to warfighters.

[3] "OSD T&E Working Oversight List," The Office of the Director, Operational Test and Evaluation, Dec 20, 2023, https://www.dote.osd.mil/Portals/97/pub/reports/oversight/osd%20te%20oversight%20current%2020231215%20www%20extract.pdf?ver=4gUVosQDP9IoINYSqt-URQ%3d%3d.

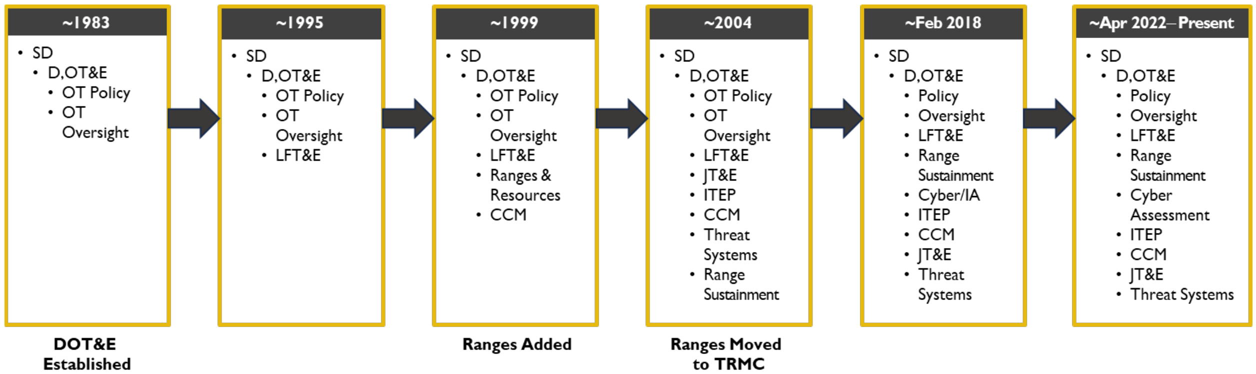


Figure 4. Brief History of OSD Oversight of Operational Test and Evaluation

In approximately 1995, LFT&E was added to DOT&E responsibilities in response to concerns about the survivability and/or lethality of conventional weapons or weapon systems during their development. Realistic survivability testing often requires firing actual or representative weapon munitions at the system under development configured for combat. Realistic lethality testing requires firing a weapon configured for combat at appropriate targets. Survivability and lethality testing for major systems and munitions programs remain legislated requirements today under Title 10, Section 4172.

A series of changes to DOT&E has occurred over the past 25 years. Ranges and resources were added to DOT&E responsibilities in 1999 and were removed in 2004 following the establishment of the TRMC. The Center for Countermeasures (CCM) was established 1999 to focus attention on adversary advancements aimed at defeating U.S. weapons and weapon systems. In 2004, Joint Test & Evaluation (JT&E) was added to focus attention on joint tactics, techniques, and procedures (TTPs) to close identified operational gaps. Further expansions in DOT&E responsibilities included the International Test & Evaluation Program (ITEP) and a formal role in assessment of cyber vulnerabilities.

A brief history of the TRMC is shown in **Figure 5**. First established in approximately 2004, TRMC was assigned oversight responsibilities for: 1) the MRTFB, which is the designated core of DoD T&E infrastructure, 2) the CTEIP; 3) the T&E science and technology (S&T) program; and 4) creating the DoD strategic plan regarding T&E resources. Organizationally, the Director, Test Resource Management Center (D,TRMC) initially reported to USD(AT&L). The responsibilities of TRMC have largely remained the same since 2004, although reporting responsibilities have shifted, most significantly in 2018 when TRMC transitioned into the OUSD(R&E) following the split of OUSD(AT&L). In April 2022, TRMC was elevated to a direct report to USD(R&E), where it remains as of this report.

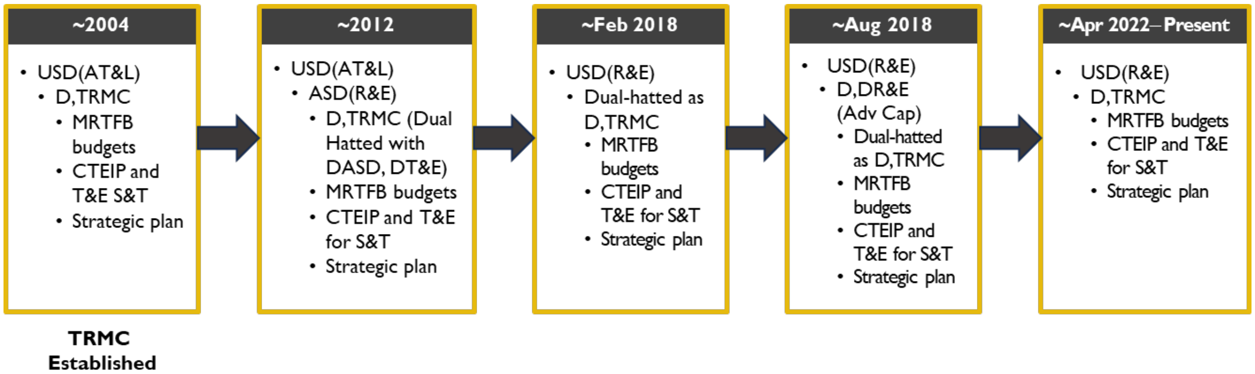


Figure 5. Brief History of the Test Resource Management Center

As seen in these abbreviated summaries (Figures 3–5) of the history of DoD T&E oversight, significant variations in T&E roles and responsibilities have occurred over time in the search for increasing efficiency and effectiveness. As called for in the ToR, the Task Force was asked to again explore opportunities for increasing efficiency and effectiveness of these OSD T&E oversight organizations.

2.3 Industry Best Practices

The Task Force was asked to consider industry T&E best practices and to determine their applicability for testing within DoD to improve test and evaluation outcomes. Essential to the success of many commercial products is the concept of continuous development, as illustrated in **Figure 6**, for a system built upon both hardware and software development. This development model is used in many different applications ranging from aircraft to elevators to traffic management. With software, the capability to develop, test, and deploy upgraded capabilities occurs on a much faster timescale than hardware changes. This approach, in which rapid software development occurs inside the timescale of hardware (or block) upgrades, is applicable to DoD system acquisition and will be addressed later in this report.

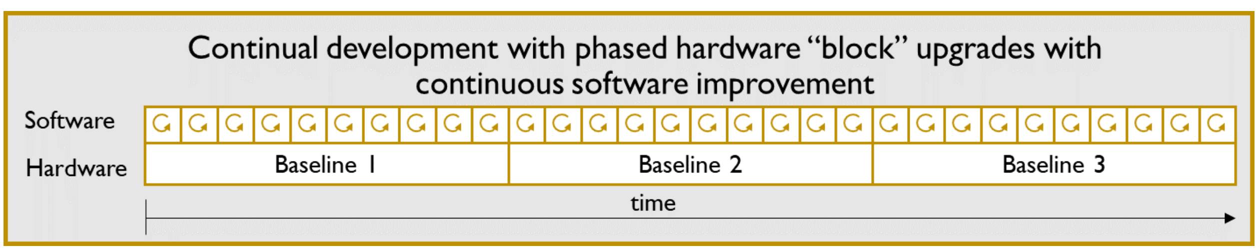


Figure 6. Industry Approach to Continual Development

Robert G. Cooper published a review of drivers of success in new product development in 2018.[4] In it, he identifies spiral or iterative development processes as the way fast-paced project teams handle dynamic environments where establishing a stable and rigid product definition is not possible during early stages of development. Using an iterative approach of "build – test – feedback – revise"

---

[4] "The Drivers of Success in New-Product Development," Cooper, R.G., Industrial Marketing Management, 76 (2019) Elsevier Inc. https://doi.org/10.1016/j.indmarman.2018.07.005.

promotes experimentation within project teams who are also encouraged to fail fast and fail cheaply. This continuous experimentation ultimately reduces technical uncertainties.

Additional industry best practices are identified in **Figure 7**. These include incorporation of built-in diagnostics and appropriately sizing infrastructure for collection and analysis of system tests. These built-in diagnostics are used in systems such as phones, computers, and cars, which enables continuous monitoring of the system and automatic reporting of system status and test results. The development of appropriate infrastructure is used to collect, monitor, and evaluate the system.

Figure 7. Key Elements of Industry Approach to Testing and Evaluation

Industry's willingness to test early in development, accept risk, and tolerate failure results in an agile development approach that often runs counter to the more structured systems engineering process described above. Advocates for this more agile approach argue that development occurs much faster with the accelerated learning made possible by early testing.

Several additional elements of industry best practices were also identified. The first is industry's continual monitoring of the competitive environment in which its systems operate. Given the efficiency of the commercial marketplace, organizations must maintain constant awareness of the environment and the usefulness of their product in that environment. This awareness feeds the continuous process of product improvement needed to remain relevant in a competitive environment.

Lastly, the development environments for many commercial systems are merging the physical and digital worlds and have found significant utility in the use of digital twins, as illustrated in **Figure 8**. The Digital Twin Consortium[5] defines a digital twin as "a virtual representation of real-world entities and processes, synchronized at a specified frequency and fidelity." A digital twin can transform business by providing a more complete understanding of an integrated system,

Figure 8. Oil Rig and its Digital Twin

[Source: SumitAwinbash, CC BY-SA4.0]

[5] "Definition of a Digital Twin," Digital Twin Consortium, November 13, 2023, accessed January 3, 2024, https://www.digitaltwinconsortium.org/initiatives/the-definition-of-a-digital-twin/.

which in turn enables improved decision making. Effective digital twins operate on real-time and historical data to represent the past, present, and simulated futures. They also provide an integrated basis for simulation, maintenance, integration, testing, and system monitoring, and often serve as the basis for lifecycle management of a product.

Commercial industry uses digital twins in a variety of circumstances, including prototype development, assessing supply chains and logistics, and modeling of retail customers. The global market for digital twins is projected to grow with a 38% compound annual growth rate (CAGR) to reach $92B-$155B by 2023.[6] Despite this, only 7% of acquisition programs under DOT&E oversight are developing a digital twin.[7]

---

[6] Adam Mussomeli et al., "Signals for Strategists: Expecting Digital Twins," *Deloitte Insights*, accessed January 1, 2024, https://www2.deloitte.com/content/dam/insights/us/articles/3773_Expecting-digital-twins/DI_Expecting-digital-twins.pdf.

[7] "FY 2022 Annual Report," Director, Operational Test and Evaluation, January 2023.

3. Challenges

In addressing the need for increasing the efficiency and speed of T&E activities, a wide range of challenges and opportunities for improvement were identified. In the following subsections, these are discussed under the broad categories of acquisition-based T&E, emerging new technologies, and mission-focused T&E.

3.1 Acquisition-Based T&E

Test and evaluation has historically been tied to the acquisition of new systems and capabilities, with T&E planning starting at program initiation and ending after proving that the system in production is satisfying its operational intent. Because the AAF aims to provide flexible acquisition pathways to accelerate development, the need has arisen to re-examine current approaches to T&E and assess opportunities for increasing efficiencies.

DoD Instruction (DoDI) 5000.02 provides the implementation guidance for the AAF.[8] An illustration of this framework is provided in **Figure 9**. The Task Force study focused on the first four pathways: Urgent Capability Acquisition, Middle Tier of Acquisition, Major Capability Acquisition, and Software Acquisition. The two remaining pathways associated with Defense Business Systems and Acquisition of Services were not addressed.

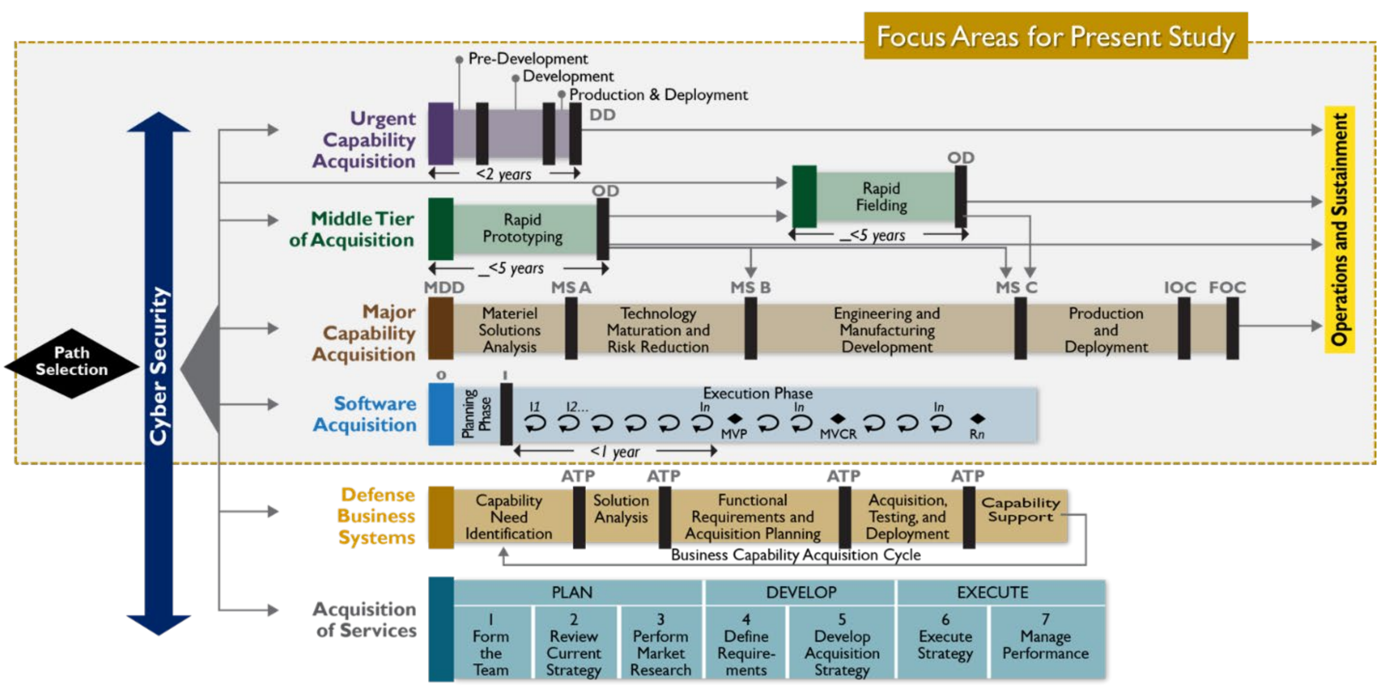


Figure 9. Adaptive Acquisition Framework Designed for Flexibility

[8] DoDI 5000.52: *Operation of the Adaptive Acquisition Framework*, Office of the Under Secretary for Acquisition and Sustainment, June 8, 2022.

3.1.1 T&E Challenges and Opportunities Regarding Major Capability Acquisitions

The Task Force evaluation of T&E challenges and opportunities started with the Major Capability Acquisition (MCA) pathway, which is used for Major Defense Acquisition Programs (MDAPs). An MDAP is a program that meets or exceeds the Acquisition Category (ACAT) I requirements, defined under DODI 5000.85[9] as requiring more than $525M in FY20 dollars for research, development, and test and evaluation (RDT&E), or more than $3.065B (FY20 dollars) for procurement. The MCA pathway is also used for ACAT II programs, which have RDT&E requirements greater than $200M or procurement requirements greater than $920M, and automated information systems not managed by other acquisition pathways.

The MCA pathway is illustrated in **Figure 10** together with several challenges and opportunities for speeding capability development. The MCA pathway begins with a Materiel Development Decision (MDD) leading to analysis of possible materiel solutions. Following an evaluation of materiel solutions, the Milestone A (MS A) decision leads to a Technology Maturation and Risk Reduction phase. A Milestone B (MS B) decision moves to formal program establishment as the program transitions into the Engineering and Manufacturing Development (EMD) phase prior to the Milestone C (MS C) decision to move to production and deployment. Initial operational capability (IOC) and full operational capability (FOC) are satisfied using production units and end-use operators.

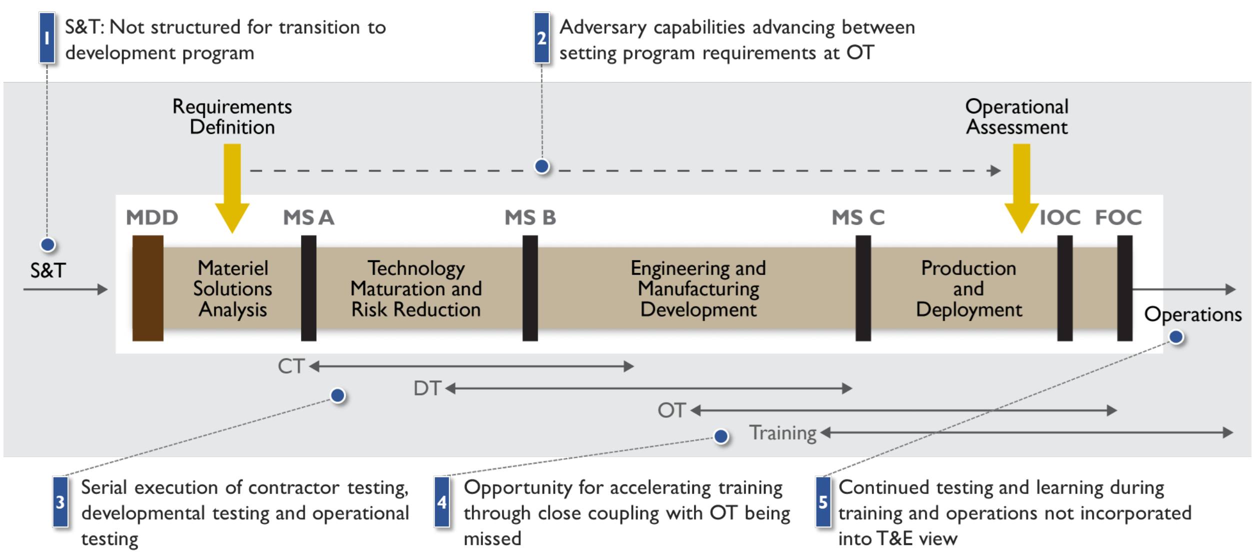


Figure 10. Challenges and opportunities for increasing efficiency in Major Capability Acquisitions.

T&E occurs throughout the MCA pathway as essential elements of the decision milestones. MS A requires definition of proposed test strategy, which is expressed using a Test and Evaluation Master Plan (TEMP) that defines approaches to DT and OT needs. Significant DT&E occurs during EMD to provide feedback on the progress of the design and compliance with contractual requirements, effective combat capability, and the ability to achieve key performance parameters (KPPs) and key system attributes (KSAs). OT organizations conduct independent evaluations, initial assessments of operational effectiveness, suitability, and survivability, and the ability to achieve KPPs and KSAs.

[9] DoDI 5000.85: *Major Capability Acquisition*, 2021.

During this phase, programs are instructed to maximize opportunities to combine contractor testing (CT) and government DT and conduct integrated OT when feasible.

The EMD phase ends when the system capabilities have been demonstrated by developmental, live fire (if appropriate), and early operational testing. The results of DT&E and early OT&E are important inputs to the MS C decision to move to production. During this phase, DT&E is completed if not already completed, and initial operational test and evaluation (IOT&E) is conducted using production or production-representative units and end-use operators.

In evaluating the MCA pathway, the Task Force identified five challenges and opportunities for improving T&E efficiencies. Several of these opportunities involve expanding the conventional definition of T&E such that it begins before MS A and extends past declaration of FOC.

S&T: Not structured for transition to development program.
Many S&T activities are not structured to support seamless transition to acquisition programs, resulting in inefficiencies in beginning phases of the acquisition process. An Independent Technology Risk Assessment (ITRA)[10] is conducted during the Materiel Solutions Analysis (MSA) phase with an update required for the MS B decision. The ITRA often requires significant compilation of otherwise existing knowledge and feeds further technology development and testing requirements as the program enters the Technology Maturation and Risk Reduction phase.

The S&T community widely uses the Technology Readiness Level (TRL) scale to gauge readiness for transition of a technology into an acquisition program, with the goal of reaching TRL 6 via a "system/subsystem model or prototype demonstration in a relevant environment." The USD(R&E) guidance for achieving TRL 6 calls for a representative model or prototype system (developed well beyond that of TRL 5) that has been tested in a relevant environment. Achieving TRL 6 is intended to represent a major step up in a technology's demonstrated readiness. Examples include testing a prototype in a high-fidelity laboratory environment or in a simulated operational environment. Supporting information includes results from laboratory testing of a prototype system that is near the desired configuration in terms of performance, weight, and volume. Questions to be addressed in assessing whether TRL 6 has been achieved include: How did the test environment differ from the operational environment? Who performed the tests? How did the test compare with expectations? What problems, if any, were encountered? What are/were the plans, options, or actions to resolve problems before moving to the next level?

