



Executive Summary

Operational and LFT&E is essential to demonstrate weapon system performance and provide DOD mission planners, commanders, operators, and maintainers with an understanding of true weapon system capabilities, and data to adequately plan and execute their mission in combat. In FY21, DOT&E provided oversight for 237 acquisition programs and published its first Science and Technology Strategy.

Major Products

In FY21, DOT&E provided operational and live fire test and evaluation oversight for 237 acquisition programs at various stages in their acquisition cycle.¹ Specifically, DOT&E reviewed and approved 26 Test and Evaluation Master Plans (TEMPs), 9 of which included a Live Fire Test and Evaluation (LFT&E) Strategy; 2 separate LFT&E Strategies; and 56 individual test plans.

DOT&E evaluates the adequacy of the Service test strategies and plans based on the degree that they will provide: 1) data to support credible evaluation of operational effectiveness and operational suitability, 2) coverage of the battlespace and threats, 3) adequate use of modeling and simulation (M&S), 4) complete cybersecurity and live fire assessments, including demonstration of system survivability and lethality against mission-relevant threats, 5) production-representative test articles, 6) operational realism, and 7) sufficient funding required to support test execution.

DOT&E published 26 reports, including 23 reports to Congress and the Secretary of Defense, and a classified annual report on the Ballistic Missile Defense Systems. In addition to the assessment of test adequacy, DOT&E reports summarize the Director's independent assessment of operational effectiveness, lethality, suitability, and survivability of DOD weapon systems in expected combat conditions. In instances where operational and live fire testing and evaluation have not yet been completed, DOT&E provides an interim assessment and identifies any risk to accomplishing the required operational performance in upcoming operational and live fire test, prior to fielding or the next acquisition decision review. DOT&E reports summarize practical recommendations intended to fix the identified deficiencies and improve the operational performance of the weapon system in expected operational scenarios and conditions to minimize risk to warfighters and maximize probability of mission success in conflict.

In FY21, DOT&E published its first Science and Technology Strategy focused on addressing the following T&E challenges: 1) software and cyber T&E, 2) next generation T&E capabilities, 3) needed integrated T&E lifecycle, 4) digital transformation, and 5) workforce expertise and partnerships. DOT&E intends for these strategic initiatives to inform emerging T&E policy and guidance and enable agile yet credible T&E that can adequately support acquisition reforms while responding to the emerging technology requirements and the increasingly complex and dynamic multi-domain operational environment.

In March 2021, in response to the FY21 Department of Defense Appropriations Act, DOT&E published a follow-on suitability assessment of MHS GENESIS. In April 2021, DOT&E testified before the Senate Armed Services Committee Readiness Subcommittee on the performance of the DOD acquisition, while in July 2021, DOT&E testified before the House Armed Service Committee, Tactical Air and Land Forces Subcommittee on the FY22 budget request for the DOD for fixed-wing tactical and training aircraft programs. In May 2021, in response to the Senate Appropriations Committee, Explanatory Statement for the Department of Defense Appropriations Bill, 2021, DOT&E published the Certification of Appropriateness of Services' Planned Test Strategies for Approved Middle Tier of Acquisition (804) and Accelerated Acquisition Programs report. Lastly, Table 1 provides the status of several completed and ongoing activities in response to the FY20 and FY21 National Defense Authorization Acts (NDAA).

¹ The number of programs on DOT&E oversight fluctuates throughout the year; 237 is the number of programs on DOT&E oversight as of September 30, 2021.

Table 1. Summary of DOT&E NDAA Activities

Section #	Title	Status
FY 2020 NDAA		
231	Digital Engineering Capability to Automate Testing and Evaluation	Ongoing; DOT&E in support of R&E
800	Authority for Continuous Integration and Delivery of Software Applications and Upgrades to Embedded Systems	Complete with publication of DOD Instruction 5000.89
FY 2021 NDAA		
112	Report on limitations of Integrated Visual Augmentation System (IVAS)	Ongoing
159	Documentation Related to F-35 Program	Ongoing
162	Briefings on Software Regression Testing for F-35	Ongoing; A&S develop quarterly briefings in consultation with DOT&E
222	Activities to Improve Fielding of Air Force Hypersonic Capabilities	Ongoing; R&E to deliver report in consultation with DOT&E
271	Modification to Annual Report of the Director of Operational Test and Evaluation	Adds 1 year to sunset date of DOT&E Annual Report
277	Independent Evaluation of Personal Protective and Diagnostic Testing Equipment	Complete
836	Digital Modernization of Analytical and Decision-Support Processes for Managing and Overseeing Department of Defense Acquisition Programs	DOT&E is a member of the Steering Committee

