EXECUTIVE SUMMARY

This page intentionally left blank.

MAJOR PRODUCTS

In FY24, DOT&E designated 25 new DoD systems for OT&E and LFT&E oversight and removed 26 systems from the T&E Oversight List. As of September 2024, DOT&E had 265 DoD systems on the T&E Oversight List for OT&E and/or LFT&E, pursuing different acquisition pathways and in different phases of their acquisition life cycles. In FY24, DOT&E:

- Reviewed and approved 24 TESs/TEMPs and disapproved 1 TEMP.
- Approved 65 individual test plans.
- Published 47 reports, including 29 reports to the Services, Congress, and/or the SECDEF providing system evaluations, a classified annual assessment of the Missile Defense System, and 17 special or legislative reports.

DOT&E completed nine legislative actions, summarized in Table 1, for which DOT&E was assigned as the Office of Primary Responsibility (OPR). DOT&E completed seven legislative actions, summarized in Table 2, for which DOT&E was assigned Office of Coordinating Responsibility (OCR).

Table 1. Summary of DOT&E Congressional Activities as OPR

Source	Title	Status
FY23 NDAA		
Sec. 217	Competitively awarded demonstrations and tests of electromagnetic warfare technology	Complete
Sec. 242	Study and report on sufficiency of operational test and evaluation resources supporting certain major defense acquisition programs	Complete
Sec. 1656	Persistent cybersecurity operations for ballistic missile defense systems and networks	Complete
FY23 Other Legislative Actions		
HASC Report	Assessment of contractor-provided test and evaluation capabilities	Complete
HASC Report	Battery testing infrastructure	Complete
HASC Report	Development and testing of body-worn equipment	Complete
HASC Report	Equipment shortfalls within the test and evaluation community	Complete
FY24 Other Legislative Actions		
SAC Bill	Assessment of the DoD's and Services' Funding of Test Infrastructure, Assets, and Personnel to Support Agreed-Upon Test and Evaluation of Programs on the DOT&E Oversight List	Complete
SAC Bill	Certification of Appropriateness and Risk Assessment of Services' Planned Test Strategies for Approved Middle Tier of Acquisition (804) and Accelerated Acquisition Programs	Complete

Acronyms: HASC – House Armed Services Committee; NDAA – National Defense Authorization Act; SAC – Senate Appropriations Committee

Table 2. Summary of DOT&E Congressional Activities as OCR

Source	Title	Status
FY22 NDAA		
Sec. 833	Pilot Program on Acquisition Practices for Emerging Technologies	Complete
Sec. 1529	Demonstration program for automated security validation tools	Complete
FY22 Other Legislative Actions		
HASC Report	Report on Testing Infrastructure to Support Strategic and Missile Defense Programs	Complete

Table 2. Summary of DOT&E Congressional Activities as OCR, Continued

Source	Title	Status
FY23 NDAA		
Sec. 240	Report of potential for increased utilization of the electronic proving grounds testing range	Complete
Sec. 1514	Operational testing for commercial cybersecurity capabilities	Complete
Sec. 1553	Plan for commercial cloud test and evaluation	Complete
FY24 Other Legislative Actions		
SASC Report	Fiscal Year 2024 Modernization Plan of Hill Air Force Base Little Mountain Test Facility	Complete

Acronyms: HASC – House Armed Services Committee; NDAA – National Defense Authorization Act; SASC – SSenate Armed Services Committee

In FY24, DOT&E completed the adjudication of all stakeholder comments on the forthcoming DoD Instruction for OT&E and LFT&E and the following DoD Manuals:

- TEMP/TES
- Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A) for OT&E and LFT&E
- OT&E and LFT&E of Software
- OT&E and LFT&E of Artificial Intelligence (AI)-Enabled and Autonomous Systems
- Full-Spectrum Survivability and Lethality T&E

These policies, which DOT&E expects to publish by early FY25, are intended to enhance DOT&E's execution of its OT&E and LFT&E roles and responsibilities through the acquisition lifecycle. They emphasize the importance of using the right type and amount of data, including from validated modeling and simulation, training events, and joint exercises, to evaluate operational effectiveness, suitability, survivability, and lethality. They also consider survivability and lethality holistically across all potential threat and target types – kinetic and non-kinetic.