Recognizing that significant testing and knowledge accrual occurs during the S&T phase of development, opportunities exist to speed the efficiency of starting a program acquisition, including the system T&E, by tailoring S&T programs to aim for a seamless and structured transition to the acquisition program. Typical knowledge products that should be developed and prepared for transiting to formal acquisition programs should include data repositories, validated modeling and simulation (M&S) tools, description and results from test approaches and lessons learned, and documented assumptions and limitations of technologies, subsystems, and/or systems. With this knowledge package produced and transitioned for acquisition, the opportunity exists for the ITRA

[10] "Technology Readiness Assessment Guidebook," Office of the Executive Director for Systems Engineering and Architecture Office of the Under Secretary of Defense for Research and Engineering, June 2023.

required at the beginning of an acquisition program to be more readily accomplished and the efficient transition from technology to acquisition to be realized.

Adversary capabilities advancing between setting program requirements and initiation of OT.
Top-level program requirements are explored during the MSA phase and adopted at Milestone A. These Top-Level Requirements are decomposed to lower system and subsystem requirements leading to a System Requirement Review. During these early phases of the program, requirements are established based on known and projected understanding of the operational environment, which incorporates adversary capabilities and needed interfaces to other systems and capabilities. These requirements then drive planned system development, including CT and DT activities.

While aspects of operational testing occur throughout a program, the final OT assessments occur much later in the program, sometimes years later. During the period between setting program requirements and conducting operational testing, the external environment often changes and a disconnect emerges between the system needs for its operational assessment and the capabilities to which the system has been designed. This disconnect between DT, which verifies that the system meets its requirements, and OT, which validates that the system meets its intended capability in an operational environment, can significantly impact program delivery costs and schedules.

Without integrated testing, serial execution of contractor testing, developmental testing, and operational testing.
Contractor testing, developmental testing, and operational testing occur sequentially during the MCA process. With the need for independent testing in the DT and OT processes, the serial nature of test activities can cause inefficient hand-offs of test results and knowledge products between CT, DT, and OT. Significant effort has been undertaken to "shift left," bringing DT and OT activities to earlier stages of the program. This is done with the goal of maximizing "integrated testing," where a single test event can satisfy aspects of CT, DT, and OT.

MCAs require adoption of a TEMP at Milestone A. The goal of the TEMP is to synchronize the CT, DT, and OT activities and maximize the opportunities for integrated testing. This goal is captured in the mantra "test once, use the data many times."

When integrated testing is not fully realized, inefficiencies in T&E slow the acquisition process. Typical reasons for not fully realizing the goals of integrated testing include government test teams not being prepared for (or sufficiently integrated into) test activities, and unforeseen schedule changes with test facilities, ranges, and operational users.

Opportunities for accelerating training through close coupling with OT being missed.
The 2023 DOT&E annual report, *DOT&E 2023,* identified training as an issue with a number of programs. Since development of training systems proceeds in parallel with the capability development, the opportunity exists to couple conventional OT of the system capability being acquired with needed training capabilities. To the extent that OT and the training system proceed independently of each other, the Department is missing an opportunity to accelerate the transition of the capability to operational forces.

Opportunities for continued testing and learning during training and operations often not incorporated into T&E view.

Conventional T&E ends with declaration of FOC and the operational testing to validate that the system meets its intended function using actual production systems. In reality, with the deployment of the system into training and operational environments, knowledge concerning the operation and performance of a system continues to accrue. When this knowledge is not formally captured and linked to the testing that occurred in earlier phases of the capability development, the opportunity to efficiently improve understanding of the system and gain insight into needed improvements is missed.

3.2 Accelerated Acquisition Pathways

Alternatives to the MCA pathway are available within the AAF to enable acceleration in acquisition of new capabilities. Through exploration of examples of activities using these alternative pathways, the Task Force identified a number of T&E challenges being encountered in the use of these pathways. **Figure 11** provides a summary of the challenges observed in the pathways using Urgent Capability Acquisition (UCA), Middle Tier of Acquisition (MTA), and Software Acquisition.

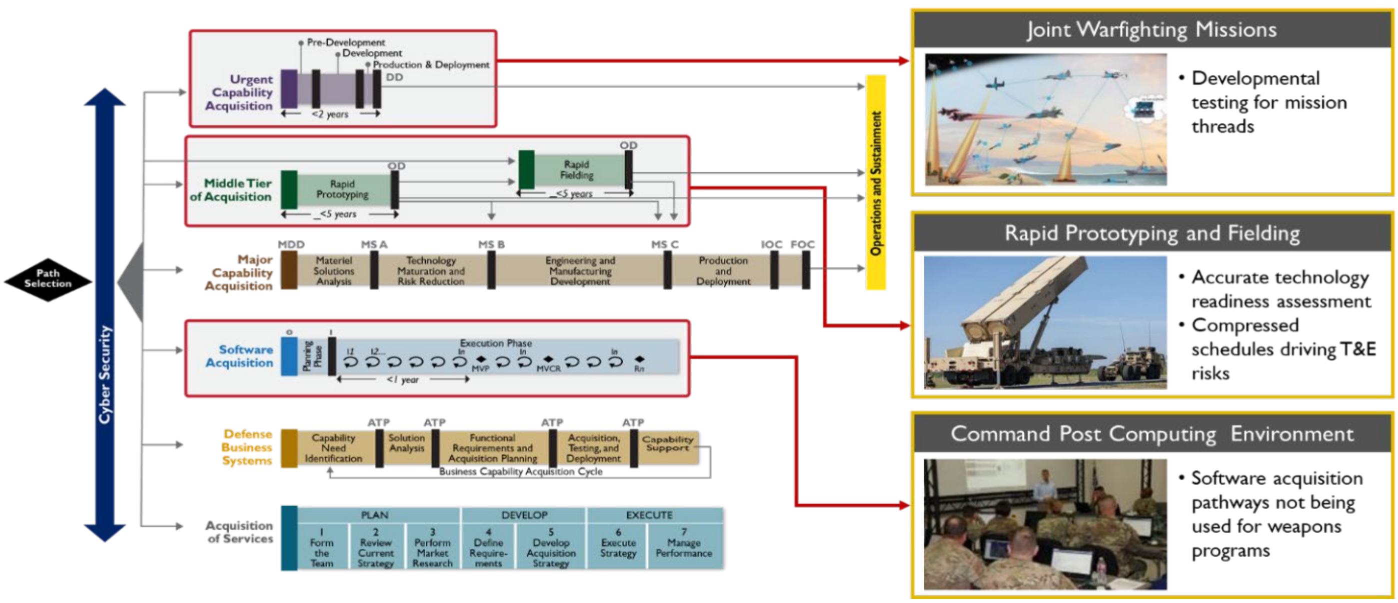


Figure 11. T&E Challenges for Adaptive Acquisition Framework Pathways Working to Accelerate Acquisition of New Capabilities

3.2.1 Urgent Capability Acquisition Pathway

The UCA pathway is available for rapid acquisition of new capabilities meeting a pressing operational need. The acquisition program duration is limited to two years, which results in significant compression of the system development and production timeline, including T&E. A direct transition to operations and sustainment is called for at the completion of the UCA.

Three types of urgent needs are defined in the Chairman of the Joint Chiefs of Staff (CJCS) Instruction 5123.01, *Charter of the Joint Requirements Oversight Council and Implementation of the Joint Capabilities Integration and Development System*:

- Joint Urgent Operational Need (JUON);
- Joint Emergent Operational Need (JEON); and
- Urgent Operational Need (UON).

Given the highly compressed development timeline of the UCA pathway, challenges involving potential T&E inefficiencies are amplified. UCA acquisitions are ideally suited to execution by relatively small, focused teams given authorities to make rapid progress. Note that the objectives of DT&E and OT&E are maintained, although processes to meet those objectives are usually tailored to enhance progress and development speed.

In reviewing activities using the UCA pathway, the Task Force observed T&E challenges associated with the development of advanced joint mission threads that are integrating both existing and new capabilities. This type of system-of-systems mission capability, which is being requested by the Joint Force to quickly address operational needs, often crosses Military Services, and may result in acquiring systems that themselves may be under development. As such, historic approaches to T&E, which are designed to be executed within a single acquisition activity, break down.

In addition to crossing Military Service boundaries, experimental and analytical efforts that do not consider the full complexity of envisioned operational requirements, as would typically be considered in a longer-term MCA activity, are used in the UCA pathway for rapid development of a system-of-systems capability. Thus, a complete set of system requirements may not be defined, which leads to shortfalls in the DT verification, which may impact the robustness of resultant systems. This challenge is often uncovered as OT&E processes work to validate the full utility of the needed system-of-systems capability.

3.2.2 Middle Tier of Acquisition Pathway

The MTA pathway is available for accelerating acquisition of new capabilities through rapid prototyping and fielding activities, each of a duration less than five years.[11] Significant acquisition flexibility is available using this pathway; MTA rapid prototyping activities may move to an MCA at Milestone B or C, into a Rapid Fielding activity, or transition directly to operations and sustainment. Similarly, the rapid fielding activity can transition to an MCA at Milestone C or directly transition to operations and sustainment. Examples of programs being executed using the MTA pathway include Conventional Prompt Strike (CPS), Air-launched Rapid Response Weapon (ARRW), B-52 Commercial Engine Replacement, F-22A Modernization, Long-Range Assault Aircraft, and 3D Expeditionary Long-Range Radar.

In reviewing programs using the MTA pathways, a number of observed T&E challenges were tied to the compressed timelines, especially in the accuracy of the initial technology readiness assessments (TRAs), which led to challenges discovered in later testing. The Task Force also noted that the compressed schedules are driving significant T&E risks. As shown in Figure 1 earlier in this report, a structured systems engineering process uses a decomposition of system requirements down to subsystem and component levels, with a structured DT&E process verifying that the component, subsystem, and system requirements are being met. With the compressed timelines of the MTA

[11] DoDI 5000.80: *Operation of Middle Tier of Acquisition (MTA)*, December 30, 2019.

activities, significant T&E risks are being taken, leading to program delays as problems are uncovered during subsequent integration tests.

3.2.3 Software Acquisition Pathway

The Software Acquisition Pathway has been established in recognition that software development is not readily adaptable to conventional milestone-based acquisition.

DoDI 5000.87[12] provides guidance for integrating modern software development practices, such as agile development, security, and operations (DevSecOps), by specifying guidance that:

- software development should not be treated as MDAPs even if MDAP size thresholds are exceeded;
- the viability and effectiveness of capabilities for operational use must be demonstrated within one year after obligation of funds;
- embedded system software must be delivered to operationally representative environments at least annually;
- modern, iterative software development methodology, modern tools (e.g., DevSecOps), and human-centric design processes are to be used to iteratively deliver software;
- software must be instrumented for monitoring of critical functions related to health, security, and operational effectiveness; and
- software development must be conducted collaboratively with end users and undergo regular assessments of software performance and risks.

DoDI 5000.87 also addresses testing strategies, to include:

- use of automated testing and operational monitoring of the software to the maximum extent possible; and
- providing continuous runtime monitoring of operational software to provide health-related monitoring (e.g., performance and security) and data collection to support test and continuous operational testing.

Note that for programs on the DOT&E oversight list, DoDI 5000.87 provides guidance, aligned with acquisition authorities, that DOT&E will:

- approve the adequacy of test strategies and test plans for OT&E in connection with the program, and
- assess test results of OT&E and whether tests confirm the program is effective, suitable, and survivable for operational use.

---

[12] DoDI 5000.87: *Operation of the Software Acquisition Pathway*, Under Secretary of Defense for Research and Engineering, 2020.

In reviewing the Software Acquisition Pathway, the Task Force investigated the software development being conducted under Kessel Run[13] for the United States Air Force (USAF) Air Operations Center (**Figure 12**), which is a program operating under DOT&E oversight. Kessel Run is executing non-safety critical development in a continuous development environment where they are, on average, achieving program deliveries of once per week, and, across all applications, averaging deliveries every 3.3 hours. In addition, the period between committing to code development and running in production is approximately 8 hours.

Observations. Kessel Run is executing with DT/OT personnel embedded and "shifted left," which has evolved to a continuous development model. These independent DT/OT functions operate in parallel with automated and team testing. The Kessel Run program notes that mismatches often occur regarding the definition of "done" with respect to testing. The program evaluates itself against meeting priority requirements for the users, and risks are routinely accepted in pushing new software to operations without formally completing DT/OT functions. In these cases, software bugs identified by operational users are merged with needs identified by testers and are fed back into the continuous development process. The Kessel Run program also notes OT often evaluates progress against legacy requirements, which typically do not match the continually changing needs of the user community.

Figure 12. 612th Air and Space Operations Center at Davis-Monthan Air Force Base, Arizona

An essential feature of the Kessel Run environment that serves to enable this continuous development is a continuous authority-to-operate (cATO). By certifying the environment, rather than individual applications, Kessel Run is provided the flexibility to ensure capabilities can be released as needed.

The definition of "done" is one challenge widely observed in agile development, especially in highly regulated environments, and many studies have been conducted on agile development in safety critical systems. Poth et. al. have proposed a systematic approach to scaling agile development within a regulated environment.[14] Introducing the concept of "levels of done" as an independent consideration in assessing product quality risk allows the development of an approach to compliance

---

[13] "AOC ("Kessel Run") Case Study," briefing provided to the T&E Task Force, Section 231 Team, dated 10 December 2021.

[14] "Systematic Agile Development in Regulated Environments," Poth, A., Jacobsen, J. and Riel, A., 27th European Conference on Systems, Software, and Service Improvement EuroSPI 2020, Yilmaz, M. et. al. (Eds) 2020. https://doi.org.10.1007/978-3-030-56441-4_14.

verification that accounts for risk, level of team accountability, and team maturity. At the present time, DoD uses a single set of compliance regulations regarding acquisition of weapon systems.

3.3 Emerging New Technologies

The next major challenge area for test and evaluation is associated with emerging new technologies, especially when dealing with technologies that represent significant leaps forward in producing new capabilities. These technologies often require development of new analytical tools and testing capabilities for verifying and validating system performance. Further, physical testing may not be possible in some cases, so V&V of new analytical and computational tools become essential to filling gaps in physical testability.

To explore the issues with T&E of emerging new technologies, the Task Force examined hypersonics, directed energy, and artificial intelligence/machine learning (AI/ML), as illustrated in **Figure 13,** as examples presenting T&E challenges.

Figure 13. Emerging Technologies that Present Challenges to Today's T&E

3.3.1 Challenges of T&E of Hypersonic Systems

Development of hypersonic systems (greater than Mach 5) is an imperative to matching and challenging the rapidly increasing defensive and offensive capabilities of our adversaries.[15] These systems will be important in the future to provide a U.S. capability that can deter major adversary aggression and defeat them if required. Hypersonic weapons provide the speed and maneuverability required to survive emerging adversary defensive systems, and to provide a prompt response should our forces be called upon to engage adversary targets at range. Because of the tremendous speed these systems achieve, they must survive in harsh environments when operated within the sensible atmosphere. These environments dictate that unique high-temperature materials be employed in the system design, as well as unique control mechanisms, structures, sensors, and mission planning capabilities that can handle these advanced capabilities.

Observations. As a result of many overlapping challenges, hypersonic systems are complex and require strong systems engineering to achieve successful deployment. An important and needed element in this success is a thoughtful test and evaluation activity to ensure system development progresses successfully through the various technical challenges and potential solutions while also efficiently demonstrating the end capability. Replicating the harsh environments of a real hypersonic

[15] "Hypersonic Weapons: Background and Issues for Congress," Congressional Research Service, updated February 9, 2024.

flight cannot be fully accomplished in ground-test facilities. Flight tests of these systems are expensive, and, because of the potential range and cross range capability, they require extensive test infrastructure to ensure safety and data collection requirements are met. Therefore, subscale, component, and subsystem testing are important, and unique facilities are required to aid in system development.

Despite the investments made in ground test infrastructure, complete representation of the hypersonic environment is usually not achieved in ground tests. A basic understanding of the reason behind this limitation is provided in **Figure 14**, where the air supply temperature needed for the ground test is shown as a function of flight velocity and altitude. For example, flight at Mach 15 and 150,000 ft. altitude requires a supply temperature of approximately 25,000°R. When faced with an inability to produce these conditions in ground-test facilities (except in extremely short-duration pulse facilities), developers resort to partial simulations in ground tests complemented by detailed analyses and numerical modeling prior to proceeding to flight tests.

In addition to the investments made in ground test infrastructure, significant investment has been made to provide test launch capabilities that can support a range of subsystem test payloads and experiments and replicate flight environments to support hypersonic system development (see **Figure 15**). The MACH-TB program, funded through USD(R&E) and TRMC, is already providing initial capability in this area. Capabilities such as autonomous flight termination subsystems are being developed and tested, which will lead to increased test efficiency and reduction of test system complexity.

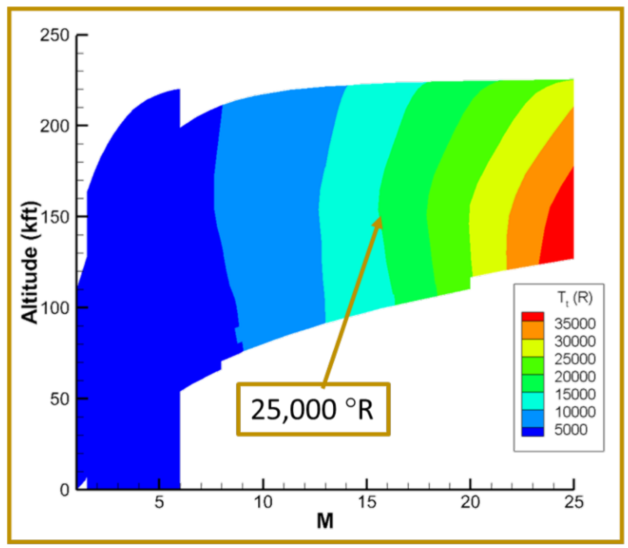


Figure 14. Total Temperature Required for Ground Test Duplication in Testing of Hypersonic Systems

Recent hypersonic development programs are trying to move fast and achieve deployed systems quickly, but these efforts have suffered from failures because of compressed test plans and strategies. The lack of structured subcomponent/subsystem tests and "test-to-learn" philosophies that buy down technical risk gradually have been replaced by complete all-up-round, full-scale tests of fully assembled systems that have mostly failed.

The Task Force notes that common test asset needs exist between legacy strategic systems and space systems test capabilities, which should be managed and leveraged across the hypersonic test regimes as well. An annual assessment of T&E capacity and assets for existing and emerging programs should be

Figure 15. Hypersonic Flight-Testing Improvements Include SkyRange and Autonomous Missile Self-Destruct System

completed to promote and manage efficient use of available test capacity.

Several observations can be made regarding the T&E challenges of hypersonic system under development. First, the concurrent design and testing of the technology under development presents very high risk to the programs. The incorporation of new materials, new flight regimes, and challenging new vehicle control systems are resulting in concurrent and interrelated risks that are proving to be points of failure in many development programs.

When combined with the inability to duplicate the flight environment, tremendous risks are being brought forward into flight tests. Overall, integrated flight tests are proving to be prone to initial failures due to system complexity and extreme environments.

Needed Best Practices. Several best practices derived from recent hypersonic system development are evident, as follows.

- Development programs should structure T&E events and testing evolution as part of an integrated development and advocate for continuous testing and “test to learn” philosophies.
- Hypersonic programs should explore production-like disciplines as early as possible to maximize test efficiency and minimize total test requirements.
- An annual assessment of T&E capacity for existing and emerging programs should be produced with advocacy for efficient use of available test capacity.
- Common requirements and infrastructure should be evaluated and leveraged between hypersonic and strategic systems testing with a goal of sharing resources across the Military Services/communities rather than allowing stovepiped efforts.

3.3.2 Challenges with T&E of Directed-Energy Systems

Directed-energy technologies and systems are being pursued for their potential use as cost-effective weapons. Both high-power microwave (HPM) and high-energy laser (HEL) technologies are being explored for use as directed-energy weapons, but both introduce unique T&E complexities that will need to be overcome.

High-Power Microwaves. HPMs create damage by transmitting an electromagnetic pulse that couples into target electronics, inducing currents of sufficient magnitude to result in burnout of circuit elements. For example, induced electric fields can result in arcing across small circuit features that disrupt or cause damage to sensitive circuits. As the feature sizes of modern electronics decrease, the electromagnetic field strength required to cause damage also decreases.

