EXECUTIVE SUMMARY

MAJOR PRODUCTS

In FY22, DOT&E provided operational and/or live fire test and evaluation oversight for 243 acquisition programs at various stages in their acquisition cycle.[1] Specifically, DOT&E reviewed and approved 27 Test and Evaluation strategies / Test and Evaluation Master Plans (TEMPs), 9 of which included a Live Fire Test and Evaluation (LFT&E) Strategy. DOT&E also approved 68 individual test plans and disapproved 3 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 adequate evaluation of operational effectiveness and operational suitability; 2) coverage of the battlespace and threats; 3) credible use of modeling and simulation (M&S); 4) complete assessments of system survivability and lethality against mission-relevant threats (e.g., kinetic; cyber; electromagnetic; and Chemical, Biological, Radiological, Nuclear, and Explosives (CBRNE)); 5) production representative test articles; 6) operational realism; and 7) sufficient funding and resources required to support test execution.

In FY22, DOT&E published 48 reports, including 37 reports to Congress and the SECDEF, and a classified annual report on the Missile Defense System. In addition to the assessment of test adequacy, DOT&E reports summarize the Director's independent assessment of operational effectiveness, lethality (where relevant), suitability, and survivability of DOD weapon and business systems in realistic operational conditions. In instances where operational and/or 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/or live fire test, prior to fielding or the next acquisition decision review. DOT&E reports include practical recommendations to fix the identified deficiencies and improve the operational performance of the weapon or business system in expected operational scenarios and conditions to minimize risk to warfighters and maximize probability of mission success in conflict.

In FY22, DOT&E responded to several National Defense Authorization Act (NDAA) tasks and other Congressional taskers, the status of which is summarized in Table 1.

1 The number of programs on DOT&E oversight fluctuates throughout the year; 243 is the number of programs on DOT&E oversight as of September 30, 2022.

Table 1. Summary of DOT&E Congressional Activities

Source	Title	Status
FY20 NDAA		
Sec. 231	Digital Engineering Capability to Automate Testing and Evaluation	Ongoing; DOT&E in support of USD(R&E)
FY21 NDAA		
*Sec. 112	Report on limitations of Integrated Visual Augmentation System (IVAS)	Complete
Sec. 159	Documentation Related to F-35 Program	Ongoing
Sec. 162	Briefings on Software Regression Testing for F-35	Ongoing; USD(A&S) develop quarterly briefings in consultation with DOT&E
Sec. 222	Activities to Improve Fielding of Air Force Hypersonic Capabilities	Ongoing; USD(R&E) to deliver report in consultation with DOT&E
FY22 NDAA		
*Sec. 115	Limitation on Availability of funds pending report on the Integrated Visual Augmentation System	Complete
Sec. 223	Development and implementation of digital technologies for survivability and lethality testing	Ongoing; program selection complete
*Sec. 235	Limitation on transfer of certain operational flight test events and reductions in operational flight test capacity	Complete
*Sec. 1046	Comparative testing reports for certain aircraft	Complete
Sec. 1529	Demonstration program for automated security validation tools	Ongoing
Other FY22 Congressional Taskers		
*Appn ES pg. 5-6	Department of Defense Test Infrastructure Investments: Detailed Spend Plan	Complete
*Appn ES pg. 12	Certification of test strategies on Middle-Tier Acquisition and Rapid Prototyping programs	Complete
*Appn ES pg. 119-120	Self-defense test ship Congressional Response	Complete
*Appn ES pg. 138-139	Certification of funding for test infrastructure and test event resources	Complete
SASC Report pg. 191-192	Electronic Health Record interoperability between DOD and Veterans Affairs	Ongoing
HASC Report pg. 54	Commercial Virtualization Technology briefing	Complete
*HASC Report pg. 70	Digital twin assessment and agile verification processes report	Complete

Table 1. Summary of DOT&E Congressional Activities

Source	Title	Status
*HASC Report pg. 268	Software academic technical expertise implementation plan	Complete

Appn ES – Appropriations Act Explanatory Statement; HASC – House Armed Services Committee; NDAA – National Defense Authorization Act; SASC – Senate Armed Services Committee; USD(A&S) – Under Secretary of Defense for Acquisition and Sustainment; USD(R&E) – Under Secretary of Defense for Research and Engineering

* These activities resulted in reports to Congress which are reflected in the Appendix.