Major Contributions

Ensure Adequate Testing in Combat Representative Conditions

In FY21, DOT&E continued to highlight and correct instances where proposed test plans were not adequate. Based on the test plans that DOT&E reviewed in FY21, common shortfalls were associated with data collection plans, deficiencies with M&S fidelity or validation, test resources constraints, and insufficient coverage of the operational environment and threats, including insufficient test scope and threat realism for cyber assessments. To address these test shortfalls, DOT&E worked with program stakeholders to improve the test adequacy of plans.

In addition, because some test shortfalls result from range infrastructure challenges, DOT&E recruited the National Academies of Sciences, Engineering, and Medicine to conduct a study on the health and readiness of the DOD test ranges and associated infrastructure for future operational and live fire testing. DOT&E also established a T&E resources and infrastructure working group responsible for cataloging and resolving operational and LFT&E resource and infrastructure shortfalls in coordination with USD(R&E) and other DOD stakeholders.

The National Academies of Sciences, Engineering, and Medicine published their report in September 2021 offering the following five major recommendations: 1) develop the “range of the future” to test complete kill chains in Joint All Domain Operational environments, 2) restructure the range capability requirements process for continuous modernization and sustainment, 3) bootstrap a new range operating system for ubiquitous M&S throughout the weapon system development and test life cycle, 4) create the “TestDevOps” digital infrastructure for future operational testing and seamless range enterprise interoperability, and 5) reinvent the range enterprise funding model for responsiveness, effectiveness, and flexibility. DOT&E is evaluating the National Academies’ recommendations and will work with DOD stakeholders to address each as appropriate, and as resources allow.

In parallel, through the newly-formed T&E resources and infrastructure working group, and in coordination with the Test Resources Management Center in USD(R&E), DOT&E initiated the development of more representative electronic warfare testing at Navy sea and land ranges, a threat torpedo capable of simulating a range of acoustic signatures, the acquisition of miniaturized instrumentation for data collection from unmanned aerial system threats, and continued the development of the next generation aerial target.

Ensure Adequate Testing Across any Acquisition Pathway

The DOD has made significant changes to its acquisition policies to support the National Defense Strategy goal of delivering performance at the speed of relevance. To support faster delivery of proven warfighting capability, in November 2020, DOT&E, in conjunction with USD(R&E), USD(A&S), and Service T&E executives, supported the publication of a DOD Instruction 5000.89, which provides T&E procedures for new acquisition pathways that include Urgent Capability Acquisition, Middle Tier Acquisition, Major Capability Acquisition, Software Acquisition, and Defense Business Systems.² Significant efforts are underway to provide the T&E community with the tools, architectures, and methods required to optimize the benefits of integrated testing, digital engineering tools and enable agile T&E without compromising the credibility of operational performance evaluation. Such improvements will be documented in the Enterprise T&E Guidebook that is being developed by DOT&E and USD(R&E) to provide the DOD Acquisition and T&E communities the tailorable guidance they require to ensure adequate developmental, operational, and live fire T&E for each of the acquisition pathways.

In the interim, in FY21, DOT&E assessed the appropriateness of test strategies for 86 programs approved by the Service Acquisition Executives to pursue accelerated acquisition authorities. DOT&E reviewed 47 test strategies (the remaining 35 were not made available for review) and certified 33 of those as appropriate, while observing the following: 1) test strategies frequently lack well-defined resources to plan and execute operational testing, or to train operators, maintainers, and cyber defenders, 2) test strategies lack the rigor typically required to demonstrate operational effectiveness, suitability, survivability, and lethality, 3) adoption of integrated test approaches with rapid test/fix/test cycles to enable agility has begun to stress the Service operational test agencies and developmental test organizations, which are currently not resourced, staffed, or trained for the continuous level of effort and reporting required by such approaches.