HPM attack vectors can be divided into “front-door” and “back-door” approaches. HPM attacks can be further divided into narrow-band and wide-band attacks, referring to the frequency bandwidth of the propagated beam. Front-door attacks couple the propagated HPM beams into the target system through existing apertures that are tied to sensors and receivers. Since these apertures generally have filters that limit the noise received, front-door HPM systems tend to be narrow-band. Back-door attacks use projected HPM beams to couple into internal electronics through gaps in the target that allow electromagnetic radiation into the system interior. Back-door attacks can be either narrow- or wide-band.

Several T&E challenges are present in HPM weapons development, including uncertainties in the effective coupling of the transmitted HPM pulse into target electronics, required diagnostics and testing equipment, and uncertainties concerning the accuracy of analytical tools being used to predict performance. Operational considerations are further challenged with understanding collateral damage issues, which include the impacts of HPM weapons on one's own forces and their equipment.

One phenomenon that is particularly challenging concerns an accurate assessment of HPM system performance, which is sensitive to atmospheric breakdown phenomena that are altitude and threat dependent. As illustrated in **Figure 16**, a few different regimes exist concerning the air electrical breakdown in the presence of high-power microwave beams in terms of ambient electric field strength, E_0, and ambient pressure, measured in Torr (note that the physical phenomenon of air breakdown is more closely tied to molecular collision rates, but collision rates are in turn closely tied to pressure for small changes in ambient temperature.) The chart shows the altitude corresponding to the ambient pressure for the 1976 standard atmosphere. The chart also shows the propagating power (in MW/cm^2) corresponding to the propagated electric field (in kV/cm).

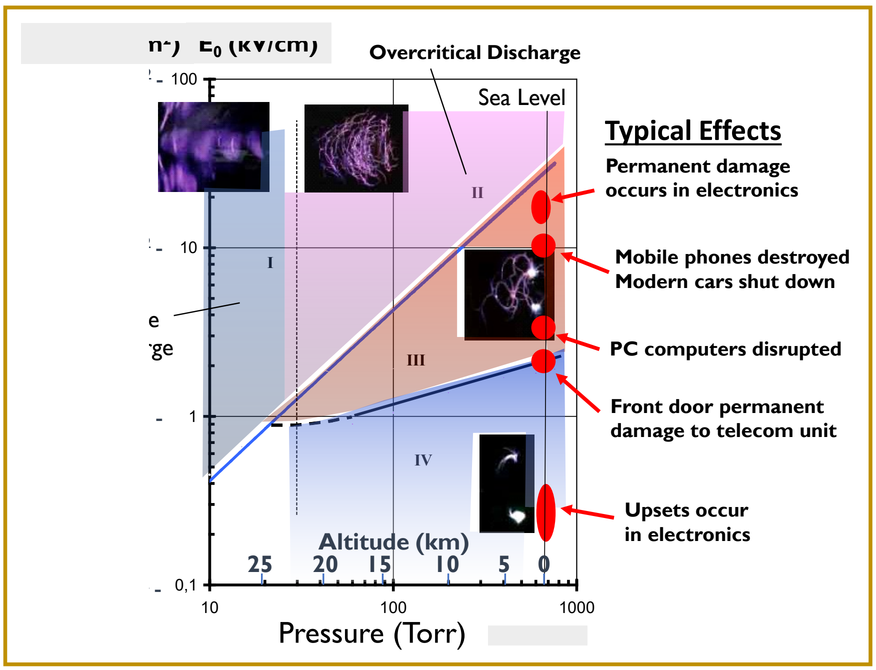


Figure 16. Microwave Breakdown in Air as a Function of Field Strength and Ambient Pressure

In **Figure 16**, the region labeled "Overcritical Discharges", designated Region II, refers to the range of conditions where electrical breakdown of air will occur at ambient conditions without a target present. The breakdown results in a filamentary structure that propagates towards the HPM source, creating the web-like structure. These filaments are highly effective resonant absorbers of the incident radiation. The electric field strength that produces air breakdown decreases with decreasing ambient pressure (i.e., increasing altitude). At high altitude (i.e., low pressures), designated Region I, diffusion of the gas away from the heated filaments produces a diffuse filamentary discharge.

At electric fields lower than that needed to create an over-critical discharge, electrical breakdown can still occur in regions of separating under-critical discharges (Region III) or non-separating under-critical discharges (Region IV). In both Regions III and IV, air breakdown occurs because of local electric field amplification due to the presence of the target. In Region III, the breakdown occurs with filaments separating from the originating source to propagate towards the source. In Region IV, the breakdown remains attached to the target feature that provides the local amplification of the ambient electric field. In both Regions III and IV, once breakdown occurs, the resulting filamentary discharge results in an efficient coupling of the transmitted HPM pulse into filamentary heating, which reduces the transmitted energy that can be coupled into the target electronics.

Bäckström and Lövstrand published results summarizing the susceptibility of electronic systems to HPM,[16] and **Figure 16** shows typical values for sea-level conditions where approximate magnitudes of the local HPM electric field that cause damage are shown. While the damage mechanisms are impacted by HPM pulse width, pulse frequency, and the design of electrical components, these results provide insight into potential challenges as the field strength needed for damage overlaps with the physical mechanisms of air breakdown.

With this brief description of the breakdown physics of microwave beams, the challenge associated with T&E of HPM weapons becomes apparent. With the target geometry and materials capable of impacting local atmospheric breakdown, the testing of HPM weapons becomes highly dependent on the target and altitude of engagement.

With the complexity of the interaction of HPM beams and targets, there exists a limited ability to predict damage thresholds for realistic military hardened systems. New approaches will be required for predicting damage and new targets will need to be developed to be used for T&E.

High-Energy Lasers. HELs mainly create damage to targets through thermal energy deposition. At increasing power levels, targets can be impacted by dazzling sensors, damaging sensors, or burning holes in the target.

Performance of HELs is sensitive to target hardening, laser power and beam control, and atmospheric environmental conditions. The propagation of lasers within the atmosphere is impacted by atmospheric turbulence levels, which impact beam quality, and the density of aerosols and particulates, which impact beam scattering and absorption. Atmospheric propagation factors cannot be well controlled in atmospheric testing, leading to challenges in conducting realistic operational assessments.

An additional complicating factor concerning HELs is the collateral damage that potentially results from testing the system. Collateral damage resulting from laser scattering due to target interactions has been studied and testing guidelines are in place. The potential collateral damage that results from beam propagation and impact of unintended targets is handled in low density and controlled environments, but this concern grows with increasing number of low-earth-orbit satellites and increasing numbers of air vehicles.

Observations. The development and operational employment of HEL and HPM weapons will present T&E challenges for both DT and OT that need to be resolved to fully understand weapon

---

[16] "Susceptibility of Electronic Systems to High-Power Microwaves: Summary of Test Experience," Bäckström, M.G. and Lövstrand, K.G., IEEE Transactions on Electromagnetic Compatibility, Vol. 46, No. 3, August 2004.

effectiveness. Currently, gaps exist in the understanding of the operational environment that will present challenges to assessment of operational effectiveness of directed-energy systems. In particular, the accurate prediction of atmospheric turbulence and aerosols/particulates that impact laser propagation and the prediction of subcritical microwave breakdown during HPM-target interactions are seen as shortfalls in the T&E of directed-energy systems.

Needed Best Practices. Several improvements in the T&E associated with directed-energy systems include:

- new classes of test equipment and test targets will be required to assess system V&V under realistic operational conditions;
- consideration of subcritical microwave breakdown resulting from the interaction of beamed microwaves and presentative targets should be assessed for testing of HPM systems;
- routine collection of atmospheric turbulence, aerosols, and particulates at available ranges should become a standard practice in range operations with the data compiled in a centralized repository and made available for improving the understanding of operational environments; and
- an annual assessment of T&E capacity for existing and emerging programs should be produced with advocacy for efficient use of available test capacity.

3.3.3 Challenges with T&E of Systems Using AI/ML

AI/ML technology is advancing rapidly and presenting challenges for the T&E of system capabilities being developed. The Task Force explored the issues associated with AI/ML technologies as an example of an emerging technology stressing T&E processes.

With DT&E focused on verifying that a system under development satisfies its design requirements and OT&E focused on validating that a system meets its capability intent, T&E of systems incorporating AI/ML technologies must have well-defined system requirements including a clear representation of the machine decisions that will be allowed.

In addressing the T&E of AI/ML technology, the 2020 *AI Ethics Principles*[17] were evaluated through the lens of driving T&E approaches and functions. As illustrated in **Figure 17**, the *AI Ethics Principles* demand that a system being developed is responsible, equitable, traceable, reliable, and governable. Each of these principles has implications for the T&E of these systems, including:

- The T&E workforce must be trained to understand AI and the testing approaches being used to train algorithms and verify system performance.
- Development of testing techniques for ensure adequacy of training data, discover unintended biases, and measuring system performance must occur.
- New capabilities for training operators and new OT&E approaches to evaluate the intended system behaviors in operational use will be needed.

---

[17] DoD Memorandum, "Artificial Intelligence Ethical Principles for the Department of Defense," Feb 2020; "U.S. Department of Defense Responsible Artificial Intelligence Strategy and Implementation Pathway," DoD Responsible AI Working Council, June 2022.

- Requirements for development and testing of DoD personnel to conduct fail-safe deactivation of the AI/ML capabilities must be in place.

	Implications for T&E
Responsible • DoD personnel remain responsible for the development, deployment, and use of AI capabilities	• Workforce trained in AI capabilities
Equitable • The Department will take deliberate steps to minimize unintended biases in AI capabilities	• Testing to discover unintended biases
Traceable • AI capabilities will be developed and deployed such that relevant personnel possess an appropriate understanding of the technology	• Trained operators
Reliable • AI capabilities will have explicit, well-defined uses, and the safety, security and effectiveness of such capabilities will be subject to testing and assurances within those defines uses across their entire lifecycle	• Requirements for machine decision authority • Testing in ensure adequacy of training data • Measurable performance testing
Governable • DoD personnel will have the ability to catch and deactivate AI capabilities that demonstrate unintended behaviors	• New type of Operational Test and Evaluation • Requirement for fail-safe deactivation

Figure 17. Alignment of 2020 AI Ethics Principles with AI T&E

Shneiderman[18] addresses approaches for bridging the gap between ethics considerations and practical steps of governance by developing a three-level structure built around teams producing reliable systems based on sound software engineering practices, ensuring a safety-oriented culture within organizations built through business management strategies, and trustworthy certifications by external reviews from independent oversight organizations.

The recent White House Executive Order calling for the trustworthy development and use of artificial intelligence[19] further defines many actions that directly impact the T&E of systems developed by DoD. Specifically, Section 2(a) states:

> "Artificial Intelligence must be safe and secure. Meeting this goal requires robust, reliable, repeatable, and standardized evaluations of AI systems, as well as policies, institutions, and as appropriate, other mechanisms to test, understand, and mitigate risks from these systems before they are put to use."

Section 2(a) further states:

> "Test and Evaluations, including post-deployment performance monitoring will help to ensure that AI systems function as intended, are resilient against misuse, or dangerous modifications, are ethically developed, and operated in a secure manner, and are compliant with applicable Federal laws and policies."

One area that has received significant attention recently is the question of trust in systems using AI/ML technologies. The concept of trust is an inherently human concept, and, as such, has been the

---

[18] Shneiderman, B. 2020. "Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-centered AI Systems," ACM Trans. Interact. Intell. Syst. 20, 4, Article 26 (October 2020).

[19] "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," White House, October 30, 2023.

subject of philosophers for millennia. Aspects of the concepts of AI/ML trust bleed into issues associated with T&E. For example, Ferrario et. al. have developed a multi-layer model of trust,[20] with its adaptation to warfighting trust in a system shown in **Figure 18**.

		Warfighter Believes System is Trustworthy	
		Yes	No
Warfighter is willing to rely on system without control	Yes	Paradigmatic Trust	Deviant Simple Trust
	No	Deviant Reflective Trust	No Trust

Figure 18. Model for Analyzing Human-Artificial Intelligence Interactions

In this model, paradigmatic trust is our normal expectation regarding trust that is both simple trust (i.e., we rely on others' trustworthiness to inform us regarding trust) and reflective (i.e., we rely on the belief that the system is trustworthy). This creation of paradigmatic trust of warfighters regarding the use of systems is one of the roles of the T&E process. Through formal means of demonstrating that the system performs as designed and expected, warfighters develop trust in using the system. This trust evolves with experience in application of the system through testing, training, and operations.

Deviant simple trust occurs when the warfighter does not believe the system is trustworthy but is willing to use the system without controls. This situation can occur as systems do not behave as expected, but other viable options are not available. Deviant reflective trust occurs when the warfighter believes that the system is trustworthy (i.e., the system will perform as anticipated), but is unwilling to use the system without controls due to unacknowledged instinctual suspicion or unconscious biases against using the capability. Both versions of deviant trust are typically handled through training.

In addressing T&E of AI/ML systems and the associated need to develop trust by the warfighters in using the technology, a need exists to explore the uniqueness of AI/ML applications relative to conventional systems to understand how T&E may be impacted. A conceptual view of the development pipeline for a ML application is shown in **Figure 19.** The essential elements of the AI/ML development pipeline include the software code, model or algorithm, data used in training the system, and ultimately the sensing system used to generate the data for a tactically deployed system. The development pipeline consists of model building, model evaluation and experimentation, generation of the production model, testing of the production system, deployment, and monitoring and observability. A system undergoing continuous learning and improvement will include the continuous monitoring and feedback loop depicted.

[20] "In AI We Trust: A Multi-layer Model of Trust to Analyze Human-Artificial Intelligence Interactions" Ferrario, A., Loi, M., and Vigano, E., Philosophy and Technology (2020) 22:523-539. https://doi.org/10.1007/s13347-019-00378-3.

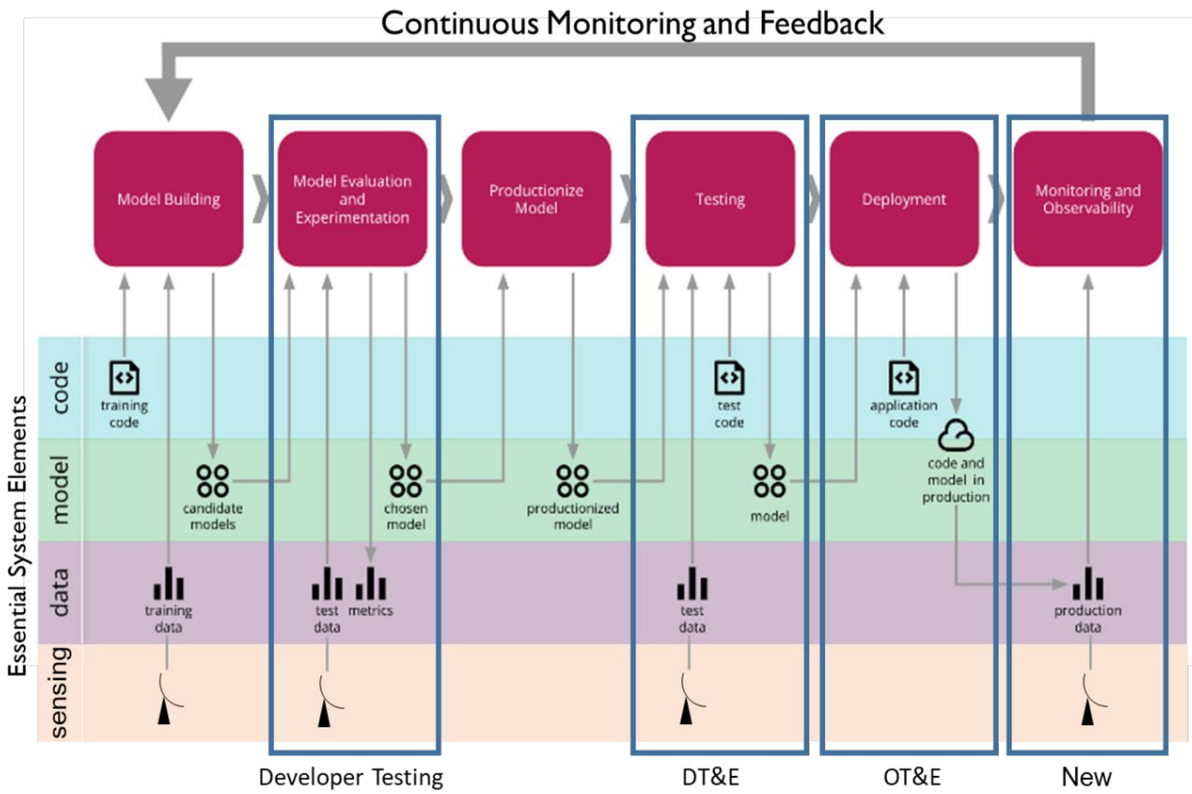


Figure 19. AI/MI Development Pipeline Illustrating T&E Functions and Need for Continuous Monitoring and Feedback [adapted from "Continuous Delivery for Machine Learning," M. Fowler, https://martinfowler.com/articles/cd4ml.html.]

In this view, the classic elements of T&E, including developer testing, DT&E, and OT&E, are readily identifiable. The new elements of the ML system shown in **Figure 19** are the continuous monitoring and feedback loop needed since the type and quality of the input data can be altered during operation deployment. Note that for a tactical system used by T&E, changes in the tactical sensing system (i.e., the system providing sampling of the real-world environment) can result in changes to the data quality which can impact the performance of the system.

With the continuous monitoring and feedback loop in an AI/ML system, the timeliness of the needed system modification becomes an important factor in the approach to T&E. **Figure 20** shows three examples requiring different update rates. First, the time available to update an automatic target recognition system looking for major combat elements may not need to be updated regularly because the major combat elements and the tactical sensing system change relatively infrequently. The algorithms and approaches to acquiring the training data are relatively well established, and training data is available across different environments with and without confusers. Further, the system performance is measurable with identifiable metrics such as target identification success rates and acceptable rates for false positives. In this case, today's T&E processes are generally adequate, and using testers-in-the-loop to conduct the independent V&V is viable.

Figure 20. Three Examples of Potential Machine Learning Applications with Differing Update Timelines

The second example illustrates a situation where an adversary is rapidly adapting, as was the case for systems aimed at suppressing improvised explosive devices (IEDs). In this case, techniques for adapting and improving the system are urgent and the requirement for producing new capability can be reduced to days or hours. In this situation, the T&E environment needs to adapt in providing rapid assessment of new system capabilities. Here, significant use of automated testing is called for and a tester-on-the-loop approach is warranted, where the tester reviews and assesses the overall results being generated by the automated testing.

The third example is one where real-time adaptation of a system will be required, and the timelines shrink below that where humans have time to react. The example shown here is a system employing intelligent and reactive electronic warfare (EW). To the extent that this capability exists in both red and blue forces, the adaptation must occur in real time as the blue system learns and adapts to observed red behaviors. Since the system is adapting and responding without human intervention, a human-out-of-the-loop approach is required. In this type of situation, the AI/ML system can still be made to obey the AI Ethic Principles through the adoption of guardrails on both inputs and outputs, although these guardrails need to be tested prior to deployment of an operational capability. In this mode of operation, the AI system may encounter situations and environments outside of the training environments for which formal T&E has been conducted, and for which the system will require further learning or human oversight.[21] To the maximum extent possible, the AI/ML system should also incorporate an out-of-distribution-detection capability that alerts operators that the system is dealing with information on which it has not been trained.

Observations. A complex range of T&E challenges is introduced with the development of AI systems resulting from the rapidly changing landscape of sensors, data, algorithms, hardware, and software systems. The current rigid approach to T&E tied to program milestones will be unsuccessful in responding to the system development and deployment needed in a dynamic environment. The approach to T&E will need to expand beyond the current acquisition system cycle where OT is tied to initial production to incorporate continuous feedback and monitoring for continual improvement.

A further challenge is associated with the large number of possible threat vectors for adversarial actions against AI systems, which impacts the type of testing required. The National Institute of

---

[21] "AI Autonomy: Self-initiated Open-world Continual Learning and Adaptation," B. Liu, S. Mazumder, E. Robinson, and S. Grigsby, AI Magazine, 2023, 44:185-199.