OT&E AND LFT&E OVERSIGHT OF DOD SYSTEMS

» ENSURED ADEQUATE OT&E AND LFT&E PLANNING AND EXECUTION

TES and Test Plan Recommendation Trends

In FY24, DOT&E evaluated the adequacy of TEMPs, TESs, and test plans to ensure they will provide: (1) data to support credible evaluation of operational effectiveness and suitability, (2) coverage of the operational environment and threats with users executing realistic mission operations, (3) adequate verification and validation (V&V) of M&S, (4) complete assessments of system survivability and lethality against relevant kinetic and non-kinetic threats, (5) production-representative test articles, and (6) sufficient funding and resources required to support test execution.

In FY24, DOT&E approved all but one TEMP. The TEMP disapproval was the result of insufficient resources available to execute the strategy as documented. Common DOT&E conditions for

document approval include: (1) testing of the supply chain and inclusion of all potential attack vectors in contested cyberspace, (2) coverage of the operational environment and threats (current and future), (3) M&S V&V plans, (4) use of latest software versions, (5) data collection processes or equipment, (6) use of operationally realistic users and maintainers exercising the most recent tactics, techniques, and procedures, (7) planning for any correction of fixes regression testing, (8) survey administration, and (9) resource sufficiency.

Test Adequacy Recommendation Trends

In FY24, DOT&E assessed the adequacy of OT&E and/or LFT&E in 28 of 29 systems evaluations.[1] DOT&E assessed 64 percent (18 of 28) of testing as adequate, 14 percent (4 of 28) as partially adequate, and 21 percent (6 of 28) as not adequate, as shown in Figure 1. By comparison, over the last 8 years (FY16 – 23), DOT&E assessed 66 percent (167 of 252) of the executed OT&E and LFT&E as adequate, 24 percent (61 of 252) as partially adequate, and 10 percent (24 of 252) as not adequate. The determinations of inadequacy or partial adequacy of OT&E and LFT&E in FY24 were caused by: (1) insufficient scope or lack of operational testing prior to early fielding, (2) early test termination or execution shortfalls, (3) insufficient data to conduct rigorous analysis, and/or (4) lack of testing against all relevant threats, including cyber and electromagnetic spectrum. DOT&E also highlighted limitations discovered in testing or in post-test analysis, including but not limited to:

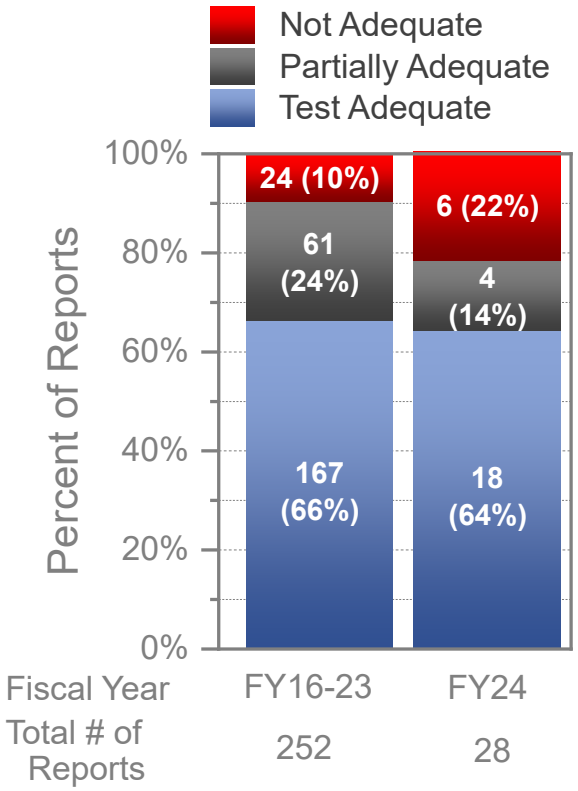

Figure 1. Test Adequacy in FY24 and Prior Years

- Lack of a complete assessment of relevant non-kinetic attack vectors or paths.
- Safety restrictions.
- Software and firmware updates.
- Reliability of an available threat target, simulator, or surrogate.
- Presentation of an operationally realistic environment that replicates the most challenging scenarios and based on current threat information.
- Insufficient amount and quality of data to conduct an adequate evaluation.
- Lack of mission effects data due to being part of a training exercise or day-to-day operations with deprioritized OT&E objectives.

DOT&E reports also provide recommendations for improving test adequacy. These recommendations include:

- Completing testing with production-representative assets to support early fielding or deployment decisions.
- Conducting operational testing across the relevant missions sets, operating conditions, and threats. Execution of robust testing continues to reveal important shortfalls that can be addressed prior to fielding.
- Testing all relevant cyber-attack paths, including supply chain on all subcomponents, and requiring test teams to execute a complete cyber restore of the system.