Lastly, DOT&E published the DOT&E Strategy Update 2022, outlining the intent to transform T&E and enable delivery of the world's most advanced warfighting capabilities at the speed of need. Driven by challenges caused largely by software-reliant systems, artificial intelligence (AI) and machine learning, Joint All-Domain Operations, data management, speed to field, culture, and talent management, this strategy seeks to advance the T&E infrastructure, processes, tools, and workforce needed to meet the T&E demands of the future. The Strategy intends to deliver on this intent by focusing on five pillars: 1) Test the way we fight; 2) Accelerate the delivery of weapons that work; 3) Improve the survivability of the DOD in contested environments, 4) Pioneer T&E of weapon systems built to change over time; and 5) Foster an agile and enduring T&E enterprise workforce. An accompanying Implementation Plan, treated as a living document that DOT&E will update annually in coordination with the T&E community, clarifies the desired end-state and specific actions and deliverables proposed to contribute to the accomplishment of the strategic intent.

MAJOR CONTRIBUTIONS

» ENSURING ADEQUATE TESTING IN COMBAT-REPRESENTATIVE CONDITIONS

In FY22, 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 FY22, common shortfalls were associated with deficiencies with M&S verification and validation (V&V), insufficient coverage of the operational environment and threats, including insufficient threat realism for cyber assessments, and inadequate data collection to support an evaluation of operational performance. DOT&E also noted a lack of production representative systems due to differences in software between the fielded system and the system under test. DOT&E worked with program stakeholders to improve the test adequacy of plans.

In FY22, at DOT&E's request, the National Academies of Sciences, Engineering, and Medicine completed their study on the health and readiness of the DOD test ranges and associated infrastructure for future operational and live fire testing. The National Academies' resulting classified Phase II report was published in August 2022, and expands upon previous findings summarized in the unclassified, Phase 1 report, published in September 2021. Specifically, it focuses on 1) improving threat modeling and prototyping to keep pace with the adversary; 2) addressing gaps in testing driven by new and emerging technologies; 3) testing as you fight with a focus on operational capability, not technical requirements; 4) formalizing a test range for multi-domain operational test at a system-of-systems level, based on live, virtual, and constructive technologies; and 5) testing at the speed of operational needs. DOT&E continues to evaluate the National Academies' recommendations and will include many of them, as appropriate, in the Implementation Plan for the DOT&E Strategy.

In parallel, through the DOT&E Resources and Infrastructure Working Group, and in coordination with the Test Resources Management Center in

USD(R&E), DOT&E received funding to upgrade T&E capabilities in support of next-generation weapons, including hypersonics, directed energy, and space technologies, and to improve realistic threats through verified and validated threat model surrogates. Additional details are available in the T&E Resources section of this report.

» ENSURING ADEQUATE TESTING ACROSS EVERY ACQUISITION PATHWAY

In FY22, DOT&E, in close coordination with the Director of Developmental Test, Evaluation and Assessments within USD(R&E), published a T&E Enterprise Guidebook to replace the current Defense Acquisition Guidebook, Chapter 8 and provide the DOD's acquisition and T&E communities with detailed guidance on adequate developmental, operational, and live fire T&E for each of the acquisition pathways. Across all acquisition pathways, the T&E Enterprise Guidebook emphasizes 1) the need for early and active engagement in acquisition programs to inform requirement development and acquisition contracts; 2) the use of any and all test events and data collection opportunities to support assessment of technical and operational performance; 3) the establishment of data storage and management processes to build accessible data repositories to support timely evaluations; and 4) the use of digital engineering and automation tools, supported by rigorous V&V processes, whenever possible for T&E planning, analysis, and reporting.