Transforming T&E Concept of Operations

The increasing complexity of U.S. weapons systems and the capabilities of our potential adversaries, compounded with the parallel, increasing complexity of the environments in which combat will be conducted, continue to underscore the importance and need for transforming T&E concept of operations. As the warfighting capability continues to evolve to support the DOD's ability to fight and dominate in a multi-domain operational environment, the T&E community will require innovative and enterprise-level approaches to enable realistic testing, both live and virtual. To support the new T&E concepts, DOT&E has emphasized the need for investments in: 1) tools to automate testing and visualize the test space and mission effects, 2) data collection, storage, and analytics improvements, 3) improved virtual environments and M&S tools that are credible and validated by live data, 4) tools and methods such as sequential testing and uncertainty quantification to optimize integrated T&E, and 5) tools and methods to test autonomous and artificial intelligence (AI) enabled systems, hypersonic weapons, directed energy weapons, space systems, and other emerging T&E challenges. To adequately focus on meeting these and similar objectives, DOT&E established a new Deputate for Strategic Initiatives, Policy, and Emerging Technologies (SIPET). Notable FY21 efforts in this domain can be grouped into five major lines of effort:

2 DOT&E and USD(R&E) are assessing the inclusion of the Acquisition of Services Pathway in the next update to DoDI 5000.89

1. Enhance Software and Cyber Testing and Evaluation

In FY21, DOT&E established a team of software and cyberspace experts from across the organization to orchestrate internal and external efforts to improve cyber T&E and DOD cyber strategic initiatives. Specific objectives include: 1) improving cyber threat representation, 2) optimizing mission-focused cyber assessments, 3) increasing the availability and integration of cyber expertise, 4) increasing the understanding and inclusion of cyber OT&E of defensive countermeasures in cyberspace, 5) enhancing OT&E of DOD's cyberspace attack and enabling capabilities, and 6) emphasizing vulnerability management in all phases of a program's lifecycle. Initiatives to improve standardization of cyber T&E data and to assess effects from the supply chain are also underway.

2. Develop Next-Generation T&E Capabilities

As AI and autonomy advance to become an integral part of the DOD mission space, DOT&E is teaming up with the Joint Artificial Intelligence Center and USD(R&E) to develop a T&E roadmap for such systems. While Industry's approaches, best practices, and technologies are informative, they do not address the suite of challenges and needs when evaluating DOD capabilities that operate in complex and degraded environments and inform strategic and tactical decisions critical to national security. The roadmap aims to ensure that the larger T&E community is developing the test strategies, practices, methods, infrastructure, data tools, workforce, and other T&E needs as AI-based systems and technologies mature.

3. Enable Optimal Integration of T&E Across the Program Lifecycle

The DoDI 5000.89 policy emphasizes the importance of integrated testing to increase the efficiency of the overall T&E program by planning test events that provide data for multiple objectives. Integrated testing provides programs with the opportunity to identify problems earlier in developmental test, improve production readiness, and shorten the acquisition timeline by leveraging more operationally relevant data across the acquisition cycle. Specifically, DOT&E has partnered with USD(R&E) to expedite the implementation of an integrated decision support key framework intended to ensure data-based acquisition decisions. This guidance will provide a more structured and standardized approach for program stakeholders to align decision points with the operational and technical evaluations and events necessary to inform decisions. Using this framework, testing could be planned in a mission context with operational end users earlier by adopting test design methodologies, such as sequential methods. Using these methods for test planning, execution, and evaluation, individual test events build upon each other and are refined based on previous test outcomes, avoiding redundancies without compromising the credibility of the evaluation.

4. Enable Digital Transformation to Advance T&E Efficiency

In partnership with USD(R&E)-TRMC, DOT&E led the selection and execution of five demonstrations showcasing applications of digital engineering paradigms as called for in the FY20 NDAA Section 231 study, "Digital Engineering Capability to Automate T&E." Demonstration results, still preliminary, suggest improvements spanning attributes such as quality, cycle time, predictability, and costs. This work spawned a number of related digital transformation initiatives, to include: FY22 NDAA Section 217, "Development and Implementation of Digital Technologies for Survivability and Lethality Testing," as well as initiatives in agile Verification, Validation, and Accreditation (VV&A), and an assessment of digital twins.