1 DOT&E did not make an adequacy determination in the Three-Dimensional Long-Range Radar Operational Assessment Interim Observation Memo.

- Evaluating system suitability and cyber survivability early in the design to increase test efficiency, discover problems early, and improve outcomes in OT&E and LFT&E.
- Developing robust and independent V&V for all M&S for use in OT&E and LFT&E.
- Increased use of telemetry for data collection and understanding interoperability mission effects.

Programs Pursuing the Middle Tier of Acquisition Pathway

In FY24, for the 97 programs approved by the Service Acquisition Executives to pursue the Middle Tier of Acquisition pathway, DOT&E received and reviewed 45 test strategies and certified 37 of those as appropriate, and 8 test strategies as not appropriate. DOT&E did not review the test strategies for the remaining 52 programs because they were either still in development or not made available for review. Test strategies were not certified as appropriate primarily due to inadequate resources for OT&E and/or LFT&E to evaluate the required performance in operationally representative environment, including in contested cyberspace and electromagnetic spectrum environments.

Adequacy of Funding Resources for Programs with Approved TEMPs or TESs

In FY24, DOT&E assessed the adequacy of OT&E and LFT&E resources required to execute the agreed upon OT&E and LFT&E, scheduled in the current year and future years defense planning. This assessment could only be made for those programs on DOT&E oversight that have approved TEMPs or TESs.

- Fifty-three percent (71 of 134) of the eligible programs were assessed to have adequate funding to support the remainder of the planned test execution. Sixteen percent (21 programs) were identified as having funding shortfalls, while 19 percent (26 programs) required updated TEMPs or TESs due to program changes that may require new or altered testing or resource requirements. Eleven percent (15 programs) have fully executed all required testing and require no current or Future Years Defense Program funding. One additional program, Public Key Infrastructure Increment 2, was not assessed despite being eligible for this assessment because funding data was not provided.
- The identified OT&E and LFT&E resource funding shortfalls were primarily related to the following: (1) flight test instrumentation, most commonly, for Open Air Battle Shaping capability, (2) funding for LFT&E events, and (3) accredited threat representation in contested environments, including space.

» PROVIDED INDEPENDENT EVALUATION OF OPERATIONAL PERFORMANCE

In FY24, DOT&E published 29 independent system evaluation reports on the operational performance of the system. System reports where DOT&E was unable to assess operational effectiveness, operational suitability, and/or survivability were based on early fielding and early operational testing, where not enough data are available to make full assessments. In those cases, DOT&E's reports comment on progress towards operational effectiveness, operational suitability, and survivability. The performance trends, discussed below, are depicted in Figure 2.

Operational Effectiveness Trends

In FY24, DOT&E was able to assess operational effectiveness for 13 of 29 systems reports. Of those 13 evaluated programs, DOT&E reported 54 percent (7 of 13) as operationally effective. By comparison, over the last 8 years (FY16 – 23), DOT&E reported 53 percent (82 of 154) as operationally effective. DOT&E assessed two FY24 programs as not operationally effective and four programs as being partially effective because the system could either not complete one or more of its primary missions, the system performed worse than the legacy capability, or had poor operational effectiveness in some operationally relevant conditions against intended threats and targets, including against realistic cyber and electromagnetic spectrum environments.

Operational Suitability Trends

In FY24, DOT&E was able to assess operational suitability for 12 of 29 systems reports. Of those 12 evaluated programs, DOT&E reported 58 percent (7 of 12) as operationally suitable. By comparison, over the last 8 years, DOT&E reported 48 percent (70 of 147) as operationally suitable. DOT&E assessed four programs as not operationally suitable and one program as being partially operationally suitable. These five programs experienced shortfalls in hardware and software reliability and availability. Other common suitability limitations included insufficient training, maintainability, and network connectivity issues.