In FY22, DOT&E evaluated the DOD and Service test resources and funding profiles needed to support agreed-upon TEMPs for Major Defense Acquisition Programs and T&E strategies for prototyping programs.

Fifty-one of 101 programs (51 percent) were found to have adequate funding to support the remainder of the planned test execution. One program, the CH-53K King Stallion, was identified as having funding shortfalls related to survivability testing planned in the live fire plan. DOT&E also identified 34 of 101 programs (34 percent) that required updated TEMPs or T&E strategies due to program changes and thus, may have new or altered testing or resource requirements. Fifteen of 101 programs (15 percent) have fully executed all required testing; no current or Future Years Defense Program funding is required or allocated.

DOT&E also assessed the appropriateness of the test strategies for 105 programs approved by the Service Acquisition Executives to pursue accelerated acquisition authorities. DOT&E received and reviewed 53 test strategies and determined 41 of those to be appropriate. DOT&E's assessment was based on the test strategy supporting demonstration of the maturity and feasibility of the system to achieve the required capability, mission-relevant system capabilities and limitations with planned operational units, operators, missions, and environments, and system survivability against mission-relevant threats. DOT&E also assessed whether the test strategy identified the funding required to support the test execution.

» TRANSFORMING T&E

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, enterprise-level approaches to enable realistic testing. Improvements in infrastructure, tools, processes and the T&E workforce are needed to ensure adequate characterization of joint warfighting concepts and support delivery of the most advanced technical capabilities at the speed of need. The initiatives required to accomplish this are summarized in five major pillars of the DOT&E Strategy Update 2022.

1. Test the way we fight

Accurate evaluation of warfighting capabilities requires an adequate, scalable, and adaptive representation of the multi-domain operational environment as well as the ability to measure the operational performance of the future Joint Force capabilities in such an environment. DOT&E seeks to enable the identification, prioritization, and tracking of key range capability and funding requirements as driven by emerging technologies and threats.

DOT&E also seeks to define the OT&E and LFT&E requirements needed to support an operational evaluation of DOD scenarios, vignettes, and mission threads, including kill webs, in addition to evaluation of individual acquisition systems within those mission threads. In FY22, DOT&E highlighted key OT&E and LFT&E gaps in test range capabilities, instrumentation, and threat representation. In addition, given the growing importance of M&S and other virtual representations, DOT&E considered the challenges preventing rigorous V&V of M&S used to supplement findings from live test events. To improve M&S V&V adequacy, DOT&E, in conjunction with USD(R&E), is developing new M&S V&V policy to enable increased use of M&S in operational evaluation. However, availability of test data needed to enable credible V&V of key M&S tools remains a challenge.

2. Accelerate the delivery of weapons that work

As the complexity of systems grow, so often does the amount of data needed to support an adequate evaluation of their operational performance. For the T&E community to optimize the use of large volume of data, it must accelerate the implementation of the DOD Data Management Strategy and the five data decrees. To increase T&E and acquisition efficiencies, including automated, near real-time, enterprise-level data management and analysis to drive new insights, T&E data must be contained in data repositories that are discoverable, accessible, and secure. Internally, in FY22, DOT&E continued working a proof of concept to improve T&E data management with goals of improved searchability and the ability to document trends in findings across reports.

DOT&E is also exploring ways to better implement digital data practices across the DOD as a way to operationalize the "Shift Left" approach. By understanding a system's performance throughout the life cycle with increased access to data, the T&E community can use tools common in digital engineering and Bayesian inference processes to more efficiently integrate testing and analyze results. DOT&E hosted a workshop with the T&E community focused on exploring methods for taking advantage of digital engineering principles in the T&E planning process. DOT&E is using feedback from across the Services to understand where model-based TEMPs and T&E strategies may be most beneficial as a tool for efficiently documenting test planning and tracking execution.