Separately, in September 2021, DOT&E kicked off an internal initiative to improve T&E data management by automating the manual processes of searching and aggregating data elements from various artifacts. An initial proof of concept will demonstrate the ability to ingest a variety of unstructured documents and automatically process them into a readable machine learning format, giving the T&E community the ability to quickly and easily identify information required to inform requirements, users, materiel developers, and acquisition decisions.

5. Prepare the T&E Workforce for the Future

In August 2021, DOT&E initiated a technical workforce assessment to better understand and develop the knowledge, skills, and abilities the T&E workforce needs to execute its mission as DOD weapon systems evolve. This effort is crucial to ensure the organization is optimally structured, organized, and postured for success. Over the coming months, DOT&E will develop a Technical Skills and Manpower Report and Strategic Workforce Plan detailing the challenges, opportunities, and actionable steps DOT&E can take to best position the T&E workforce to meet the mission of today and the future. Our greatest asset is our workforce, and this assessment will help DOT&E provide its people the support and resources they need to stay ahead of evolutions and revolutions in T&E.

In addition to the workforce assessment, DOT&E partnered with Cyber Test Teams across the Services to complete the first year of Software and Cyber Network of Excellence for Testing (SCyNET) pathfinding activities. These pathfinding initiatives are identifying requirements, defining the business case, and documenting lessons learned for institutionalizing a long-term capability, provided in the form of university-based service providers, to address strategic T&E gaps. This real-time DOD operator and university researcher connection has both solved problems and developed a DOD-university relationship for future work and collaboration.

Demonstrating the Value of T&E

T&E is essential to demonstrate weapon system performance and provide DOD mission planners, commanders, operators and maintainers with an understanding of true weapon system capabilities and data to adequately plan and execute their missions in combat. Examples of this can be found in the Joint Technical Coordinating Group for Munition Effectiveness, Joint Test and Evaluation, and Cyber Assessment Program sections of this report. Specifically, DOT&E cyber-related activities have helped the DOD characterize cyber effects on mission performance, identify network and system vulnerabilities, assess operational concepts and procedures, enhance cyber team capabilities, update guidance and methodologies, facilitate operational assessment of offensive cyber capabilities, and inform the Department on cyber considerations of initiatives and technologies such as the move to commercial cloud-based computing. DOT&E cybersecurity assessments have uncovered important vulnerabilities that, if corrected, will improve the Department's resilience against cyberattacks. T&E, in general, identifies warfighting performance shortfalls that could and should be addressed prior to weapon system fielding or the next acquisition decision. This identification permits corrective action to be taken before large quantities of a system are procured and avoids expensive retrofit of system modifications. An example includes the full ship shock trial testing on the CVN 78 that identified several CVN 78 design shortfalls that, if addressed, could improve the survivability of the CVN 78 against underwater torpedo or mine engagements. The performance trends section below provides additional detail on the value of T&E.

Major Findings

Test Adequacy Trends

Consistent with DOT&E reports from previous years, in FY21, DOT&E reported that 62 percent (13 of 21) of programs conducted adequate operational testing, as detailed in Figure 1.³ Of the eight programs assessed as not adequate or partially adequate, five programs reported cyber testing inadequacies due to limited breadth of coverage; insufficient collection of data on mission effects and the ability to prevent,