Standards and Technology has recently published a taxonomy for attack vectors.[22] These include: (i) taking control of training data to insert or modify training data; (ii) taking control of the model parameters; (iii) taking control of testing data to influence test outcomes; (iv) controlling labels used in supervisory training; and (v) controlling queries accessing when using cloud environments. Attackers can further disrupt AI-enabled systems through manipulating the inputs to the sensing element in an operational system, which is a modern-day version of using camouflage to confuse and deny identification.[23]

The challenges associated with T&E of AI systems are amplified when dealing with systems that must be updated either rapidly or continuously. Again, continuous monitoring and feedback will be essential elements.

Needed Best Practices. After reviewing needed T&E elements for AI/ML systems, best practices have been identified as follows.

- The requirements for AI systems should be developed with a clear view towards testability with access to data and performance metrics.
- The requirements should address decision authority granted to system, performance, and test requirements.
- For systems anticipated to require rapid or real-time learning, guardrails and out-of-distribution detection will be required to enhance trust in system.
- Continual improvements in infrastructure investments should be pursued to capture and centralize data together with associated tools for data access and system performance refinement.
- An annual assessment of T&E capacity for existing and emerging AI capabilities should be produced with advocacy for efficient use of available test capacity.

3.4 Mission-Focused T&E

The final area considered that is providing stressors to the T&E enterprise concerns the desire for creating joint mission-focused capabilities. As illustrated in **Figure 21**, joint mission threads pull together individual system capabilities to create end-to-end mission capabilities. Historically, this system-of-systems challenge has been addressed through a focus on system interoperability within individual system acquisitions. The integration across systems to produce end-to-end effects has largely fallen to warfighters.

---

[22] Vassilev, A. Oprea, A., Fordyce, A., Anderson, A. (2024) "Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations.," (National Institute of Standards and Technology, Gaithersburg, MD) NIST Artificial Intelligence (AI) Report, NOST Trustworthy and Responsible AI NIST AI 100-2E2023. https://doi/org/10.6028/NIST.IA.100-2e2023 .

[23] "Adversarial Patch Camouflage Against Aerial Detection," A. Adhikari, et. al., arXiv:2008.13671v1, 31 Aug 2020. [2008.13671] Adversarial Patch Camouflage against Aerial Detection (arxiv.org).

Development of joint mission threads, including their T&E, is challenged under three different situations. First, the T&E of the joint mission threads are stressing as the need exists to cross Military Service acquisition boundaries with systems still under development. Each Military Service has test organizations established to conduct DT and OT in the framework of the capability acquisition. Further, some acquisition activities are underway that have an embedded mission focus to be developed. Examples include the United States Navy (USN) Aegis system, which integrates anti-air warfare and ballistic missile defense missions, and the F-35 program with its focus on air dominance. The development of joint warfighting mission threads and their T&E challenges arise with the crossing of Military Service boundaries, often leaving critical engineering development and T&E approaches weakly defined.

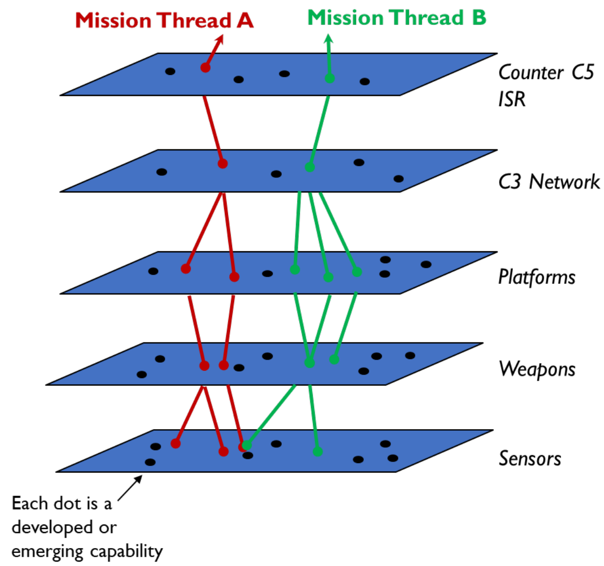


Figure 21. Joint Mission Threads with Emerging New Capabilities will Stress Conventional T&E Approaches

The second situation is the desire for rapid development of mission capabilities to satisfy existing warfighting needs as expressed by JUONs or JEONs. Combatant Commands often conduct experiments and demonstrations aimed at satisfying these needs as adjunct activities to major training exercises. These experiments and demonstrations often use technically immature capabilities, but also have shown potential operational capabilities within the limitations present in the training event. The T&E challenges become evident as rigorous assessments of the technology's readiness for use under true operational conditions occur. The acquisition-based approach to T&E is not well positioned to rapidly accelerate development of these urgently needed capabilities.

The third situation is associated with a desire to construct end-to-end mission capabilities by dynamically connecting individual and emerging systems. This approach of creating dynamic mission threads is believed to create an innate resilience in the overall system-of-systems since single points of failure are not present. The T&E challenge encountered with this approach is that the testing must evaluate both the individual system capability and its ability to meet the objectives of the end-to-end mission. As the number of systems becomes larger and ever evolving, the enabled mission complexities and ever-increasing number of pathways will challenge the T&E enterprise, which is not positioned for this future. The mantra of "test like you fight" will likely not be possible due to the complexity of this environment. Furthermore, individual systems become a smaller portion of the overall system-of-systems capabilities, with the attendant need for focusing activities on mission-based T&E.

To summarize the challenges (**Figure 22**) that are stressing today's approaches to the DoD T&E enterprise, we see:

- an increasing pace of adversary threat capabilities;
- a need for increasing speed and efficiency of introducing new warfighting capabilities;
- the emergence of new technologies for which historical T&E approaches are inadequate;
- an increasing use of artificial intelligence and autonomous systems;
- an increasing use of software-intensive systems; and
- an increasing focus of joint system-of-systems for execution of warfighting mission threads.

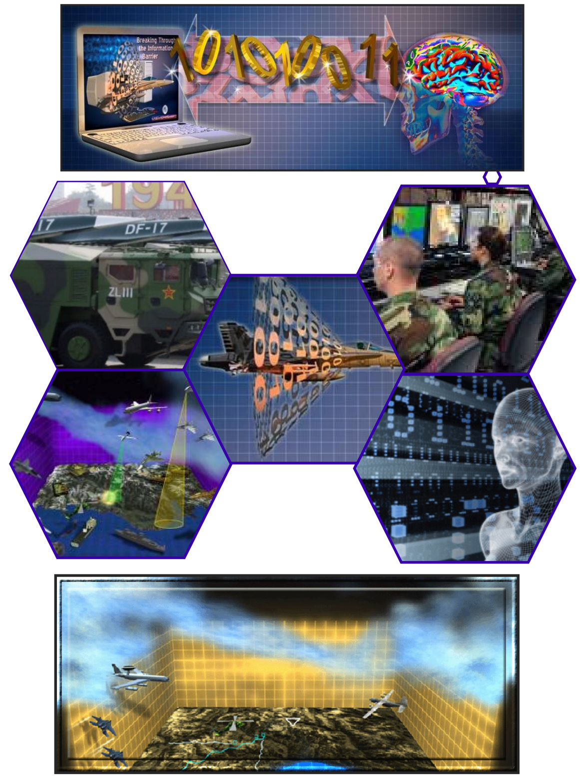


Figure 22. Illustration of Today's T&E Challenges.

4. Findings and Recommendations

4.1 A Need for a Strategic Shift in Test & Evaluation

With the wide range of challenges facing today's approach to T&E, the Task Force developed five overarching findings and recommendations (F&R) with the first focused on a need for a strategic shift in the approach for T&E. The Task Force found that today's challenges associated with T&E are unlikely to be overcome without a strategic shift in the current approach of acquisition-based T&E. To close gaps and increase efficiencies regarding T&E, DoD's approach to T&E needs to be broadened beyond today's acquisition-based T&E to extend the concept of T&E both before programs are formalized and with a continuation of T&E activities following system introduction into training and operational environments.

F&R 1: Joint Warfighting Mission Threads

Finding 1. Rapidly emerging technologies, increasingly capable adversaries, and joint warfighting mission threads demand a significant strategic shift in the focus of T&E to augment today's approach of acquisition-based T&E.

Recommendation 1. **USD(A&S)**, **USD(R&E)**, and **DOT&E** should expand T&E activities to address needed required strategic shift.

1.1. Shift T&E to include S&T development leading into programs.

- **USD(R&E)** and **USD(A&S)** should revise Technology Readiness Levels and standards for Technology Readiness Assessment to set assessment to include data, software, M&S tools, and test results readiness for transition to acquisition programs.

1.2. Incorporate warfighter input early and use operational testing to inform and seamlessly transition into training.

- **USD(A&S) should direct Service Acquisition Executives** to create plans to fully leverage OT events for informing system training.

1.3. Ensure robust testing of joint warfighting mission threads.

- **USD(R&E) should direct the Director for Multi-Domain Joint Operations** to ensure sufficient system-of-systems mission engineering is conducted to form the basis of detailed integrated mission testing strategies.
- **USD(R&E) should direct DTE&A** to generate guidance for DT of all defined joint warfighting mission threads.
- **DOT&E** should ensure operational robustness of joint warfighting mission threads using live and virtual environments.

1.4. Develop capability for continuous data and user feedback from training and operations.

- **USD(A&S) should direct Service Acquisition Executives** to enable collection of system data from training venues and operational events together with user feedback to inform continual system improvement.

Recommendation 1.1 calls for an expanded view of T&E, which will require coordination across USD(A&S), USD(R&E), and DOT&E, and addresses the needed shift of T&E to address S&T activities prior to initiation of formal acquisition programs. As the desire for new system capabilities draws on an increasing speed of technology adaptation and system T&E becomes more heavily reliant on M&S for system V&V, the S&T conducted prior to initiation of a formal acquisition program becomes critical to speeding technology adaption and prevention of duplication of effort.

Prior to initiation of a formal acquisition program, technology development naturally involves experimentation and model development with ongoing V&V activities aimed at improving the understanding of and ability to predict the underlying physical phenomena. This knowledge gained serves as the foundation of efforts to incorporate a technology into development of a new system. As an example, the knowledge gained from model V&V activities conducted during S&T programs should flow seamlessly into acquisition programs without the need for repetition of V&V activities. Currently, the knowledge accrual process that occurs prior to the initiation of a program is not viewed as part of the T&E process; therefore, the opportunity is lost for increasing efficiency and accelerating transition of the technology using better transition products.

Furthermore, USD(R&E) and USD(A&S) need to revise TRLs and standards for technology readiness assessment to recognize the value of data, software, M&S tools, and test results readiness in the transition to acquisition programs. As Title 10, Section 4272, states:

> "With respect to a major defense acquisition program, the Secretary of Defense shall conduct or approve and independent technical risk assessment (1) before any decision to grant Milestone A approval for the program… that identifies critical technologies or manufacturing processes that need to be matured; and (2) before any decision to grant Milestone B approval for the program… to enter low-rate production or full-rate production… the identification of any critical technologies or manufacturing processes that have not been demonstrated in a relevant environment."

Currently, DoDI 5000.88 defines the process for the engineering of defense systems.[24] Accordingly, USD(R&E) is responsible for establishing the policy and guidance of ITRAs consistent with Title 10, Section 2448(b). USD(R&E) is also responsible for the conduct and approval of ITRAs for ACAT 1D programs and determination of the ITRA approval authority for ACAT 1B and 1C programs and establishing policy for oversight authority over all DoD uses of developmental prototyping, except that conducted under MTA.

The Technology Readiness Assessment Guidebook[25] provides guidance on initiation, organization, and conduct of TRAs on acquisition programs. Building on the concepts of Critical Technology Elements (CTEs), TRLs provide a structured framework for identifying risks and technology maturation needs. The TRL scale is used widely and is generally defined as the following for hardware-related CTEs:

---

[24] "DoDI 5000.88: *Engineering of Defense Systems*," Office of the Under Secretary of Defense for Research and Engineering, Nov. 18, 2020.

[25] "Technology Readiness Assessment Guidebook," Office of the Executive Director for Systems Engineering and Architecture, OUSD(R&E), June 2023.

- TRL 1: Basic principles observed and reported.
- TRL 2: Technology concept and/or application formulated.
- TRL 3: Analytical and experimental critical function and/or characteristics proof of concept.
- TRL 4: Component and/or breadboard validation in a laboratory environment.
- TRL 5: Component and/or breadboard validation in a relevant environment.
- TRL 6: System/subsystem model or prototype demonstration in a relevant environment.
- TRL 7: System prototype demonstration in an operational environment.
- TRL 8: Actual system completed and qualified through test and demonstration.
- TRL 9: Actual system proven through successful mission operations.

The S&T community generally aims to mature CTEs to a TRL 6, and acquisition programs generally conduct assessments of technology readiness at their onset to ensure TRL 6 readiness. Examples provided of typical supporting information include:

> "Results from laboratory testing of a prototype system that is near the desired configuration in terms of performance, weight, and volume. How did the test environment differ from the operational environment? Who performed the test? How did the test compare with expectations? What problems, if any were encounters? What are/were the plans, options, and actions to resolve problems before moving to the next level?"

Through redefining the definition, description, and supporting information associated with TRLs and specifically the TRL 6 goal used in starting programs, a faster, more efficient, and seamless transition of technologies from S&T activities to acquisition programs can result. The needed additional components of the CTE transition (generally performed by S&T components) and then used within TRA processes should include:

- formal assessment of the CTE range of applicability including environmental limitations;
- documentation of CTE testing and results including formal assessment of test result uncertainties;
- CTE M&S toolsets prepared for transition with accompanying documentation and formal assessment of applicable environments and underlying performance uncertainties; and
- documentation and delivery of all non-standard data specific to a CTE that will be required in subsequent acquisition activities (examples include properties of novel materials, fuels, and fabrication processes).

This recommendation addresses challenges presented to the Task Force associated with inadequate technology readiness assessment at the start of programs, especially programs operating through the MTA pathways.

It should be noted that for software-related CTEs, the TRL scale is adjusted in TRLs 4-9 as follows:

- TRL 1: Basic principles observed and reported.
- TRL 2: Technology concept and/or application formulated.

- TRL 3: Analytical and experimental critical function and/or characteristics proof of concept.
- TRL 4: Module and/or subsystem validation in a laboratory environment (i.e., software prototype development environment).
- TRL 5: Module and/or subsystem validation in a relevant environment.
- TRL 6: Module and/or subsystem validation in a relevant end-to-end environment
- TRL 7: System prototype demonstration in an operational, high-fidelity environment.
- TRL 8: Actual system completed and mission qualified through test and demonstration in an operational environment.
- TRL 9: Actual system proven through successful mission-proven operational capabilities.

Software development is quite different than hardware development in that software technology can rapidly transition through the TRLs. As such, the recommendations regarding software intensive systems are addressed separately under Finding/Recommendation 4. For software associated with hardware CTEs, the issues of code stability, supporting data, and state of V&V remain the same as hardware-related CTEs. For example, the data sets used to train AI/ML models during S&T development shown now become evaluated for their immediate transition and incorporation into the acquisition system development.

Recommendation 1.2 calls for incorporating warfighter input early and using operational testing to inform and seamlessly transition into training. This recommendation is aimed at expanding the definition of T&E in filling the seam that exists between T&E and warfighter training and operations. Currently OT events are used to validate the operational utility, thereby bringing operators into the T&E environment. The Task Force notes that the 2022 *DOT&E Annual Report* often identifies training issues associated with shortfalls in OT:

- *Command Post Computing Environment (CPCE):* "Training provided to soldiers did not prepare them to make full use of advanced features, troubleshooting, and employment of CPCE Increment 1 in a collaborative manner."
- *Distributed Common Ground System – Army (DCGS-A) Capability Drop 2*: "The Army should… Develop tools, technology, and training for personnel to support testing of advanced data-centric systems such as DCGS-SA CD2 and to prepare for the advanced data analytics and Artificial Intelligent systems of the future."
- *Electronic Warfare Planning Management Tool (EWPMT)*: "The Army should… Refine training to emphasize troubleshooting and help leaders and staff understand EEPMT INC1 capabilities and operational employment."
- *Joint Air-to-Ground Missile (JAGM)*: "The Joint Program Manager and Navy should… Continue development and integration testing of the JAGM Captive Aircrew Training Missile while developing unique TTPs to ensure aircrew effectiveness."
- *Limited Interim Missile Warning System (LIMWS)*: "The Army should… Continue collecting data and evaluate system suitability and make system and/or training modifications as needed."

- *Terrain Shaping Obstacles (TSO)*: "The Army should… Address deficiencies found in the training materials to ensure that soldiers emplacing the XM204 understand the sensor limitations and avoid terrain features that would impact system performance."
- *Amphibious Combat Vehicle – Command and Control Variant (ACV-C)*: "The Marine Corps should… Train ACV-C crews to adequate support the C2 mission."

With a goal of expanding the vision of T&E and increasing efficiency in the development and employment of new capabilities, this recommendation is aimed at fully leveraging OT events as an early evaluation of training system effectiveness. The Task Force realizes this vision will require closer coordination of programs, operational test agencies, and training centers.

Recommendation 1.3 calls for ensuring robust testing of joint warfighting mission threads. Currently, T&E is authorized, organized, and executed around acquisition programs. The Combatant Commands and Joint Staff are defining joint warfighting mission threads that bring together multiple systems to produce new mission capabilities. This system-of-systems approach for new mission capabilities crosses multiple acquisition programs and multiple Military Services, and often relies on emerging new capabilities. Given the integration of systems of varying maturity and a dynamic adversary environment, many challenges are being realized in the V&V of the end-to-end mission capability. By extending the viewpoint of T&E beyond individual acquisition programs, the role of T&E in development of joint warfighting mission threads must be considered.

Much of the mission engineering for these emerging joint capabilities is being conducted under the purview of the OUSD(R&E) Director for Multi-Domain Joint Operations. This effort must ensure sufficient system-of-systems mission engineering is conducted to form the basis of detailed integrated mission testing strategies. Using structured systems engineering processes to deconstruct the mission thread requirements (i.e., the "left side" of the systems engineering Vee), the requirements for DT and OT (i.e., the "right side" of the systems engineering Vee) can be defined. Significant attention should be paid to system interface specifications and information flows, which will be heavily leveraged in needed T&E activities.

With the joint mission threads defined, the appropriate process for conducting the DT of the warfighting capability can be developed. In this case, USD(R&E) should direct DTE&A to generate the needed guidance for DT of all defined joint warfighting mission threads. The DT guidance should provide needed information on test strategies, process controls, data requirements, and data flows.

Finally, significant challenges will be realized in the OT of joint mission threads, which will stress the ability to "test like you fight." DOT&E should ensure operational robustness of joint warfighting mission threads, which will likely rely heavily on the use of live and virtual environments given the challenges of replicating true operational environments.

Recommendation 1.4 calls for the development of the capability for continuous data and user feedback from training and operations. In particular, USD(A&S) should direct Service Acquisition Executives to enable collection of system data from training venues and operational events together with user feedback to inform continual system improvement.

This recommendation recognizes the inherent need and value in continued evidence accrual following formal OT assessment and production decisions. Through expanding the formal roles of T&E and expanding beyond acquisition system T&E, capturing the knowledge that continues to accrue during training and operations is critical. This knowledge is essential in further refining the true capabilities of systems which will flow into continual system improvements.

F&R 2: Improving T&E Efficiency Using Continuous Threat Engineering and Assessment

Finding 2.1. Rapid adversary capability development is resulting in disconnects between threat assessments used to define program requirements and the threats used for operational assessment, often conducted years later.

Finding 2.2. Many complex weapon system lifetimes will exceed the knowledge of current and projected threat capabilities and a physics-based threat assessment can be used to allow consideration of future threats as technologies evolve.

Finding 2.3. A deep tie between the system developers, intelligence community, and the DoD threat assessment community is critical to ensure intelligence is relevant and adversary advancements are reflected in system assessments.

- DoD threat assessment community must understand both intelligence and state of capability development.

Recommendation 2.1. **Service Acquisition Executives should direct Program Executive Officers** to develop, maintain, and fund threat assessment teams knowledgeable of existing and developing systems to coordinate with the Service and Combatant Command intelligence components, Defense Intelligence Agency, and other Intelligence Community elements to conduct continuous evaluation of emerging threats.

Recommendation 2.2. **USD(R&E) should direct ASD(Mission Capabilities)** to coordinate with the Strategic Intelligence and Analysis office to provide continuous assessments of emerging threats to joint warfighting mission threads.

The Task Force's second major set of findings and recommendations to support the need for a strategic change in T&E addresses the challenges associated with rapid advancements in adversary capabilities. With the internationalization of science and technology and the accelerating pace of technology adoption for new military capabilities, there is a widening gap between the threats evaluated in setting initial program requirements and those that exist as the program nears production and operational capabilities. These threat advancements naturally create T&E challenges between the DT conducted to validate that a system meets its initial requirements and OT testing conduct to ensure operational relevance.