DOT&E is working with USD(R&E) on updated policy for TEMPs and T&E strategies. The updated policy places a new emphasis on the Integrated Decision Support Key as a tool that finds opportunities for shifting left through integrated test designs (i.e., T&E events designed to meet developmental, live fire and operational test objectives) that provide useable data for multiple evaluations. This policy 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. The Integrated Decision Support Key also lays the groundwork for further research into statistical application of sequential methods, development or relational databases, and similar tools that could be used to optimize the use of available data collected under different but relevant conditions.

3. Improve the survivability of the DOD in contested environments

To improve DOD survivability, DOT&E is committed to helping minimize the number of mission critical vulnerabilities in fielded systems through continuous mission-based risk assessments and rigorous evaluations, as both U.S. systems and the threats they face evolve. DOT&E continues to find that many programs are not survivable against operationally relevant threats. As systems become more interoperable, the attack surface increases exponentially, while also complicating the assessment of synergistic effects of multiple threats across a given mission thread. In FY22, DOT&E considered the survivability of systems against the full spectrum of potential threats. In our initial response to FY22 NDAA Section 223, we chose four programs to serve as pilots for full spectrum survivability testing approaches: 1) the Army's Future Long Range Assault Aircraft, 2) the Air Force's LGM-35A Sentinel (Ground Based Strategic Deterrent), 3) the Navy's DDG 51

Flight III, and 4) the Joint F-35 program. Each of the programs uses some form of digital technologies, in addition to live testing, for their evaluations. DOT&E will continue exploring the overall utility of live, virtual, and constructive methods in standardizing the assessment of mission-based survivability in FY23 as part of a more detailed response to the Section 223 task.

DOT&E continues to emphasize cyber and electromagnetic spectrum survivability as attack surfaces multiply. DOT&E also sees space as an increasingly congested and contested environment. In FY22, DOT&E, in conjunction with USD(R&E), drafted an update to cyber T&E and electromagnetic spectrum operations T&E policies. The T&E community must consider supply chains, software factories and pipelines, and an array of cloud solutions in survivability assessments. DOT&E observed in FY22 that test teams across the Army, Navy, and Air Force demonstrated the capability to test non-IP systems during operational test. Other new tools and processes may be necessary to effectively assess countermeasures and other self-defense solutions.

4. Pioneer T&E of weapon systems built to change over time

AI-based systems require the T&E community to continuously monitor and evaluate the system' behavior, including in theater post-fielding, to ensure their ethical, effective, and safe use as they're exposed to new operating environments that they may have not been trained and tested on. Similarly, with the increased prevalence of software-reliant programs planning to deliver capabilities more frequently, DOT&E has considered how some T&E practices may need to evolve to enable continuous monitoring and evaluation of their operational performance. DOT&E is committed to enabling operationally relevant and timely evaluations for these rapidly changing systems to ensure they continue to be effective, lethal, suitable, and survivable as both they and the threat change. In particular, DOT&E plans to issue policy on T&E of software and AI-enabled systems. The policy places emphasis on early involvement in requirements development and program planning. It encourages early integration of end users to understand the complex dynamic of how warfighters use systems to accomplish their missions and how those might supplement automated test results.

The use of digital twins and commercial virtualization technology may be a means to enable continuous evaluation of operational performance. Such technologies may optimize the use of model-based system engineering supporting early, continuous, and automated T&E across the life-cycle of the system. In FY22, DOT&E responded to the House Armed Services Committee (HASC) task on digital twin practices and commercial virtualization technology, assessing that that a consensus on what constitutes a digital twin for the purposes of OT&E and LFT&E has not yet been reached. DOT&E also assessed that only a small fraction (about 7 percent) of the acquisition programs currently on DOT&E oversight have developed or are developing some version of a digital twin. While use of digital twins for OT&E and LFT&E is not yet common practice, DOT&E is researching the work on the verification, validation, and accreditation of these along with the benefits and challenges of these approaches.