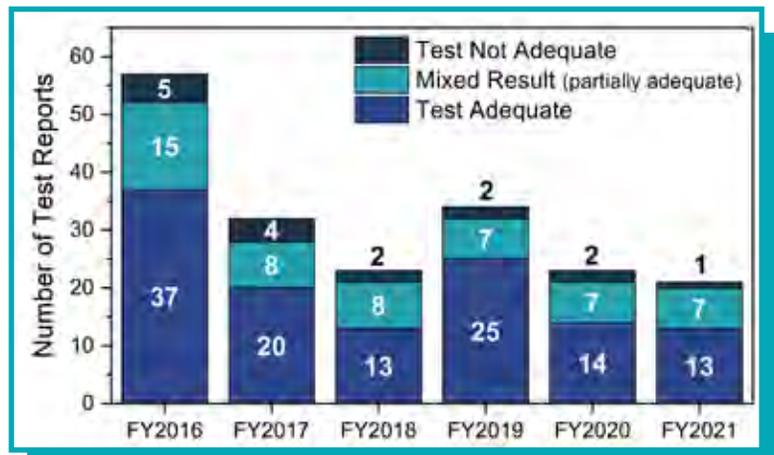


Figure 1. Test Adequacy Trends in DOT&E Reports

³ Five FY21 reports were excluded where DOT&E did not make a test adequacy assessment.

mitigate, and recover from attacks; and lack of sufficient funding. Three programs reported deficiencies with M&S fidelity or model validation. Two programs reported that the most challenging threats were not considered or the threat used was not portrayed properly. Other test adequacy issues included contractor support that is not combat representative, failure to collect or deliver all required data, and system developmental delays that led to incomplete testing.

In addition to test adequacy concerns, DOT&E reports identified other test execution limitations. Common test limitations included inadequate data collection, test range environmental restrictions that prevented a robust operational assessment, unrealistic maintenance due to overreliance on field support representatives, limited doctrinal training resources that prevented full use of new system capabilities, and operational testing being limited to one environment or not covering all required threats, such as electronic attack. For eight programs, COVID-19 hindered full test participations of all T&E stakeholders, which affected data collection and the availability of supporting assets and other resources.

DOT&E-approved test plans also provide insights into known test limitations. As shown in Figure 2, survivability was the most common type of test plan limitation, followed, in order, by limitations that affected DOT&E's assessment of effectiveness and suitability. The majority (93 percent) of survivability limitations were due to cybersecurity. Twenty-four of the 56 test plans in FY21 were focused only on cybersecurity and all but two identified cybersecurity limitations. Common cyber test limitations included lack of advanced cyberattack capabilities by cyber Red Teams, inadequate coverage of all attack vectors due to concerns with safety or the compromise of live operational networks, the need for a more robust supply chain assessment, missing test resources such as data connection cables, insufficient time for Red Teams to probe all possible threat vectors, and lack of an available full-up system. Other common test limitations included M&S or model VV&A deficiencies that were sometimes due to lack of an available full-up system.

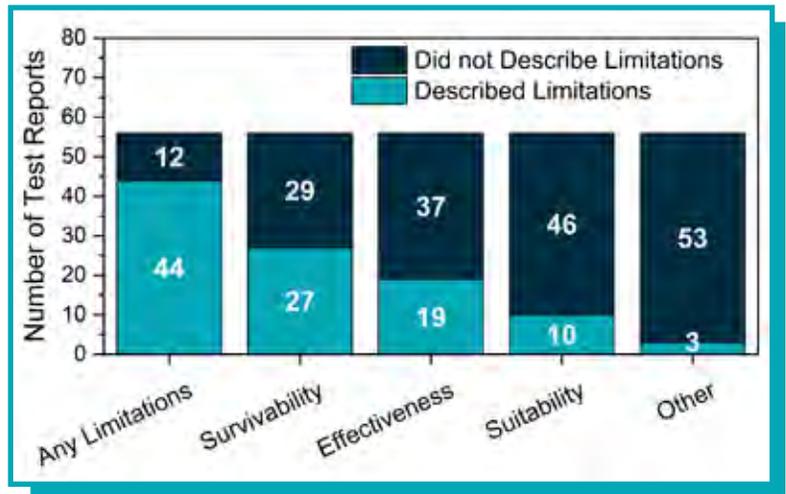


Figure 2. Limitations in DOT&E FY21 Test Plans by Area

Performance Trends

Figures 3 through 5 show the result of DOT&E assessments of operational effectiveness, operational suitability, and survivability since FY16. The figures exclude reports where DOT&E did not make an assessment because the test event was too early in the acquisition cycle, was narrow in scope, or had limitations that precluded an assessment of operational performance.