It must also be recognized that many complex weapon system lifetimes will exceed the knowledge of current and projected threat capabilities. This limitation can be somewhat overcome by using a physics-based threat assessment to set program requirements to allow consideration of future threats as technologies naturally evolve, but it should be anticipated that predictions concerning future capabilities will never be fully accurate.

To address this widening gap between DT and OT, there is a need for continual threat engineering and assessment over the course of an acquisition program. This continual assessment will require a deep tie between the system developers, the intelligence community, and the DoD threat assessment community to ensure intelligence is relevant and adversary advancements are reflected accurately in blue system assessments.

One view of the development timeline for a typical adversary weapon capability and a typical acquisition program aimed at defeating that weapon is shown in **Figure 23.** The responsibility for development of high-fidelity representations of threat capabilities lies with the intelligence community and can often take years to develop. In many cases, aspects of threat capabilities are driven by physical processes and constraints are often defined much faster compared to the subtle nuances of sensor operation or decision logic. In this situation, physics-based threat definitions can be used to set acquisition program requirements together with broad capability requirements set for the aspects of threat systems that are expected to change on much faster timescales.

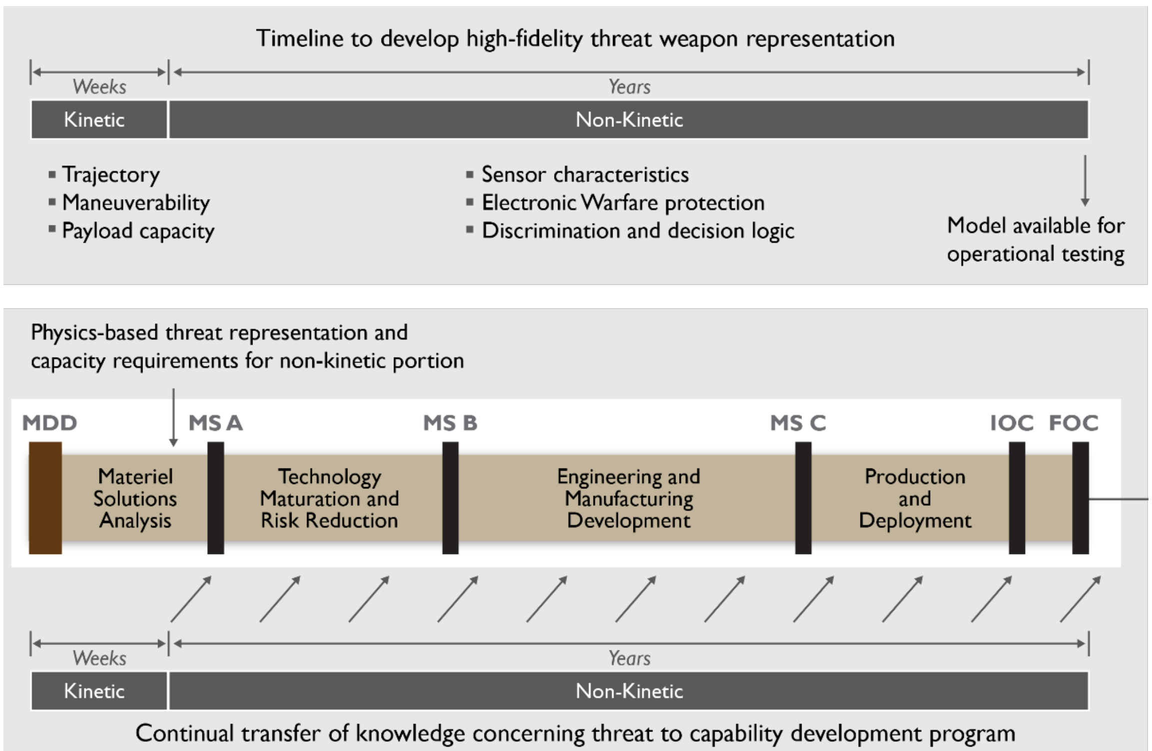


Figure 23. Threat Assessment and Continual Assessment During Program Execution

In the situation illustrated in **Figure 23**, a broad knowledge of threat systems is used to set the requirements for a program, but continual threat assessment is conducted to both ensure the physics-based threat assessment remains valid and to flow additional information concerning the information-based elements into the blue capability development, which can react to changes on much faster timescales. In cases where the threat capabilities develop beyond that planned in the system acquisition, a formal risk assessment should be conducted with either acceptance of the risk or development of a plan for a blue system block upgrade to handle the emerging threat capability.

Recommendation 2.1 states that the Service Acquisition Executives should direct Program Executive Officers (PEOs) to develop, maintain, and fund threat assessment teams knowledgeable of existing and developing systems. These teams should coordinate with the Defense Intelligence Agency and other intelligence community elements to conduct continuous evaluation of emerging threats. The rationale for recommending this action directly to PEOs is that PEOs are best positioned to both understand blue systems under development and to drive needed intelligence across a family of capabilities under development.

Recommendation 2.2 states that USD(R&E) should direct ASD(Mission Capabilities) to coordinate with the Strategic Intelligence and Analysis office to provide continuous assessments of emerging threats to joint warfighting mission threads. With joint mission threads crossing individual acquisition programs and Service boundaries, the equivalent continuous threat assessment for joint mission threads is best conducted under USD(R&E), where engineering development is already focused for these new mission capabilities.

F&R 3: Improving System Performance Through Continuous Testing

Finding 3.1. Continuous testing offers a pathway for improving T&E efficiency through continual evidence accrual.

- A structured DE framework is essential for aligning development, simulation, and testing activities.

Finding 3.2. Significant automated testing will be required to cover the operational “envelope” of complex systems.

Recommendation 3.1. USD(A&S) should direct Service Acquisition Executives to structure new programs to:

- incorporate testability requirements in components, subsystems, and systems to speed evidence accrual;
- maximize use of automation to increase testing for systems and subsystems; and
- develop approaches to report system status and data to enable feedback for improving system performance.

Recommendation 3.2. USD(R&E) should direct DTE&A to develop and promulgate DT guidance to ensure system capability to use automated developmental testing to the maximum extent possible.

Recommendation 3.3. USD(A&S) should develop a repository for collection and exploitation of system data collected after delivery to training and operational units.

As discussed under Findings 1 and 2, a strategic shift in T&E is needed to better reflect the knowledge of a new system capability. This shift involves incorporating information generated prior to the start of an acquisition program and extends past formal production decisions to incorporate information gathered during training and operations. Further, continual assessments of threat system capabilities are needed as the environment under which blue systems must operate are continually changing.

This environment leads to a need to change the perspective of T&E to one of continuous testing. Under OUSD(R&E), DTE&A has published a paper on "Test and Evaluation as a Continuum,"[26] discussing the needed transition to continuous testing building on the attributes of: 1) capability and outcome focused testing; 2) an agile scalable evaluation framework; and 3) enhanced test design.

As part of the strategic shift in T&E, the Task Force found that continuous testing offers a pathway for improving T&E efficiency through continual evidence accrual. This testing begins prior to the formal start of an acquisition program and extends well past its completion as systems enter training and operation. An enabler for this continuous testing environment is a structured DE framework for aligning development, simulation, and testing activities.

Examples of testing as a continuum include:

- testing conducted under S&T development informs uncertainties inherent in M&S tools that flow directly into acquisition systems assessments;
- testing lessons learned during a prototype demonstration conducted prior to initiation of an acquisition program flow directly into test planning to include needed diagnostics, data acquisition and management successes and challenges, and resulting in uncertainties in technology performance and operability;
- contractor testing results and supporting data seamlessly transition to DT evaluations;
- testing results and lessons learned across related blue system capabilities and acquisitions are shared in common frameworks;
- lessons learned during OT events are seamlessly integrated into operator training; and
- data collection is enabled during training and operations to enable continual assessment of system performance and capabilities.

With the continual testing approach, significant use of automated testing will be required to cover operational "envelope" of complex systems.

Recommendation 3.1 states that USD(A&S) should direct Service Acquisition Executives to structure new programs to: a) incorporate testability requirements in components, subsystems, and systems to speed evidence accrual; b) maximize use of automation to increase testing for systems and subsystems; and c) develop approaches to report system status and data to enable feedback for improving system performance.

While the test ability of system requirements is a fundamental precept of sound systems engineering, new approaches are emerging for the automated testing of complex systems. With the increasing dimensionality of complex systems and an exploding parameter space for conventional testing, new approaches will be required for system testing within reasonable cost and schedules. A

[26] "Test and Evaluation as a Continuum," Collins, C.C. and Senechal, K., The ITEA Journal of Test and Evaluation, 44(1), March 2023. https://itea.org/journals/volume-44-1/test-and-evaluation-as-a-continuum.

taxonomy for the use of artificial intelligence in driving testing approaches for complex systems[27] addressed four levels of autonomy of AI in system testing.

- Level 0 — AI is not applied.
- Level 1 — AI algorithms assist humans by performing (semi-)automated testing tasks.
- Level 2 — AI replaces or mimics human behavior.
- Level 3 — System testing is done fully automated by AI agents.

As system complexity increases, the need to transition towards Level 3 use of AI for testing becomes paramount.

T&E activities often end at the completion of the OT&E activities, with initial operational test & evaluation used to support a full-rate production decision and follow-on test and evaluation (FOT&E) to verify operational effectiveness and suitability of the production system.

It is widely recognized that significant additional learning about the capabilities of a new system occur as the system is deployed into training and operational environments. As compared to a specification-based description of the system performance, actual system performance becomes apparent over time. Further, the operational use of a system matures as operators find opportunities for exploiting additional performance from the integration of a system into a system-of-systems construct.

As discussed in Section 2.3, modern commercial entities continuously test systems and work to improve them through post-deployment operations using automated testing, assessment, and reporting. User feedback and lessons learned directly feed continual improvement of the product, and this post-deployment feedback is an essential characteristic of systems designed to change, adapt, and learn over their lifetime.

A block diagram of this continual feedback process is shown in Figure 24 for the context of a new capability operating within a system-of-systems mission thread. The current T&E process employs a coupling of the capability development and DT process, ultimately demonstrating that the initial system requirements are being met. OT mostly follows establishment of the baseline system performance and serves to demonstrate performance in an operational environment. As shown in the block diagram, significant opportunity exists for continued evidence accrual through training and operational deployment. With a strategic shift to view T&E as a continual process, approaches to continuously improve knowledge of a system's performance, or improve the system performance itself, enable continual product growth.

[27] Felderer, M., Enoiu, E.P., Tahvili, S. (2023). Artificial Intelligence Techniques in System Testing. In: Romero, J.R., Medina-Bulo, I., Chicano, F. (eds) Optimizing the Software Development Process with Artificial Intelligence. Natural Computing Series. Springer, Singapore. https://doi.org/10.1007/978-981-19-9948-2_8

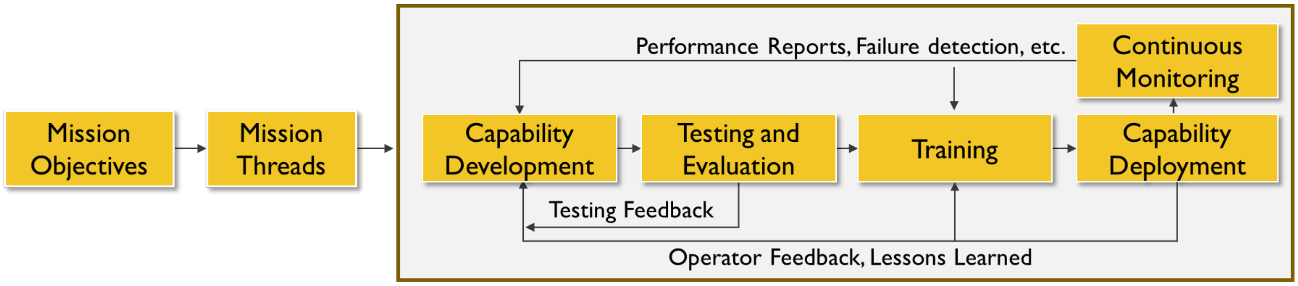


Figure 24. Continual system improvement through feedback from training and operations.

Recommendation 3.2 states that DTE&A should develop and promulgate DT guidance to ensure system capability to use automated developmental testing to the maximum extent possible. This DT guidance should work towards development of standard data reporting requirements and test interfaces implemented. With data standards and reporting requirements established, mechanisms for reporting, compiling, and analyzing post-deployment performance data can be enabled. Further, by automating DT and by leveraging standard interfaces, AI-based tools can be developed to explore and test a more complete sampling of the full operational domain of complex systems both before and after deployment of the new capability.

Recommendation 3.3 states that USD(A&S) develop a repository for collection and exploitation of system data collected after delivery to training and operational units. With systems designed to collect and report data after deployment to training and operational environments, a repository to collect and hold this data and information is needed.

The information held in this type of repository will enable refinement of the M&S tools in continual development to facilitate a system-of-systems analysis capability. Access to this repository will enable system developers and researchers to explore options for system upgrades and improvements in a continuous process mimicking that used in commercial practices.

The Task Force recommendations on continual testing will:

- close testing gaps between S&T and acquisition development test activities;
- close gaps in T&E activities with discrete transitions from contractor testing, DT, and OT;
- enable a continuation of a T&E mindset after deployment of capability into training and operations;
- provide a continuous and efficient capability development pipeline that mimics commercial best practices; and
- prepare DoD for a future of complex system-of-systems mission-focused operations where historical T&E approaches will not be sufficient and the need will emerge to leverage automated AI-driven testing.

F&R 4: Improving T&E Speed and Efficiency of Software Intensive Systems

Finding 4. For software being delivered to weapon systems, today's milestone-based T&E inhibits intended use of software designed pathways.

- At present, no weapon system program is being executed using the Software Acquisition Pathway.
- An execution inconsistency results in the Software Acquisition Pathway instructions regarding continuous cybersecurity testing and mandated requirement to use the 177th Information Warfare Aggressor Squadron (177 IWS) for adversarial assessments.

Recommendation 4.1. USD(R&E) should establish development guidelines for software-intensive systems to be built using an architecture that allows segmentation of the software stack (i.e., application, middleware, operating system interface, firmware, hardware) to enable testability of all layers.

Recommendation 4.2. USD(A&S) should direct new software-intensive system development to segment safety critical functions from non-safety critical functions and execute test-driven development of non-safety critical system development using the Software Acquisition Pathway.

One of the biggest challenges associated with current T&E processes is the acquisition of new capabilities associated with software-intensive systems. The DSB conducted a study[28] of this issue in 2018 with findings regarding the role of continuous iterative development, lessons to be learned from the commercial world, and misalignments that exist in the Department's acquisition of software. Recommendations were provided regarding requirements for software factories, continuous iterative development, risk reduction and metrics, needed workforce, software sustainment, and the emerging need for independent V&V of machine learning.

Since the 2018 DSB study, the Software Acquisition Pathway has been defined and is available, but the current Task Force found that agile development processes remain in conflict with the T&E requirements for weapon system acquisition programs. The Task Force found that for software being delivered to weapon systems, milestone-based T&E inhibits the intended use of software designed pathways, and, at present, the benefits of the Software Acquisition Pathway are not being realized in weapon system programs.

The challenge of software-intensive system development was seen in the Air Operation Center (AOC) discussed earlier. With the AOC weapon system structured as a software acquisition pathway program under DOT&E oversight, milestone-based T&E activities are conducted and reviewed. In parallel, an agile software development process with embedded DT and OT elements is creating capabilities that could be delivered widely to operational environments on timescales much faster than the milestone-based review process enables. Similar challenges exist in other large weapon system development programs that include significant software-intensive components (e.g., Aegis Modernization and Joint All-Domain Command and Control).

These challenges between the timeline differences of an agile software development environment and the milestone-based T&E process is most clearly seen in an execution inconsistency in instructions regarding continuous cybersecurity testing and the mandated requirement to use the

[28] "Design and Acquisition of Software for Defense Systems," Defense Science Board, Office of the Under Secretary of Defense for Research and Engineering, Washington DC. February 2018.

177 IWS for adversarial assessments. Software developed in a DevSecOps environment undergoes nearly continuous testing of cyber vulnerabilities using automated static and dynamic testing tools. Current OT requirements for assessing system survivability place a heavy emphasis on red teaming to assess vulnerabilities. It should be recognized that the two different types of vulnerability testing are not incompatible with each other, and vulnerabilities found in either should be quickly remedied within the agile environment.

Recommendation 4.1 states that USD(R&E) should establish development guidelines for software-intensive systems to be built using an architecture that allows segmentation of the software stack (i.e., application, middleware, operating system interface, firmware, hardware) to enable testability of all layers. Through effective segmentation of the architecture and control of interfaces, automation techniques are enabled for testing at the interface level without disrupting the other system layers. With this effective segmentation of the software developed for systems, agile processes can be fully utilized for speeding system development.

Recommendation 4.2 states that USD(A&S) should direct new software intensive system development to segment safety critical functions from non-safety critical functions and execute test-driven development of non-safety critical system components using the Software Acquisition Pathway. Major acquisition systems are large complex systems that often bundle both safety critical and non-safety critical functions within the same acquisition program. This approach results in a single set of development milestones for the entire program with the attendant T&E requirements established by statute. The authority for this recommended segmentation currently exists under Title 10, Section 4203, which states:

> "If the Secretary of Defense determines that a major defense acquisition program requires the delivery of two or more categories of end items, which differ significantly from each other in form and function, the Secretary may designate each such category of end items as a major subprogram for the purposes of acquisition reporting."

By segmenting safety critical and non-safety critical functions in a software-intensive system development, different T&E processes can be employed. The non-safety critical functionality can be acquired using the Software Acquisition Pathway, fully implementing DevSecOps in an environment of rapid and continual improvement. The safety critical software-intensive process can still be conducted in an agile DevSecOps environment, but those capabilities can be subjected to more intensive testing and certification processes. Looking forward to needed changes in authorities and acquisition instructions, further refinements will be required to fully realize the ability to operate in agile environments. Considerations of acceptable risks and "levels of doneness" will need to be structured to account for risks, development team accountability, and team maturity.

As one looks to a future of increasingly complex system behaviors and desire for increasing speed of new capability development, a further strategic shift in T&E of software intensive systems will be required. The human ability to test systems will be outpaced by the automation provided by intelligent machine testers due to the many orders of magnitude of testing that can be accomplished. As this future is realized, a transition will likely be needed to shift the testing of even safety critical systems to this automated environment with human-in-the-loop oversight.

The Task Force recommendations regarding development of software-intensive systems will:

- increase the speed and efficiency of T&E for software intensive systems;
- provide continuous test practices that mimic commercial best practices; and
- prepare DoD for a future of complex system-of-systems operations where the most complex behaviors are contained in the software being used with those systems.

F&R 5: Emerging Technologies: Linking Technology Development and Test Strategies

Finding 5.1. For emerging technologies, close coupling of the technology development with its testing approach is a key enabler for increasing the speed and efficiency of T&E.

- Testing approaches will need to advance and mature in parallel with the technology itself.

Finding 5.2. Consideration of existing and potential adversary countermeasures will also need to mature in parallel with the technology itself.

Recommendation 5. **USD(R&E)** should ensure timely development of test techniques and strategies for critical technologies.

- Require the Principal Directors for Critical Technologies and TRMC to generate assessments of test approaches and requirements and advocate for needed resources.
- Require the DTE&A to develop guidance for acceptable test strategies for all critical technologies to ensure readiness for transition to formal programs.
- Ensure warfighter inputs into technology development to accelerate transition to programs and subsequent operational testing.

As discussed in Section 3.3, systems using new technologies stress the T&E process, often requiring new analytical tools and techniques. In conducting a deep dive into the emerging technologies associated with hypersonics, directed energy, and AI/ML, the Task Force found that close coupling of a technology development with its testing approach will be required and will be key to increasing the speed and efficiency of T&E for systems employing emerging technologies. Often, the testing approaches will need to advance and mature in parallel with the technology itself early in development. Realization of the synergism between development of the technology and its test needs leads to one of the previously discussed drivers for continuous testing.

By definition, emerging new technologies will involve physical processes and techniques not previously explored within conventional T&E processes. While existing T&E capabilities can be leveraged to the maximum extent possible, it should be expected that new test diagnostic equipment and new analytical tools will be required to conduct the T&E function. Since these new capabilities are being explored during the technology development phase, it is essential that knowledge gained during those activities seamlessly transitions into the acquisition process.