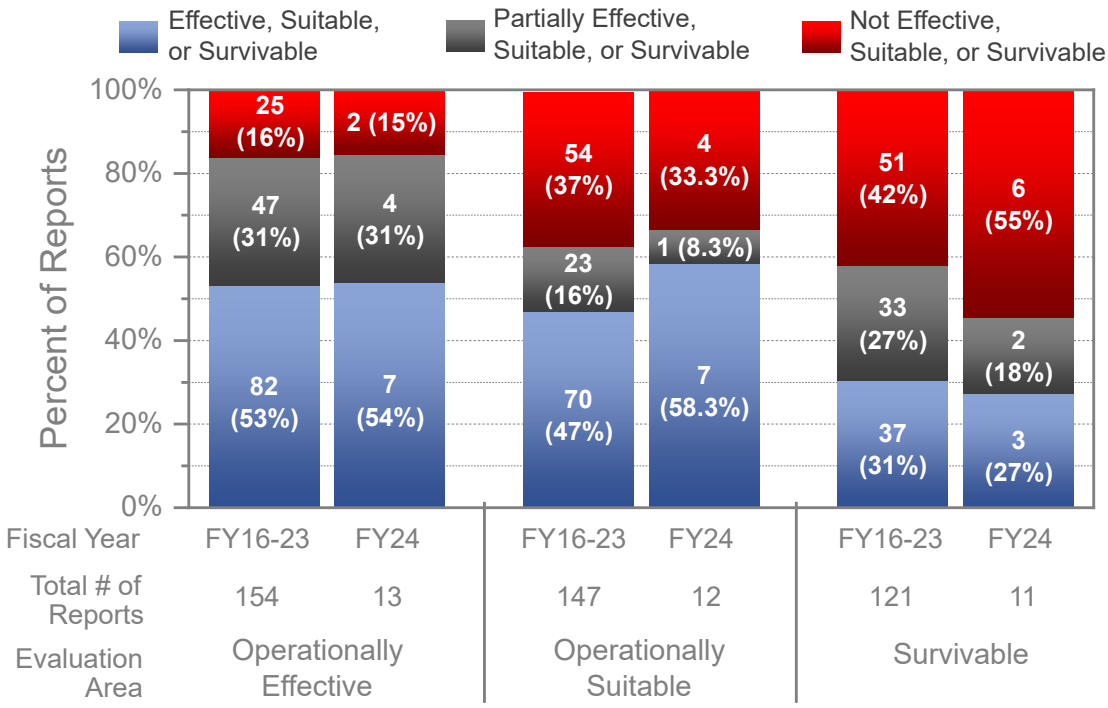

Figure 2. Operational Performance Trends in FY24 and Prior Years

Survivability Trends

In FY24, DOT&E assessed survivability for 11 of 29 systems reports. Of those 11 evaluated programs, DOT&E reported 27 percent (3 of 11) were survivable and 18 percent (2 of 11) were partially survivable. By comparison, over the last 8 years, DOT&E assessed 31 percent (37 of 121) as survivable and 27 percent (33 of 121) as partially survivable, primarily due to vulnerabilities in contested cyberspace. Cyber threats remain the most common threat type tested against in comparison to testing against kinetic; electromagnetic spectrum; or chemical, biological, radiological, and nuclear (CBRN) threats in OT&E and LFT&E.

Recommendations Trends

DOT&E reports include practical recommendations to fix the identified deficiencies and improve the operational performance of the DoD systems in expected operational scenarios and conditions to minimize risk to warfighters and maximize probability of mission success. Examples of common problems discovered in OT&E and LFT&E include immature software, poor reliability, poor network availability and connectivity, not survivable against cyber-attacks, poor system performance in all threat and operational environments, deficient human systems integration, and insufficient training and technical manuals. DOT&E commonly makes recommendations to fix system deficiencies in these problem areas prior to fielding.

» DOT&E ACTIVITIES SUMMARY

In FY24, DOT&E continued to manage the Cyber Assessment Program (CAP) alongside the following field activities: (1) Center for Countermeasures (CCM), (2) Joint Aircraft Survivability Program (JASP), (3) Joint Technical Coordinating Group for Munitions Effectiveness (JTCG/ME) that includes the Joint Live Fire (JLF) program, (4) Joint Test and Evaluation (JT&E), and (5) Test and Evaluation Threat Resources Activity (TETRA). These efforts supported the advancement of DOT&E's Strategy Implementation Plan (I-Plan), published in April 2023, which focused on integrating emerging technologies and adapting our workforce to future challenges. The year marked

a shift toward more innovative and forward-thinking testing approaches. Below is a summary of how these field activities have enhanced the DoD's T&E infrastructure, tools, processes, and workforce in FY24. Further details are available in the DOT&E Strategy I-Plan Update, T&E Resources, and DOT&E-Managed Activities sections of this Annual Report.