Effectiveness

In FY21, DOT&E evaluated 67 percent of programs to be operationally effective without any caveats. Reasons for systems being not operationally effective included: system, software, or integration deficiencies; training limitations that affected operator performance or unit effectiveness; and

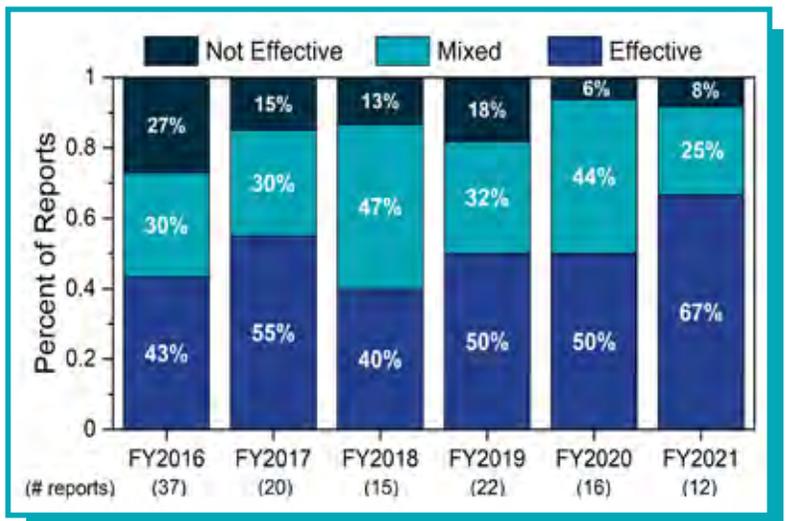


Figure 3. DOT&E Operational Effectiveness Trends

shortcomings when operating in particular environments, mission areas, or against specific threats. Programs that conducted early user testing, including operational assessments before Milestone C, were able to identify operational problems early, providing a greater opportunity to influence the design and make corrections prior to fielding. For example, the IVAS program conducted several Solider Touch Point events in order to test system prototypes in an operational, mission-based environment, and obtain early feedback from military users to support design refinements. In contrast, DOT&E has observed the consequences of not conducting early operational assessments. The F-35 program produced and fielded aircraft, avionics changes, and software releases prior to completing operational test (OT) and analysis. As a result, the OT and user communities continue to discover significant problems with the F-35, through both testing and actual employment in the field.

Suitability

In FY21, DOT&E assessed approximately half of programs to be operationally suitable without any caveats, a trend that has been relatively consistent since FY16. Suitability shortfalls were spread across Human System Integration (HSI), reliability, availability, and safety. Most notably, 80 percent of programs that assessed human factors reported HSI deficiencies. The most common causes of degraded HSI were training deficiencies resulting from incomplete or inaccurate documentation, poor usability, and high workload. Operators and maintainers frequently reported that they would benefit from additional hands-on training. Fifty percent of reports that included a determination on reliability found that the system was reliable enough to support the mission without caveats. Reliability shortfalls resulted from both hardware and software deficiencies. A larger percentage of reports found systems to be maintainable (71 percent) and available (77 percent) without caveats.