In parallel with the development of new testing approaches for emerging technologies, one must also consider the impact of existing and potential adversary countermeasures. Since the external environment is never static, this analysis capability will also need to mature in parallel with the technology itself. Techniques for improving predictive capabilities including adversary adaptations will be required.

Recommendation 5 states that USD(R&E) should ensure timely development of test techniques and strategies for critical technologies. In particular, the Principal Directors for Critical Technologies and TRMC should be required to generate assessments of test approaches and requirements and advocate for needed resources. The development of the test approaches should be an explicit segment of technology development road-mapping activities.

This recommendation aims to augment activities conducted under the TRMC TEST program where S&T activities are conducted to position the T&E community for dealing with future technologies. With significant resources going into the emerging technologies themselves, benefits of coupling the development and testing communities will accrue.

This recommendation also calls for requiring DTE&A to develop guidance for acceptable test strategies for all critical technologies to ensure readiness for transition to formal programs. By formalizing approaches and policies in parallel with the technology development and prior to initiation of formal acquisition programs, a more efficient transition from technology development to acquisition can be realized.

Finally, this recommendation calls for activities to ensure warfighter inputs are early in technology development to accelerate transition to programs and subsequent operational testing. Often, the warfighter's inputs are not considered early enough in development and therefore do not sufficiently guide follow-on development.

The Task Force recommendations on T&E of emerging technologies will:

- increase the speed and efficiency of T&E for technologies as they transition to acquisition programs;
- provide a proper focus on development of the needed T&E methodologies for emerging new technologies; and
- provide proper focus on approaches needed to address actions that adversaries may take to counter or attack the emerging technologies.

4.2 Enablers for Realizing Strategic Shift in T&E

In the last section, five major findings and recommendations were identified regarding a strategic shift in DoD T&E needed to realize changes associated with speed and efficiency of system acquisition, development of mission-focused warfighting capabilities, and adaptation of emerging new technologies. To accomplish this strategic shift, the Task Force identified an additional four overarching findings and recommendations regarding enablers that can speed the desired strategic shift, which are discussed in the following subsections.

F&R 6: Improving T&E Using Digital Engineering

Finding 6.1. Digital engineering will be essential to augment and complement T&E but will not replace the need for live testing.

Finding 6.2. Increasing T&E efficiencies through digital engineering will require aligning T&E and M&S tools to well-defined interfaces and its use as the authoritative source of system data with continual evidence accrual.

Finding 6.3. Government program offices and the T&E community are often not prepared, educated, and resourced to operate in a fast-paced digital engineering environment.

Recommendation 6.1. **DTE&A** should publish best practices on use of digital engineering principles to fully realize the opportunity for increasing developmental testing efficiency and speed.

- Align system and subsystem interfaces to existing standards where available; facilitate development of critical new interface standards.
- Align M&S tools to system and subsystem interfaces to enhance testing strategies.
- Plan for structured evidence accrual during development and testing to validate performance.

Recommendation 6.2. **USD(A&S) should work through Service Acquisition Executives** to:

- educate, empower, and facilitate Program Offices to be fully prepared to execute T&E functions using digital engineering principles; and
- direct programs to develop and deliver M&S tools aligned to identified system and subsystem interfaces to enhance testing strategies.

Recommendation 6.3. **TRMC** should accelerate development of the infrastructure to enable seamless, multi-level secure data capability to facilitate evidence accrual across contract, DT, and OT environments to reduce redundancy and testing inefficiency.

The development of a DE environment, which incorporates M&S data in developing designs of complex systems, is the first enabler necessary for increasing the speed and efficiency of DoD T&E. A parallel DSB study[29] focused widely on the application of digital engineering for automating T&E. The

[29] "Digital Engineering Capability to Automate Testing and Evaluation," Defense Science Board, February 2024.

DSB DE Task Force investigated digital engineering from the perspective of its potential impact for increasing the speed and efficiency in T&E.

This Task Force found that digital engineering will be essential to augment and complement T&E, but it will not replace the need for live testing. Currently, this finding is embedded in Title 10, Section 4171(h), which states that the term operational test and evaluation "does not include an operational assessment based exclusively on (1) computer modeling; (2) simulation; or (3) an analysis of system requirements, engineering proposals, design specification, or any other information contained in program documents."

Digital tools are essential for efficiently conducting T&E under the following circumstances:

- When the dimensional space required for testing a component, subsystem, or system exceeds the capacity for physical testing due to schedule, cost, or resource constraints; or
- When the required physical operating environment cannot be produced in physical testing due to either to an inability to produce the relevant testing environment, lack of adequate physical representation adversary capabilities, unwillingness to expose system capabilities in live testing, or the complexity of integrated system being testing.

These circumstances arise for most complex systems to some degree, resulting in T&E being conducted using a combination of digital tools and physical testing. In these situations, physical testing is used to verify and validate requirements at sampled test conditions. As system dimensional complexity increases, it can be expected that the physical testing will become an increasingly small portion of the V&V testing. The use of digital tools will be used to fill this void in needed testing.

At present, there are two principal challenges associated with digital representation of systems for the purposes of T&E. The first challenge is due to aspects of real-world phenomena not being adequately captured in an accurate and efficient manner within digital tools used for testing. This challenge is further complicated by the lack of structured tracking of the uncertainties in component, subsystem, or system performance predictions. Note that this need extends beyond the approaches used for statistical sampling in Monte Carlo simulations and includes those needed to ensure all relevant sources of physical and operational uncertainties are captured. Formal uncertainty quantification is emerging in selected communities to serve as models for the tracking needed in T&E.[30],[31]

The second challenge with digital tools concerns the cost of high-fidelity representation of complex systems wherein the use of digital tools can be extremely expensive and exceed available budgets, schedules, and computational capabilities. In some cases, the costs of using digital tools exceeds the costs of equivalent live testing. This situation was encountered in the shock trials of the *USS Gerald Ford* Aircraft Carrier, where an M&S-based assessment was projected to cost approximately

---

[30] National Research Council. 2012. Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification. Washington, D.C.: The National Academies Press

[31] "Dakota: Bridging Advanced Scalable Uncertainty Quantification Algorithms with Production Deployment," Swiler, L.P., Eldred, M.S., and Adams, B.M., Handbook of Uncertainty Quantification, R. Ghanem et al (eds), Springer International Publishing, Switzerland 2015. DOI 10.1007/978-3-213-11259-6_52-1.

three times the cost of the physical test. Ultimately, live verification testing was conducted as shown in **Figure 25**.

Figure 25. Shock Trial of the USS Gerald Ford Aircraft Carrier.

Despite the challenges of the digital environment, the structured use of DE principles can speed the transition from S&T to acquisition programs, create seamless integration of continuous threat evaluations and continuous testing, and serve as the platform for structuring T&E of joint warfighting mission threads. As such, digital engineering can become the basis for many of the needed changes to DoD T&E. An example of using a model-based approach for driving test and evaluation is provided by Dunning et. al.,[32] which demonstrated automatic generation of test cases directly from a model-based representation of the system.

Realizing the desired increases in T&E efficiencies through digital engineering will require alignment of T&E and system M&S tools to well-defined interfaces, which is the foundation of Model-Based Systems Engineering (MBSE). With this alignment, requirements, digital tools, and testing approaches all converge to the common interfaces. DE environments provide the means for becoming the authoritative source of system data. T&E efficiencies result from a continual evidence accrual realized through testing and simulations. Continuous improvements in the digital tools proceeding in parallel with testing both drive performance uncertainties lower and improve confidence in those uncertainties carried forward into system integration activities.

During the Task Force data collection, reports were received of situations where government program offices and the T&E community were not prepared, educated, and resourced to operate in a

[32] Dunning, R., Matteson, W., Wise, R., and Sharpe, J., "Using a Model-Based Approach for Test and Evaluation," 2020 NDIA Ground Vehicle Systems Engineering and Technology Symposium, August 2020.

fast-paced DE environment. In this situation, inefficiencies resulted with challenges observed in formally tracking and reviewing T&E progress.

One exemplar of the tremendous advantages of employing digital tools for T&E is the Joint Simulation Environment (JSE) built for the F-35 program. The JSE has demonstrated the viability of a virtual-constructive environment for T&E of state-of-the-art aviation weapon systems. Advances in integrated, multi-domain warfighting capabilities are challenging the ability of militaries to rapidly design, develop, field, support, test, and train next generation warfighting capabilities. Today's physical test ranges are not sufficient to meet the testing and training needs of integrated weapon systems. Future test capabilities must be augmented with state-of-the-art M&S technologies to form a virtual range capable of meeting development, test, and training needs in a rapid and cost-effective manner. The synthetic test range must be able to produce the ability to test the kill chain across multiple domains, including air, space, sea, land, and cyber. The synthetic test range must address the multi-domain aspects of future systems and their associated warfighting capabilities. Additionally, the nature of a multi-domain, multi-national, complex array of sensors distributed across the air, space, and cyber domains presents a daunting test challenge.

JSE provides an unprecedented battlespace environment for enhanced test and training needs. The JSE, infrastructure, and threat environment are government-owned and available to support additional DoD test and training needs. JSE presents a multi-service opportunity to revolutionize high end, complex test and training.

JSE provides fifth generation operational test and high-end tactics training in the world's highest fidelity, highest density threat environment while not exposing any viable capabilities to our adversaries. Full fifth generation assessments are not possible on our open-air ranges. It is virtually impossible to scale testing encompassing a few live aircraft to accurately assess many theater-wide platform capabilities. Modern system-of-systems capability complexity makes open-air testing prohibitively expensive and logistically challenging. Realistic environments for high-end, multi-platform tactics training are severely limited. To provide the DoD's premier simulation environment for fifth generation (and beyond) test and training, JSE enables platform operational test and mission effectiveness assessments, tactics development in a high-density threat environment, large force training, high-fidelity, multi-platform warfighting capability assessments, and is used for new capability research and design.

JSE is a simulation environment comprised of a software battlespace that is highly extensible, modular, and builds on a solid foundation of existing DoD M&S technologies. It has a physical computing infrastructure that implements the battlespace. JSE has one or more ownerships (i.e., F-35, F-22) that constitute the system under test. It contains cockpits and visual display systems that provide the pilot interface, including planning/control/briefing rooms that facilitate mission execution. Lastly, it has an overarching facility that securely contains all the above and the manpower to operate it. JSE will now continue as both a T&E resource and training resource.

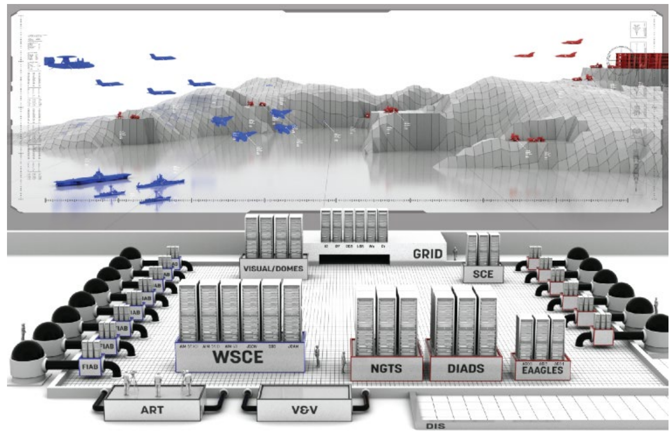


Figure 26. F-35 Joint Simulation Environment

JSE in a box is developed by the government and all the components are depicted in **Figure 26** (minus the F-35 IAB) and are deployed on a small footprint rack for use at other government and industry partner locations. This allows the government-owned environment to be used for platform integration well before arrival at a primary JSE site, reducing risk and timeline. F-35 in a box is developed by the original equipment manufacturer (OEM) and is referred to as blue platforms – virtual pilot-in-the-loop and rehosted Operational Flight Program (OFP) code from the actual platforms provides the highest fidelity blue platform and pilot experience. The F-35 eight-ship configuration is currently available for use. The goal is to develop a platform in a box for all current fighter/attack aircraft and all future aircraft including the B-21 and Next Generation Air Dominance (NGAD) initiative.

JSE has now met accreditation requirements for the F-35 IOT&E and has been formally accredited by the Air Force Operational Test and Evaluation Center (AFOTEC). The crux of using an M&S environment for test purposes relies upon its credibility with the decision-making test body. By its very nature, M&S generally utilizes assumptions to focus modelling efforts on features and functions within the system/environment believed to be critical in supporting the test objectives. This approach serves multiple purposes, including focusing limited dollars and effort on areas that are believed to be critical drivers in assessing test objectives while also balancing cost and risk in developing only as much functionality and fidelity as required to support a credible test activity. The JSE approach to V&V is to compare F-35 flight test results to those exact same missions flown in the synthetic environments.

National Air and Space Intelligence Center (NASIC), Missile and Space Intelligence Center (MSIC), and Office of Naval Intelligence (ONI) have been valuable partners with the Navy and Air Force in developing JSE. This team has been the lead government integrator, while the core components have been largely designed and built by government. The government owns and manages the interfaces and architectures. The architectures are open enough to accommodate many government and contractor systems and architectures, but still can entertain proprietary models through government-controlled interfaces and with enough understanding to accredit for intended use.

JSE is a unique facility for multiple reasons, including the fact that it was developed by multiple organizations within the government. The Navy, Air Force, and intelligence centers have worked seamlessly to create the most advanced simulation facility within DoD. Although primarily funded by the F-35 Joint Program Office (JPO), the JSE was designed to be F-35 agnostic. This means that other platforms can integrate into the JSE without design changes, which is already happening with F-22 and F-18, and more platforms are expected to follow. Government ownership also allows for the JSE architecture to be deployed across a wide variety of DoD M&S facilities.

Following the Task Force review of digital engineering and our deep dive on the JSE, three recommendations are provided.

Recommendation 6.1 states that DTE&A should publish best practices on use of DE principles to fully realize the opportunities for increasing DT efficiency and speed. In particular, guidance should be promulgated regarding alignment of system and subsystem interfaces to existing standards where available. Further DTE&A should advocate for and facilitate development of critical new interface standards where increases in speed and efficiency of development testing can be expected.

DTE&A should also promulgate guidance to development programs regarding alignment of digital tools to system and subsystem interfaces to enhance testing strategies and testing efficiencies. This guidance should address the need for a plan for structured evidence accrual during development and testing to verify requirements.

Recommendation 6.2 states that USD(A&S) should work through Service Acquisition Executives to educate, empower, and facilitate program offices to be fully prepared to execute T&E functions using DE principles and to direct programs to develop and deliver digital tools aligned to identified system and subsystem interfaces to enhance testing strategies. Contracts should be structured to ensure delivery of digital tools and contractor test results, which the government can then use for required independent testing. Programs should also ensure that program and test staff have the needed digital tools and environments available and have been trained in their use.

Title 10, Section 4401, states:

> "A major defense acquisition program that receives Milestone A or Milestone B approval after January 1, 2019, shall be designed and developed, to the maximum extent practical, with a modular open system approach to enable incremental development and enhance competition, innovation, and interoperability."

Alignment of testing and evaluation requirements to the interfaces developed in these modular open systems will lead to significant increases in T&E efficiencies through development of standardized test equipment and automation of testing.

Recommendation 6.3 states that TRMC should accelerate development of the infrastructure to enable a seamless, multi-level, secure data capability to facilitate evidence accrual across contract, DT, and OT environments to reduce redundancy and testing inefficiencies. While individual programs can build and sustain their separate digital environment, DoD will need the infrastructure to integrate the digital representations of individual systems as joint warfighting mission threads are developed and evaluated.

F&R 7: Modernizing the T&E Infrastructure

Finding 7.1. The DoD test infrastructure will need to be modernized to meet the future needs for T&E.

Finding 7.2. Promising advancements are being made in modernization of infrastructure through certification of autonomous technologies with transition to autonomous collection platforms including space-based telemetry collection.

Finding 7.3. Development of a modern test data collection and distribution system augmented with AI-driven automated test procedures and analysis tools will be required to meet the future need for T&E in complex and dynamic environments.

- Robust approaches for multi-level security operations will be essential.

Recommendations 7.1. **USD(R&E)** and **USD(A&S)** should direct a comprehensive re-evaluation of T&E infrastructure needs to support future needs, including:

- TRMC should accelerate transition to using autonomous platforms for data collection fully leveraging commercial advancements in autonomous vehicles and sensing.
- TRMC should engage with the Chief Digital and Artificial Intelligence Officer (CDAO) to accelerate development of the TRMC knowledge management system.

Recommendations 7.2. **USD(R&E)** should explore approaches to increase priority of MILCON appropriation for the T&E needs.

The Terms of Reference requested the Task Force assess the research, development, test, and evaluation infrastructure master plan required under Section 252 of the NDAA for FY 2020. Prior to discussing the Task Force findings and recommendations regarding infrastructure, it is necessary to evaluate the master plan.[33]

The major elements of the DoD RDT&E Infrastructure Master Plan include:

- infrastructure deficiencies;
- existing and emerging RDT&E mission areas for modernization investments;
- specific infrastructure projects;
- DoD plans for execution of infrastructure projects; and
- regulatory and policy barriers to implementing the master plan.

The Task Force reviewed and assessed these elements and concluded that the legislative requirements within Section 252 of the FY 2022 NDAA were satisfied. The Task Force did observe that a significant test infrastructure backlog exists with approximately $6.5B in unfunded military construction (MILCON) projects. The MILCON funding for the MRTFB over the past two decades is shown in **Figure 27**. Funding for the MRTFB has averaged approximately 0.75% of total MILCON spending over the 2000-2023 period but has risen to approximately 1.5% over recent years with annual funding of slightly more than $100M. Simple arithmetic reveals that a $6.5B backlog would

[33] "Fiscal Year 2021 National Defense Authorization Act Section 252 Infrastructure to Support Research, Development, Test, and Evaluation Missions," Pham, P. K., Director, Infrastructure and Technology Transfer, OUSD(R&E), June 2021.

require 65 years to clear at a spending rate of $100M/year.

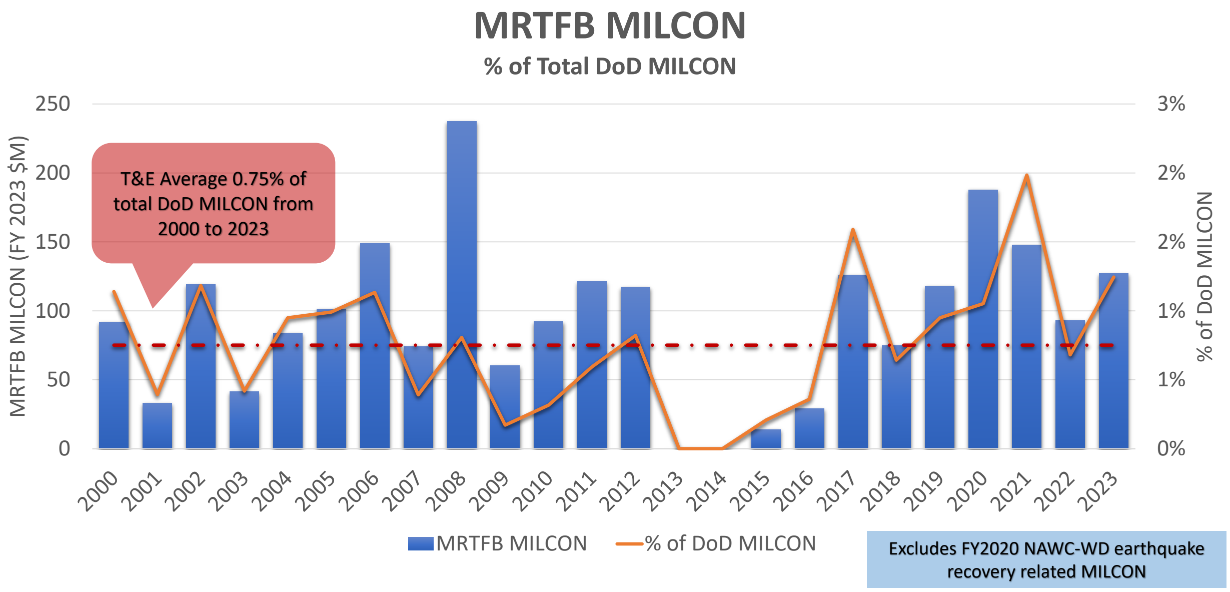


Figure 27. MRTFB Infrastructure Funding 2022-2023.

The Task Force found that a modernized DoD test infrastructure is an enabler to realize the strategic shift in T&E called for by this Task Force. Given the challenge with funding the backlog in MRTFB infrastructure projects, a different approach to T&E is needed.