T&E Infrastructure

DOT&E field activities have improved the DoD's T&E infrastructure by introducing advanced technologies and modernizing facilities. For example, JTCG/ME has implemented cloud-hosted environments that enable real-time analysis of weapon effectiveness data. In partnership with the Test Resource Management Center, they have also provided new infrastructure to support realistic and rapid cyber testing. JLF expanded DoD-wide data repositories to support review, approval, and access of lethality and vulnerability data and documentation. JASP coordinated development of a threat launch simulator for testing two-color infrared missiles warning systems, which will be a vital asset for maturing countermeasure systems. TETRA has delivered high-fidelity electronic warfare (EW) threat models to enhance EW and space system testing. The Integration Lab, launched by JT&E, introduced a digital transformation framework, promoting AI systems and digital twin workflows development.

T&E Tools

This year saw significant enhancements in T&E tools. Key advances included the use of AI and machine learning by groups like JTCG/ME, which created predictive models for fragment penetration and lethality assessments. JTCG/ME also expanded existing lethality and vulnerability data repositories and used modern software methods to enhance weaponeering tool capabilities and interfaces. TETRA is investigating use of an intelligence digital ecosystem to analyze threat intelligence data, supporting more efficient test design and threat modeling across multiple domains, including cyber and space. TETRA is also initiating pilot activities to develop AI-driven EW threat models that simulate complex adversary systems for testing. JASP introduced and validated new tools to enhance aircraft survivability assessments, including capability to simulate multi-domain engagements, EW, cyber threats, surface-to-air missiles, and high-energy lasers. CCM continued to support numerous test events in FY24 by providing threat simulators and other tools to characterize platform survivability.

T&E Processes

FY24 improvements in test processes have been substantial, incorporating modern statistical methods, AI, automation, and real-time analytics. TETRA provided roadmaps to improve test designs for space asset survivability. JTCG/ME developed a workflow management tool to streamline targeting data. JASP partnered with the Navy to improve the efficiency and accuracy of cyber analysis using a digital twin for the P-8A. JT&E introduce introduced an agile test process to expedite T&E of joint warfighter concepts and tactics, techniques, and procedures.

T&E Workforce

DOT&E invested heavily in workforce development, expanding training programs and internships to prepare staff for emerging technological challenges. JTCG/ME's training events have enhanced operational proficiency in new weaponeering technologies. JASP continued to enhance training on aircraft combat damage assessments, focusing on real-time forensics in anti-access/area denial environments. TETRA continues to manage configuration control boards to bring together experts from communities and disciplines to foster a workforce capable of addressing modern multi-domain threats. DOT&E revised its competency model and is working on a course catalog that will map specific trainings to the competencies to provide DOT&E Action Officers (AOs) the resources and guidance they need to increase their skills.

» CONTINUED SUPPORT TO GLOBAL T&E PARTNERSHIPS

The International Test and Evaluation Program (ITEP) has been making significant strides in strengthening international partnerships. In FY24, ITEP signed 16

new agreements, bringing the total number of active agreements to 33. These agreements cover a wide range of testing activities, including EW, tactical armored personnel vehicle testing, data fusion, reciprocal use of facilities, and more. The partners involved are from Australia, Canada, Germany, Italy, Japan, the Netherlands, Norway, and the United Kingdom. These projects aim to improve capabilities and instrumentation among U.S. allies in areas including EW, autonomy, and survivability.

Outside of ITEP, DOT&E increased collaboration with allies and partners in areas such as experimentation, co-development, research, testing, and evaluation. Key strategic partnerships were identified, focusing on the development of AI-enabled systems, synthetic ranges, and the integration of the DOT&E Strategy with the United Kingdom's "T&E Transformation Programme." Short term initiatives aim to enhance verification, validation, and accreditation of AI-enabled systems. A memorandum of understanding was finalized with the United Kingdom, establishing a United Kingdom liaison position within DOT&E to further advance collaboration on these strategic initiatives.

» CONTINUED SUPPORT TO T&E WORKFORCE

DOT&E offered its annual AO Course from September 30 to October 4, 2024. This annual training covers AO duties, their role within the test community, and the basics of DOT&E's legal obligations. This year, the course was offered in a hybrid format and comprised briefings on over 30 topics, to include policy overviews, technical topics, exercises, and panels. Approximately 100 people registered for the course, including staff from DOT&E, DOT&E's FFRDC-support, and other agencies, such as Service operational test agencies and foreign partners. Post-course survey results showed that attendees were satisfied with the course, finding many presentations engaging and informative. Attendees also found the course effective and well-organized. They self-reported an increased knowledge about all DOT&E topics, with a particular increase noted for electromagnetic spectrum operations; chemical, biological, radiological, and nuclear; and AI. Future courses plan to include additional reference materials and examples.