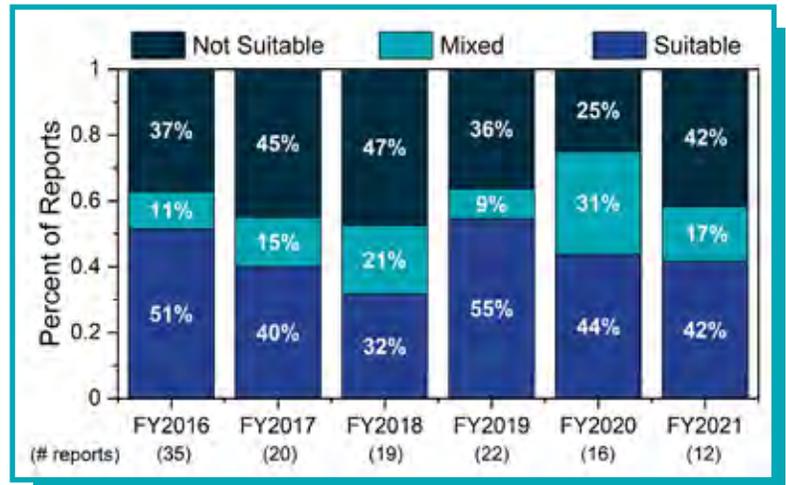


Figure 4. DOT&E Operational Suitability Trends

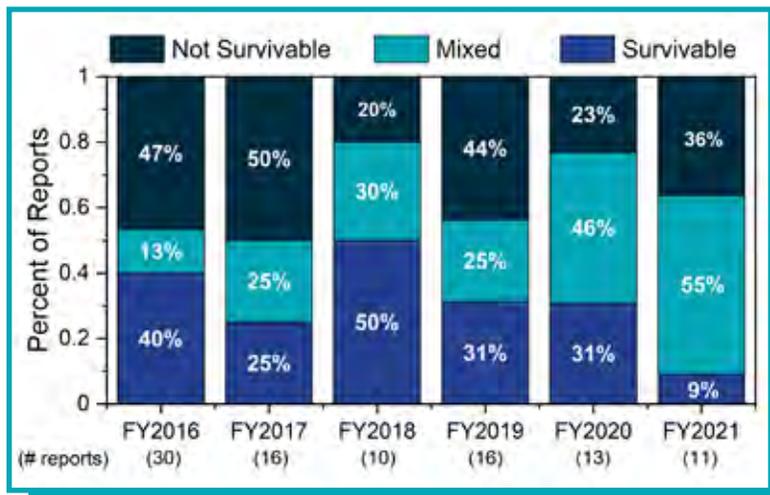


Figure 5. DOT&E Survivability Trends

Survivability

In FY21, DOT&E assessed nine percent of programs to be survivable without any caveats, a significantly lower percentage than in FY16. Given the complexity of the multi-domain operational environment, the cyber threats, and the contested electromagnetic spectrum environment, survivability assessments are becoming increasingly multi-faceted, and the fraction of programs demonstrating poor survivability has increased over time. Cybersecurity was the most common survivability problem. Cybersecurity issues included supply chain vulnerabilities, unencrypted software, and system-unique vulnerabilities to a wide spectrum of cyber threats. Other survivability shortfalls

included challenges with operating in a contested electronic warfare environment and vulnerabilities to specific kinetic threats unique to the system designs.

Recommendations

The following recommendations would better posture a program for success during operational testing:

1. Program managers should develop robust cybersecurity T&E strategies, which include an assessment of supply chain vulnerabilities; consideration of cybersecurity in the design phase to reduce potential attack vectors; collection of data to evaluate mission effects and the ability to prevent, mitigate, and recover from attacks; sufficient coverage of the system's attack surface; and early correction of deficiencies to improve the likelihood of being assessed as survivable during operational testing.
2. Program managers should develop adequate M&S, as a complement to live testing, supported by an independent VV&A process that uses credible and relevant data. M&S is increasingly necessary for development, integration, and mission-level evaluation due to the complexity of DOD systems, the importance and difficulty of representing complex operating environments, and the growing sophistication of our adversaries' weapon systems.
3. Program managers should ensure adequate rigor of HSI assessments by evaluating HSI early in the design phase and throughout development so that deficiencies can be discovered and addressed prior to operational testing. Program managers should also plan for sufficient operator and maintainer training commensurate with the level of system complexity. For many systems, the degree of hands-on and unit collective training should be expanded, and more attention should be paid to improving reliability and developing, refining, and validating operator and maintenance manuals prior to operational testing.
4. Program managers should conduct early, operationally realistic test events, including Operational Assessments, Limited User Tests, and Integrated Testing, where possible. When conducted early in a program's development and when adequately resourced across the acquisition cycle, operationally-realistic T&E offers a unique opportunity to identify and correct problems before the program matures. Early problem discovery allows the program manager to manage cost and schedule later in the process, and fix problems early so that they are not discovered for the first time in the final operational test, the field, or worse, in combat. For this to work, program managers must structure their contracts to require demonstration of operationally relevant, mission-level goals during early testing, instead of focusing solely on specification compliance.