As discussed earlier, a transition is already underway regarding increased use of digital environments for testing, which is in turn reducing reliance on physical testing. In parallel, promising advancements are also being made in the modernization of physical testing infrastructure through certification of autonomous technologies and a transition to autonomous collection platforms, including space-based telemetry collection.

A modernized test infrastructure should heavily leverage a modern test data collection and distribution system augmented with AI-driven automated test planning and execution procedures and analysis tools. This development environment is evolving as a commercial best practice and will be required to meet the future need for DoD T&E in complex and dynamic environments. A key feature of this modernized infrastructure is a robust approach for multi-level security operations, which is viewed as essential for T&E of complex systems of today and in the future.

A modernized T&E infrastructure reinforces recommendations from the 2021 report of the National Academies of Sciences, Engineering, and Medicine (NASEM) that addressed DoD range capabilities needed to ensure operation superiority of U.S. defense systems.[34] The NASEM report's recommendations included: 1) develop the "range of the future" to test complete kill chains in joint all-domain operation environments; 2) restructure the range capability requirements process for

[34] National Academies of Sciences, Engineering, and Medicine, *Necessary DOD Range Capabilities to Ensure Operational Superiority of U.S. Defense Systems Testing for the Future Fight* (Washington, D.C: National Academies Press, 2021) https://doi.org/10.17226/26181.

continuous modernization and sustainment; 3) bootstrap a new range operating system for ubiquitous M&S throughout the weapon system development and test life cycle; and 4) create the "TestDevOps" digital infrastructure for future operational and seamless range enterprise interoperability.

Recommendation 7.1 states that USD(R&E) and USD(A&S) should direct a comprehensive reevaluation of T&E infrastructure to support future needs. This reevaluation includes having TRMC accelerate the transition to using autonomous platforms for data collection, fully leveraging commercial advancements in autonomous vehicles and sensing. TRMC has initiated work in this direction with systems such as RangeHawk, but a comprehensive assessment of emerging commercial capabilities that could be leveraged is warranted.

TRMC should also engage with the CDAO to accelerate development of the TRMC knowledge management system. Bringing together test data to facilitate application of advanced analytical tools and enable system-of-systems evaluations will require this capability.

Recommendation 7.2 states that USD(R&E) should explore approaches to increase priority of MILCON appropriation for T&E needs. Recognizing that funding T&E infrastructure can fall short in MILCON prioritization discussions, often due to a lack of understanding of the underlying challenges and risks being accrued, this recommendation aims at the exploration of and advocacy for additional funding pathways that would address this essential need.

F&R 8: Ensuring a Trained T&E Workforce

Finding 8. Today's T&E workforce will be challenged in responding to the needed strategic shift in T&E.

- The pace of rapidly advancing technologies will require a flexible, adaptive, and educated T&E workforce.
- The T&E workforce, like much of the DoD enterprise, is challenged in maintaining its workforce due to hiring and retention problems.
- MRTFB operators and maintainers are not considered part of the T&E Acquisition Workforce, so they are ineligible for certifications/credentials.

Recommendation 8. **USD(A&S)** and **USD(R&E)** should develop a comprehensive approach to developing the T&E workforce for the future that includes the following elements:

- Maximizes utilization of available life-long and credentialed courses in emerging technologies at both Defense Acquisition University (DAU) and major research universities.
- Provides experiential learning opportunities through rotational assignments in program development offices and industry while leveraging existing DoD schoolhouses such as Test Pilot Schools.
- Develops a career progression for credentialing MRTFB operators and maintainers.
- Increases the talent pipeline through enhancement of outreach pathways to include underrepresented communities.

The Task Force found that the T&E workforce will be challenged in responding to the needed strategic shift in T&E; thus, ensuring a trained T&E workforce is a critical enabler of the future T&E enterprise. The challenges encountered include:

- keeping pace with rapidly advancing emerging technologies;
- effectively using an information technology (IT) landscape that is continuously evolving providing capability to collect, store, and analyze ever-increasing volumes of data;
- an efficiency to be gained by the accelerating adaptation of commercial technology to DoD T&E to include sensing and communication technologies; and
- a need for increasing the pace of T&E through the application of advanced AI/ML techniques.

These are exacerbated by other widespread challenges such as:

- high staff turnover with the T&E workforce, like much of the DoD enterprise, challenged in maintaining its workforce due to hiring and retention problems; and
- challenges with effective training of the T&E community. For example, MRTFB operators and maintainers are not considered part of the T&E Acquisition Workforce and are therefore ineligible for coursework towards certifications and credentials.

A major revision to the current workforce development was formally announced by USD(A&S) in September 2020 as the “Back-to-Basics” (BtB) initiative that modernized the Department’s implementation of the Defense Acquisition Workforce Improvement Act (DAWIA) to emphasize career-long learning. The BtB goal reduced required certification training to a functional area. Elective defense acquisition credentials supplemented certification training to ensure a technically component workforce.

USD(R&E) is the functional leader for T&E training, and DAU has developed T&E courses to include both foundational and practitioner courses (e.g., Fundamentals of T&E and T&E for Practitioners). DAU is in the process of developing additional courses to support T&E needs such as T&E for Software, T&E of AI, T&E of Autonomous Systems, and Cyber T&E.

In addressing the observed workforce challenges and the process for ensuring a trained workforce, the Task Force identified a series of steps that should be taken to ensure a well-trained workforce consistent with the needed strategic shift in T&E.

Recommendation 8 states that USD(A&S) and USD(R&E) should develop a comprehensive approach to developing the T&E workforce for the future. First, this approach should maximize the use of available life-long and credentialed courses in emerging technologies at both the DAU and major research universities. DAU courses are being developed to address emerging technologies, but available funding limits the pace of development. This approach should be augmented using the courses major U.S. research universities are developing and making available in virtual environments across a wide range of emerging technologies. As an example, the USD(R&E) Joint Hypersonic Transition Office has facilitated a compilation of available university courses produced under the University Consortium for Applied Hypersonics.

To further provide experiential learning opportunities, the comprehensive approach to training should develop options for rotational assignments in concert with program development offices, laboratories, and industry. For example, existing DoD schoolhouses such as Test Pilot Schools should be explored for providing training in applied T&E.

The comprehensive approach should also address the workforce that is often overlooked in a talent development pipeline. For example, a career progression and credentialing for MRTFB operators and maintainers should be developed, as this important segment of the T&E workforce also needs to adapt to future roles and technologies.

Finally, the comprehensive approach should address approaches to increasing the talent pipeline through enhancement of outreach pathways, including underrepresented communities. One pathway to increase the T&E talent pipeline is to provide more T&E talent development opportunities within the DoD Science, Mathematics, and Research for Transformation (SMART) scholarship program for science, technology, engineering, and mathematics (STEM) students. Another pathway is for DoD to develop a formal partnership with organizations that provide undergraduate or post graduate scholarships with internship opportunities to underrepresented students in engineering and computer science, such as the National Action Council for Minorities in Engineering (NACME) and the Consortium of Graduate Degrees for Minorities in Engineering (GEM). Furthermore, with many of the test facilities and ranges located far from major urban centers, there is an opportunity to tap into local communities to hire and develop technicians and administrative support with specialized training.

F&R 9: OSD T&E Oversight Organization

Finding 9.1. An increase in the speed and efficiency of T&E is unlikely to be realized with any evaluated options for realigning OSD T&E oversight, and significant challenges are likely to be generated if pursued.

Finding 9.2. To manage the strategic shift needed for future T&E, timely and effective collaboration will be required across USD(A&S), USD(R&E), and DOT&E.

Recommendation 9.1. The existing OSD T&E oversight roles and organizational structure should be maintained.

Recommendation 9.2. **USD(R&E) should direct DTE&A** to provide guidance on DT requirements to enable efficient software development, rapid acquisition of emerging technologies, and joint mission threads.

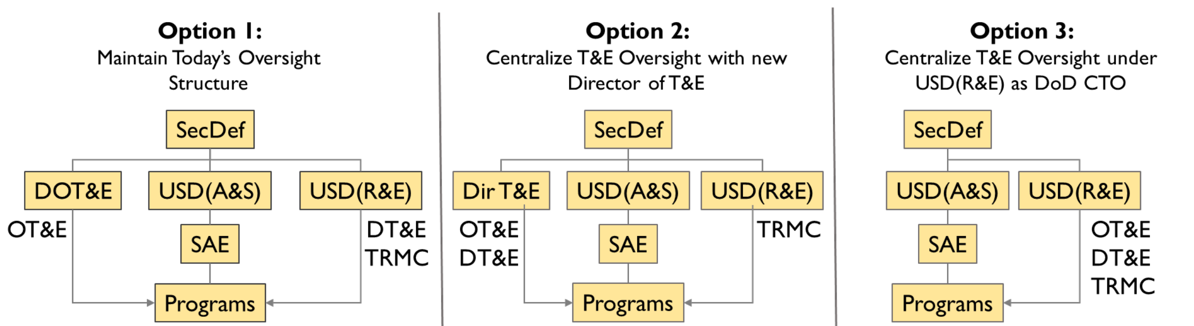
Recommendation 9.3. **DOT&E** should ensure early and timely definition of OT requirements to support effective dynamic mission threads, transition of emerging new technologies, and seamless transition of OT event into training.

The ToR for this study requested that the Task Force:

- Assess test and evaluation oversight organizations within OSD, including their authorities, responsibilities, activities, resources, and effectiveness, including with respect to acquisition programs of the military departments and defense agencies.

- Develop and assess potential courses of action to improve the effectiveness of oversight of developmental testing, operational testing, and integrated testing activities, and test and evaluation resources within OSD, including as one such course of action establishing a single integrated office with such responsibilities.

Prior to addressing specific findings and recommendations, The Task Force assessed options for OSD T&E oversight organizations. As presented in Section 2.2, the T&E oversight function is divided between DOT&E and USD(R&E), with the DTE&A and TRMC organizations under OUSD(R&E). The Task Force evaluated several other organization options in addition to the current structure, eventually consolidating to three options, shown in simplified form in **Figure 28.** Each of the three options are discussed below and on the following pages.



	Option 1	Option 2	Option 3
Advantages	• No changes in authorities needed • Maintains independence of DT&E and OT&E oversight roles	• Centralized point-of-contact for OT/DT oversight • Direct input to SecDef on prioritization of T&E needs	• Integrated T&E and infrastructure oversight • Integrated with work on emerging capabilities and Joint Mission Thread development

Disadvantages	• DOT&E organizationally separated from
	  • Emerging technology development
	  • Joint Mission Thread development
		• Independent OT&E and DT&E assessment roles lost	• Requires changes in authorities
			• Independence of DOT&E lost

Figure 28. Three Organizational Options for T&E Oversight Including Strengths and Weaknesses

Option 1: Maintain Today's Oversight Structure. Option 1 maintains today's organizational structure with DOT&E responsible for OT guidance and oversight, USD(R&E) responsible for DTE&A and TRMC functions, and program oversight flowing through USD(A&S) down through Service Acquisition Executives to individual programs.

The advantage of Option 1 is that no changes in authorities are required, and this option maintains the independence of the OT oversight role in assessing the operational readiness of new system capabilities.

It also has the advantage of clearly recognizing a difference between the DT and OT roles. DT is inherently an integral part of the system development and therefore intimately tied to the program development activities and testing, while OT is more closely tied to a grading of readiness for production and deployment.

The disadvantage of this option is that DOT&E is organizationally separated from the emerging technology development and joint warfighting mission thread development, which are largely being coordinated through USD(R&E) activities. As discussed earlier in this report, these two functions are important drivers in the needed strategic changes to DoD T&E.

Some may also argue that oversight guidance regarding T&E comes from two different OSD organizations (DOT&E and DTE&A), which may be an organizational disadvantage. The Task Force felt that this separation was not a true disadvantage given the separate functions of DT and OT as described throughout this report.

Option 2: Centralize T&E Oversight with New Director of T&E. This option creates a new organization under a Director for T&E that would report directly to the Secretary of Defense. Within this new organization, the OT&E and DT&E functions would be consolidated. DTE&A would transfer from USD(R&E), but TRMC would remain within USD(R&E). Legislature changes would be required to disestablish DOT&E and its roles and re-establish the new roles and responsibilities of the Director for T&E.

Two advantages of this option are the creation of a centralized point-of-contact for OT/DT oversight and direct input to the SecDef on prioritization of T&E needs. Advocates for this option believe that unifying DT and OT in a single organization would ensure more comprehensive and consistent guidance from OSD governing testing throughout the Department. Additionally, input to the SecDef would provide authoritative counsel concerning system development maturity, as well as system operational effectiveness and suitability. This authoritative counsel is not achieved through today's reporting by DOT&E and the Service Secretaries and Chiefs.

A disadvantage of Option 2 is that independent OT&E and DT&E roles are lost when brought under one organization. As discussed as an advantage of Option 1, separating OT&E and DT&E have advantages as DTE&A requires a closer coordination with development programs as the system is being matured, compared to the grading assessment that occurs with OT&E. Independent OT evaluations have been a concern of the Congress since DOT&E was established in 1983, and bringing DTE&A into the same organizational structure could degrade the close working arrangements between the DTE&A function and programs.

Option 2 also maintains the disadvantages of Option 1 in that it may create a disconnect between Department activities associated with the development of emerging technologies and joint warfighting mission threads.

Option 3: Centralize T&E Oversight under USD(R&E) as DoD CTO. Under Option 3, the function of OT&E is transferred to USD(R&E) under the CTO role to enable a single organization encompassing the three major elements of T&E oversight. Note that this option is the only one of the three options evaluated that satisfies the ToR request to identify a "course of action establishing a single integrated office."

The advantages of this option are that it integrates T&E and infrastructure oversight roles into a single organization, which is also integrated with work on emerging technologies and joint mission thread development. With these areas representing the bulk of the emerging challenges with T&E, this option offers the opportunity to close potential gaps as new T&E needs emerge.

The disadvantage of Option 3 is that it requires changes in authorities now provided to an independent DOT&E, sacrificing the independent OT&E function. As such, the disadvantages of Option 2 are also seen with this option.

Following the identification and evaluation of the three organizational options described, the Task Force found that an increase in the speed and efficiency of T&E is unlikely to be realized with any of the evaluated options that require realigning OSD T&E oversight. Further, the Task Force found that significant challenges are likely to be generated if reorganization options are pursued. The biggest challenge with reorganizational changes would be an anticipated breakdown of true collaboration between program development and DTE&A.

The Task Force found that to manage the strategic shift needed for future T&E, timely and effective collaboration will be required across USD(A&S), USD(R&E), and DOT&E. Evidence that this collaboration can occur is seen in the 2023 DOT&E *Strategy Implementation Plan,*[35] which focuses on five pillars: test the way we fight; accelerate the delivery of weapons that work; improve the survivability of DoD in a contested environment; pioneer T&E of weapon systems built over time; and foster an agile and enduring T&E enterprise workforce. This Strategy Implementation Plan was endorsed in writing by the USD(R&E), Hon. Heidi Shyu, USD(A&S), Hon. William LaPlante, Secretary of the Army, Hon. Christine Wormuth, Secretary of the Navy, Hon. Carlos Del Toro, and Secretary of the Air Force, Hon. Frank Kendall. The Task Force fully endorses this collaboration using today's organization structure.

The Task Force developed three recommendations regarding the OSD organization oversight, as follows.

Recommendation 9.1 states that existing OSD T&E oversight roles and organizational structure should be maintained for the reasons outlined above. The existing organizational structure provides the flexibility required for the needed strategic shift in T&E. Further, the disadvantages of pursuing this approach are viewed as manageable through collaboration, as evidenced in the DOT&E *Strategy Implementation Plan*. The alternative options were viewed as presenting new challenges that would be difficult to overcome while also attempting the needed strategic shift in T&E.

Recommendation 9.2 states that USD(R&E) should direct DTE&A to provide guidance on DT requirements to enable efficient software development, rapid acquisition of emerging technologies, and joint mission threads. Even within today's organizational structure, further roles and responsibilities need to be defined. This recommendation addresses the need to close gaps in development of software-intensive systems, emerging technologies, and joint mission threads and the DT required to verify systems. Note that this recommendation is provided in recognition that the DTE&A function needs to expand beyond that defined entirely by the framework for acquisition systems.

Recommendation 9.3 states that DOT&E should ensure early and timely definition of OT requirements to support effective dynamic mission threads, transition of emerging new technologies, and seamless transition of OT event into training. This recommendation is provided in recognition that OT&E needs to also expand beyond the roles defined by the framework of acquisition systems. OT&E considerations need to be leveraged as early as possible as new capabilities emerge.

[35] "DOT&E Strategy Implementation Plan – 2023," HON Nickolas Guertin, Director, Operational Test and Evaluation, April 2023.

5. Conclusion

T&E is widely recognized as an essential component of a structured systems engineering process that has proven to be capable of producing military systems with sufficient confidence to be operated in complex environments. However, today's approaches to T&E are being challenged by rapidly increasing adversary capabilities, emerging new technologies, and significant emerging commercial capabilities. These challenges are amplified when dealing with the acquisition of software-intensive weapon systems and the development of joint warfighting mission threads that require a system-of-systems integration, especially when the integration crosses Military Service acquisition boundaries with programs in differing states of maturity. Further, emerging new technologies provide unique challenges as these technologies stress today's approaches to T&E.

To adequately respond to current challenges, a strategic shift in DoD's approach to T&E is needed. Five findings and recommendations are provided concerning the needed shift in expanding T&E beyond the acquisition-based framework of today, and moving to a continuous process of development, testing, and assessment of adversary capabilities. Further recommendations are provided regarding software-intensive system development and approaches to linking advanced technology development and testing strategies.

The Task Force identified four enablers to achieving the needed strategic shift in T&E. These enablers include improving the use of digital engineering for T&E, modernizing the T&E infrastructure, ensuring a trained T&E workforce, and refining the DoD organizational oversight of the T&E enterprise.

T&E will continue to be an essential element of system development as the future unfolds. As the pace of development increases and acquisition transitions from being platform focused to mission focused, T&E will need to evolve to keep pace. The findings and recommendations provided herein represent a starting point to that needed evolution.

THIS PAGE LEFT INTENTIONALLY BLANK

Appendix A. Terms of Reference

UNDER SECRETARY OF DEFENSE
3030 DEFENSE PENTAGON
WASHINGTON, DC 20301-3000

RESEARCH AND ENGINEERING

10 JUL 2022

MEMORANDUM FOR CHAIRMAN, DEFENSE SCIENCE BOARD

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

The 2022 National Defense Authorization Act (NDAA) legislated that the Secretary of Defense direct the Defense Science Board (DSB) to "carry out a study on the resources and capabilities of the test and evaluation organizations, facilities, and laboratories of the Department of Defense."

As such, I am tasking the DSB—through the establishment of the Task Force on Test and Evaluation ("the Task Force")—to carry out the following tasks:

- Assess the effectiveness of current developmental testing, operational testing, and integrated testing within the Department of Defense (DoD) in meeting statutory objectives and the test and evaluation requirements of the Adaptive Acquisition Framework;
- Identify industry and government best practices for conducting developmental testing, operational testing, and integrated testing;
- Determine potential applicability of industry and government best practices for conducting developmental testing, operational testing, and integrated testing within the Department to improve test and evaluation outcomes;
- Identify duplication of efforts and other non- or low-value added activities that reduce speed and effectiveness of test and evaluation activities;
- Assess test and evaluation oversight organizations within the Office of the Secretary of Defense (OSD), including their authorities, responsibilities, activities, resources, and effectiveness, including with respect to acquisition programs of the military departments and Defense Agencies;
- Assess the research, development, test, and evaluation infrastructure master plan required under section 252 of the NDAA for Fiscal Year 2020;
- Develop and assess potential courses of action to improve the effectiveness of oversight of developmental testing, operational testing, and integrated testing activities, and test and evaluation resources within the OSD, including as one such course of action establishing a single integrated office with such responsibilities, and;

- Develop such recommendations as the DSB may have for legislative changes, authorities, organizational realignments, and administrative actions to improve test and evaluation oversight and capabilities, and facilitate better test and evaluation outcomes.

The Task Force findings, observations, and recommendations will be presented to the full DSB for its thorough, open discussion and deliberation at a properly noticed and public meeting subject to Government in the Sunshine Act exemptions. The DSB will provide its findings and recommendations to the Under Secretary of Defense for Research and Engineering as the Sponsor of the DSB. The nominal start date of the study period will be within 30 days of the initial appointment of Task Force members. In no event will the duration of the Task Force exceed 12 months from the start date.

The Task Force members are granted access to those DoD officials and data necessary for the appropriate conduct of their activities. As such, the Office of the Secretary of Defense and Component Heads are requested to cooperate and promptly facilitate requests by DSB staff regarding access to relevant personnel and information deemed necessary, as directed by paragraphs 5.1.8. and 5.3.4. of DoD Instruction 5105.04, "Department of Defense Federal Advisory Committee Management Program," and in conformance with applicable security classifications.

The DSB and the Task Force will operate in accordance with the provisions of the Federal Advisory Committee Act (5 United States Code (U.S.C.), Appendix), the Government in the Sunshine Act (5 U.S.C. §552b), and other applicable federal statutes, regulations, and policy. Individual DSB and Task Force members do not have the authority to make decisions or recommendations on behalf of the DSB nor report directly to any Federal representative. The members of the Task Force and the DSB are subject to certain Federal ethics laws governing conflicts of interest, including 18 U.S.C. § 208, and the Standards of Ethical Conduct regulations in 5 Code of Federal Regulations (CFR) Part 2635.

Heidi Shyu

2

Appendix B. DSB Membership

Dr. Eric Evans, Chair

Mr. Michael Appelbaum

Dr. Jennifer Bernhard

Dr. Alison Brown

Dr. Kimberly Budil

Mr. James Carlini

Dr. Tomás Díaz de la Rubia

Mr. Fred Dixon

Adm William Fallon, USN (ret.)

Ms. Laetitia de Cayeux

Mr. Robert Giesler

Dr. Johney Green

Dr. Robert Grossman

Dr. Daniel Hastings

Dr. Ayanna Howard

Dr. Evelyn Hu

Hon. Shirley Ann Jackson

Dr. Ashanti Johnson

Dr. Paul Kaminski

Dr. Ann Karagozian

Dr. John Manferdelli

Dr. Katherine McGrady

Dr. James Miller

Dr. DJ Patil

Dr. Gary Polansky

Dr. Sanjay Raman

Dr. David Relman

Gen Paul Selva, USAF (ret.)

Dr. Nashlie Sephus

Dr. Reshma Shetty

Dr. Alfred Spector

Dr. Vincent Tang

Dr. Dorota Temple

Dr. Jan Tighe

Dr. Bradford Tousley

Dr. David Van Wie

Ms. Mandy Vaughn

Dr. Dinesh Verma

Dr. Steven Walker

Dr. Robert Wisnieff

Appendix C. Task Force Membership

Task Force Co-Chairs

Dr. Johney Green
Dr. David Van Wie

Task Force Members

Dr. Panos Datskos
ADM William Fallon
Mr. Ed Greer
Hon. Shirley Jackson
Dr. Ann Karagozian
Mr. Steve Lewia
Ms. Heidi Perry
Dr. Gary Polansky
Dr. Nashlie Sephus
Dr. Robert Wisnieff

Executive Secretary

Mr. Chris Collins, DTE&A (OUSD(R&E))

Government Advisors

Mr. Orlando Flores, DTE&A (OUSD(R&E))
Mr. Robert Butterworth, DOT&E

DSB Secretariat

Ms. Betsy Kowalski, Designated Federal Officer (DFO)
Dr. Troy Techau, Alternate DFO
Mr. Kevin Doxey, DFO (former)

Support Staff

Ms. Hannah Gonzalez (SAIC)
Mr. Paul Normolle (SAIC)
Ms. Brenda Poole (SAIC)

Appendix D. Summary of Findings and Recommendations

	Findings	Recommendations
Strategic Shift in T&E		

1
Joint Warfighting Mission Threads
	• Rapidly emerging technologies, increasingly capable adversaries, and joint warfighting mission threads demand a significant strategic shift in the focus of T&E to augment today's approach of acquisition-based T&E.	• **USD(A&S)**, **USD(R&E)** and **DOT&E** should expand T&E activities to address needed required strategic shift.
		  • Shift T&E to include S&T development leading into programs.
		    • **USD(R&E)** and **USD(A&S)** should revise Technology Readiness Levels and standards for Technology Readiness Assessment to set assessment to include data, software, M&S tools, and test results readiness for transition to acquisition programs.
		  • Incorporate warfighter input early and use operational testing to inform and seamlessly transition into training.
		    • **USD(A&S) should direct Service Acquisition Executives** to create plans to fully leverage OT events for informing system training.
		  • Ensure robust testing of joint warfighting mission threads.
		    • **USD(R&E) should direct the DASD for Multi-Domain Joint Operations** to ensure sufficient system-of-systems mission engineering is conducted to form the basis of detailed integrated mission testing strategies.
		    • **USD(R&E) should direct DTE&A** to generate guidance for DT of all defined joint warfighting mission threads.
		    • **DOT&E** should ensure operational robustness of joint warfighting mission threads using live and virtual environments.
		  • Develop capability for continuous data and user feedback from training and operations.
		    • **USD(A&S) should direct Service Acquisition Executives** to enable collection of system data from training venues and operational events together with user feedback to inform continual system improvement.

2 **Continuous Threat Assessment**	• Rapid adversary capability development is resulting in disconnects between threat assessments used to define program requirements and the threats used for operational assessment, often conducted years later.	• **Service Acquisition Executives should direct Program Executive Officers** to develop, maintain, and fund threat assessment teams knowledgeable of existing and developing systems to coordinate with the Service and Combatant Command intelligence components, Defense Intelligence Agency, and other Intelligence Community elements to conduct continuous evaluation of emerging threats.

	Findings	Recommendations
	• Many complex weapon system lifetimes will exceed the knowledge of current and projected threat capabilities and a physics-based threat assessment can be used to allow consideration of future threats as technologies evolve. • A deep tie between the system developers, intelligence community, and the DoD threat assessment community is critical to ensure intelligence is relevant and adversary advancements are reflected in system assessments. • DoD threat assessment community must understand both intelligence and state of capability development.	• **USD(R&E) should direct ASD(Mission Capabilities)** to coordinate with the Strategic Intelligence and Analysis office to provide continuous assessments of emerging threats to joint warfighting mission threads

3
Continuous Testing
	• Continuous testing offers pathway for improving T&E efficiency through continual evidence accrual.
	  • A structured digital engineering framework is essential for aligning development, simulation, and testing activities.
	• Significant automated testing will be required to cover operational "envelop" of complex systems.
		• **USD(A&S) should direct Service Acquisition Executives** to structure new programs to:
		  • incorporate testability requirements in components, subsystems, and systems to speed evidence accrual;
		  • maximize use of automation to increase testing for systems and subsystems; and
		  • develop approaches to report system status and data to enable feedback for improving system performance.
		• **USD(R&E) should direct DTE&A** to develop and promulgate DT guidance to ensure system capability to use automated developmental testing to the maximum extent possible.
		• **USD(A&S)** should develop a repository for collection and exploitation of system data collected after delivery to training and operational units.

4
Software Intensive Systems
	• For software being delivered to weapon systems, today's milestone-based T&E inhibits intended use of software designed pathways.
	  • At present, no weapon system program is being executed using with the Software Acquisition Pathway.
	  • An execution inconsistency results in the Software Acquisition pathway instructions regarding continuous cybersecurity testing and mandated requirement to use the 177 IWS for adversarial assessments.
		• **USD(R&E)** should establish development guidelines for software-intensive systems to be built using an architecture that allows segmentation of the software stack (i.e., application, middleware, operating system interface, firmware, hardware) to enable testability of all layers.
		• **USD(A&S)** should direct new software-intensive system development to segment safety critical functions from non-safety critical functions and execute test-driven development of non-safety critical system development using the Software Acquisition Pathway.

5 **Emerging Technologies**	• For emerging technologies, close coupling of the technology development with its testing approach is a key enabler for increasing the speed and efficiency of T&E.	• **USD(R&E)** should ensure timely development of test techniques and strategies for critical technologies.

	Findings	Recommendations
	• Testing approaches will need to advance and mature in parallel with the technology itself. • Consideration of existing and potential adversary countermeasures will also need to mature in parallel with the technology itself.	• Require the Principal Directors for Critical Technologies and TRMC to generate assessments of test approaches and requirements and advocate for needed resources. • Require the DTE&A to develop guidance for acceptable test strategies for all critical technologies to ensure readiness for transition to formal programs. • Ensure warfighter inputs into technology development to accelerate transition to programs and subsequent operational testing.
Enablers to Future T&E		

6
Digital Engineering
	• Digital engineering will be essential to augment and complement T&E but will not replace the need for live testing.
	• Increasing T&E efficiencies through digital engineering will require aligning T&E and M&S tools to well-defined interfaces and its use as the authoritative source of system data with continual evidence accrual.
	• Government program offices and the T&E community are often not prepared, educated, and resourced to operate in a fast-paced digital engineering environment.
		• **DTE&A** should publish best practices on use of digital engineering principles to fully realize opportunity for increasing developmental testing efficiency and speed.
		  • Align system and subsystem interfaces to existing standards where available; facilitate development of critical new interface standards.
		  • Align M&S tools to system and subsystem interfaces to enhance testing strategies.
		  • Plan for structured evidence accrual during development and testing to validate performance.
		• **USD(A&S) should work through Service Acquisition Executives** to:
		  • educate, empower, and facilitate Program Offices to be fully prepared to execute T&E functions using digital engineering principles; and
		  • direct programs to develop and deliver M&S tools be aligned to identified system and subsystem interfaces to enhance testing strategies.
		• **TRMC** should accelerate development of the infrastructure to enable seamless, multi-level secure data capability to facilitate evidence accrual across contract, DT, and OT environments to reduce redundancy and testing inefficiency.

7
Infrastructure
	• The DoD test infrastructure will need to be modernized to meet the future needs for T&E,
	• Promising advancements are being made in modernization of infrastructure through certification of autonomous technologies with transition to autonomous collection platforms including space-based telemetry collection.
	• Development of a modern test data collection and distribution system augmented with AI-driven automated test procedures and analysis tools will be required to meet
		• **USD(R&E)** and **USD(A&S)** should direct a comprehensive re-evaluation of T&E infrastructure needs to support future needs including:
		  • TRMC should accelerate transition to using autonomous platforms for data collection fully leveraging commercial advancements in autonomous vehicles and sensing.
		  • TRMC should engage with the Chief Digital and Artificial Intelligence Officer (CDAO) to accelerate development of the TRMC knowledge management system.

	Findings	Recommendations
	the future need for T&E in complex and dynamic environments. • Robust approaches for multi-level security operations will be essential.	• **USD(R&E)** should explore approaches to increase priority of MILCON appropriation for the T&E needs
8 **Workforce**	• Today's T&E workforce will be challenged in responding to the needed strategic shift in T&E. • The pace of rapidly advancing technologies will require a flexible, adaptive, and educated T&E workforce. • The T&E workforce, like much of the DoD enterprise, is challenged in maintaining its workforce due to hiring and retention problems. • MRTFB operators and maintainers are not considered part of the T&E Acquisition Workforce, so they are ineligible for certifications/credentials.	• **USD(A&S)** and **USD(R&E)** should develop a comprehensive approach to developing the T&E workforce for the future that includes the following elements: • Maximizes utilization of available life-long and credentialed courses in emerging technologies at both DAU and major research universities. • Provides experiential learning opportunities through rotational assignments in program development offices and industry while leveraging existing DoD schoolhouses such as Test Pilot Schools. • Develops a career progression for credentialing MRTFB operators and maintainers. • Increases talent pipeline through enhancement of outreach pathways to include underrepresented communities.
9 **OSD Oversight Organization**	• An increase in the speed and efficiency of T&E is unlikely to be realized with any evaluated options for realigning OSD T&E Oversight, and significant challenges are likely to be generated if pursued. • To manage the strategic shift needed for future T&E, timely and effective collaboration will be required across USD(A&S), USD(R&E), and DOT&E.	• The existing OSD T&E Oversight roles and organizational structure should be maintained. • **USD(R&E) should direct DTE&A** to provide guidance on DT requirements to enable efficient software development, rapid acquisition of emerging technologies, and joint mission threads. • **DOT&E** should ensure early and timely definition of OT requirements to support effective dynamic mission threads, transition of emerging new technologies, and seamless transition of OT event into training.

Appendix E. Briefings Received

Meeting 1 (16 Aug 2022)

Perspective on Developmental Testing
Developmental Test, Evaluation, and Assessments, OUSD(R&E)

Perspective from the former Chief Scientist of DOT&E
Office of the Director, Operational Test & Evaluation (DOT&E), OUSD(R&E)

Perspective on Operational Testing
DOT&E, OUSD(R&E)

Perspective from TRMC
Test Resource Management Center (TRMC), OUSD(R&E)

Meeting 2 (17 Oct 2022)

Scientific Test and Analysis Techniques Center of Excellence (STAT COE)
STAT COE

Statistical Methods Applied to Cyber Testing
IDA/STAT COE

DOT&E Chief Science Advisor / Chief Scientist Initiatives
DOT&E, OUSD(R&E)

T&E Workforce Development
Systems Engineering and Architecture (SE&A), OUSD(R&E)

Meeting 3 (7-8 Nov 2022)

Digital Engineering Impact on the Future of Testing
Chair, Digital Engineering Task Force

Data Science at TRMC
TRMC

Meeting 4 (17 Jan 2023)

Army Perspective on Test & Evaluation
Director T&E, U.S. Army

Air Force Perspective on Test & Evaluation
Director T&E, U.S. Air Force

MDA Perspective on Test & Evaluation
Missile Defense Agency

Navy Perspective on Test & Evaluation
Deputy for Test and Evaluation, U.S. Navy

Meeting 5 (23 Feb 2023)

Space Force Perspective on T&E
U.S. Space Force

T&E Cyber and Historical Perspective
VADM T.J. White (USN, ret.)

Meeting 6 (29-30 Mar 2023)

Test and Evaluation at U.S. Army Operational Test Command (OTC)
Operational Test Command, U.S. Army

USAF Air Operations Center (231 Program)
Test & integration Chief, AFLCMC/HBB "Kessel Run", U.S. Air Force

AI Application for T&E
Chief of Test & Evaluation, Chief Digital and Artificial Intelligence Office (CDAO), Directorate for AI Assurance

Perspective on Challenges within T&E
CDAO, Directorate for AI Assurance

Former DOT&E Perspective
Deputy Director for Strategic Initiatives, Policy and Emerging Technologies (SIPET)

Current DOT&E Perspective
Former Director DOT&E

Meeting 7 (13 Apr 2023)

Test and Evaluation at CMU
Carnegie Mellon University, Software Engineering Institute

International T&E Programs
DOT&E

NASEM Report Briefing
Massachusetts Institute of Technology/National Academies of Sciences, Engineering, and Medicine

Meeting 8 (9 May 2023)

Joint Simulation Environment
Director, Joint Simulation Environment, Naval Air Warfare Center Aircraft Division (NAWCAD)

Navy's Operational Test and Evaluation Force Perspective
Operational Test & Evaluation Force (OPTEVFOR), Department of the Navy

Industry Best Practices – Lockheed Martin
VP/GM Strategic & Missile Defense Systems Lockheed Martin

Meeting 9 (20 June 2023)

Test and Evaluation in Industry and NDIA Industrial Committee on Test and Evaluation (ICOTE) Overview
ICOTE

T&E Organizational History
ICOTE

Joint Staff J7
Executive Director Joint Force Development and Design, Joint Staff J7

Meeting 11 (18 July 2023)

Air Force Operational Test and Evaluation Center (AFOTEC) Perspective
Executive Director, AFOTEC

Army Test and Evaluation Command (ATEC) Perspective
ATEC

Test and Evaluation at SpaceX
Director, Flight Reliability, SpaceX

Appendix F. Acronym List

177 IWS	177th Information Warfare Aggressor Squadron
AAF	Adaptive Acquisition Framework
AARW	Air-launched Rapid Response Weapon
ACAT	acquisition categories
ACV-C	Amphibious Combat Vehicle – Command and Control Variant
AFOTEC	Air Force Operational Test and Evaluation Center
AI	artificial intelligence
AI/ML	artificial intelligence/machine learning
AOC	Air Operation Center
ASD	Assistant Secretary of Defense
BtB	Back-to-Basics
C2	command and control
CAGR	compound annual growth rate
cATO	continuous authority-to-operate
CCM	Center for Counter Measures
CDAO	Chief Digital and Artificial Intelligence Officer
CJCS	Chairman, Joint Chiefs of Staff
ConOps	Concept of Operations
CPCE	Command Post Computing Environment
CPS	Conventional Prompt Strike
CT	contract testing
CTE	critical technology elements
CTEIP	Central Test and Evaluation Investment Program
D, DR&E(Adv Cap)	Director of Defense Research & Engineering for Advanced Capabilities
D, SE&A	Director, Systems Engineering and Architecture
D, TRMC	Director for Test Resource Management Center
D, TE	Director of Test and Evaluation
D, TRMC	Director, Test Resources Management Center
DD, DR&E	Deputy Director of Defense Research and Engineering
DASD	Deputy Assistant Secretary of Defense
DAU	Defense Acquisition University
DAWIA	Defense Acquisition Workforce Improvement Act
DCGS-A	Distributed Common Ground System- Army

DE	digital engineering
DevSecOps	development, security, and operations
DoD	Department of Defense
DoDI	Department of Defense Instruction
DOT&E	Director, Operational Test and Evaluation
DSB	Defense Science Board
DT	developmental testing
DT&E	developmental test and evaluation
DTE&A	developmental test, evaluation, and assessments
EMD	Engineering and Manufacturing Development
EW	electronic warfare
EWPMT	Electronic Warfare Planning Management Tool
FFRDCs	Federally Funded Research and Development Centers
F&R	findings and recommendations
FOC	full operating capability
FOT&E	follow-on test and evaluation
FY	Fiscal Year
HAPCAT	Hypersonic Aero Thermal and Propulsion Clean Air Testbed
HEL	high-energy laser
HPM	high-power microwave
IC	intelligence community
IED	improvised explosive device
IOC	initial operating capability
IOT&E	initial operational test & evaluation
IT	information technology
ITEP	International Test & Evaluation Program
ITRA	Independent Technology Risk Assessment
JAGM	Joint Air-to-Ground Missile
JEON	Joint Emergent Operational Need
JPO	Joint Program Office
JUON	Joint Urgent Operational Need
JSE	Joint Simulation Environment
JT&E	joint test & evaluation
KPPs	key performance parameters
KSAs	key system attributes

LFT&E	Live Fire Test and Evaluation
LIMWS	Limited Interim Missile Warning System
LVC	live, virtual, constructive
M&S	modeling and simulation
MBSE	Model-Based Systems Engineering
MCA	Major Capability Acquisition
MDAP	Major Defense Acquisition Programs
MDD	Materiel Development Decision
MILCON/MilCon	military construction
MPAH	Mid-Pressure Arc-Heater
MRTFB	Major Range Test and Facility Base
MS	Milestone
MSA	Materiel Solutions Analysis
MSIC	Missile and Space Intelligence Center
MTA	Middle Tier of Acquisition
NACME	National Action Council for Minorities in Engineering
NASEM	National Academies of Sciences, Engineering, and Medicine
NASIC	National Air and Space Intelligence Center
NDAA	National Defense Authorization Act
NGAD	Next Generation Air Dominance
OEM	original equipment manufacturer
OFP	Operational Flight Program
ONI	Office of Naval Intelligence
OSD	Office of the Secretary of Defense
OT	operational test
OT&E	operational test and evaluation
OTA	operational test agencies
PEO	Program Executive Officer
RDT&E	research, development, test, and evaluation
S&T	science and technology
SD	Secretary of Defense
SE&A	systems engineering and architecture
SMART	Science, Mathematics, and Research for Transformation
STEM	science, technology, engineering, and mathematics
T&E	test and evaluation

TEMP	Test and Evaluation Master Plan
TEST	Test and Evaluation Science and Technology
TLR	top level requirements
ToR	Terms of Reference
TRA	Technology Readiness Assessment
TRL	Technology Readiness Level
TRMC	Test Resource Management Center
TSO	Terrain Shaping Obstacles
TTPs	tactics, techniques, and procedures
UCA	Urgent Capability Acquisition
UON	Urgent Operational Need
USAF	United States Air Force
USD(A)	Under Secretary of Defense of Acquisition
USD(A&S)	Under Secretary of Defense for Acquisition and Sustainment
USD(AT&L)	Under Secretary of Defense of Acquisition, Technology, and Logistics
USD(R&E)	Under Secretary of Defense for Research and Engineering
USN	United States Navy
V&V	verification and validation
WSARA	Weapon System Acquisition Reform Act

Defense Science Board

DSB

Department of Defense