5. Foster an agile and enduring T&E workforce

DOT&E requires a trained and equipped workforce, prepared to meet the toughest T&E challenges, with access to continuous learning opportunities. In June 2022, DOT&E completed a workforce assessment to identify existing strengths and weaknesses and clarify follow-on actions needed to ensure that DOT&E is optimally structured, organized, and postured to meet the demands of the future. As a result, DOT&E developed core DOT&E workforce competencies that will be used to develop and implement a continuous learning curriculum and future operating model to support DOT&E's future mission execution and keep pace with the evolving T&E environment. Time to train, speed to hire, and easy access to in-demand expertise remain a challenge.

In addition to the workforce assessment, DOT&E partnered with Cyber Test Teams across the Services to complete the Software and Cyber Network of Excellence for Testing (SCyNET) pathfinding activities.

In July 2022, DOT&E responded to the HASC task by providing an implementation plan for a Cyber and Software Test and Evaluation Center of Excellence. The plan provides a strategy for a scalable, one-stop-shop for cutting edge research and development, thought leadership, and collaboration across the T&E enterprise with three major objectives: 1) nurture a culture of information exchange across the T&E enterprise; 2) drive continuous innovation across cyber and software T&E; and 3) build a pipeline for talent acquisition, training, and retention through multiple key actions.

DOT&E also responded to an NDAA Section 235 tasker to evaluate the effect of the Navy's proposed manpower reductions on naval aviation OT. DOT&E found that OT squadrons (VX-1 and VX-9) have adequate physical and organizational infrastructure but, if the proposed manpower reductions get implemented, the squadrons' capacity utilization will exceed 100% precluding either squadron to complete the planned and required operational testing. DOT&E recommended against delegating OT to non-OT units citing several reasons that could cause a negative effect on the cost, schedule, and capacity of Naval Aviation OT. Instead, DOT&E recommended that the VX-1 OT squadron maintain its FY22 manning and that VX-9 OT squadron add 106 maintainers to its FY22 manning.

» DEMONSTRATING THE VALUE OF T&E

T&E is essential to demonstrate weapon system performance and provide DOD mission planners, commanders, and operators and maintainers with an understanding of true system capabilities to adequately plan and execute their missions. 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 cyber-attacks. 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. Examples of common problems discovered in OT&E include poor system performance, poor interoperability with Joint Partners, poor human systems integration, insufficient training, and various hardware failures. The performance trends section below provides additional detail on the value of T&E.

MAJOR FINDINGS

Figures 1 through 4 summarize the trends in DOT&E assessments of test adequacy, operational effectiveness, operational suitability, and survivability since FY16. While DOT&E published 48 reports in FY22, 28 reports focused on acquisition programs. Of those 28 reports, not all included an assessment of final determination of operational performance due to either maturity of the program, test limitations, or multi-level classifications. More specifically, all 28 reports included an assessment of test adequacy,15 included an assessment of operational effectiveness, and 13 included an assessment of operational suitability and survivability. As discussed below, operational testing continues to reveal challenges with effectiveness, suitability, and survivability that would not be observed by developmental testing alone.

» TEST ADEQUACY TRENDS

In FY22, DOT&E reported that 75 percent (21of 28) of programs conducted adequate operational and/or live fire testing, as detailed in Figure 1. This was similar to prior years where the fraction of programs conducting adequate testing ranged from 57 to 74

percent. The majority of programs (6 of 7) assessed as not adequate or partially adequate were early fielding reports for programs that did not complete operational testing prior to fielding. DOT&E assessed one program as partially adequate because the test did not include all relevant threats and operational environments.

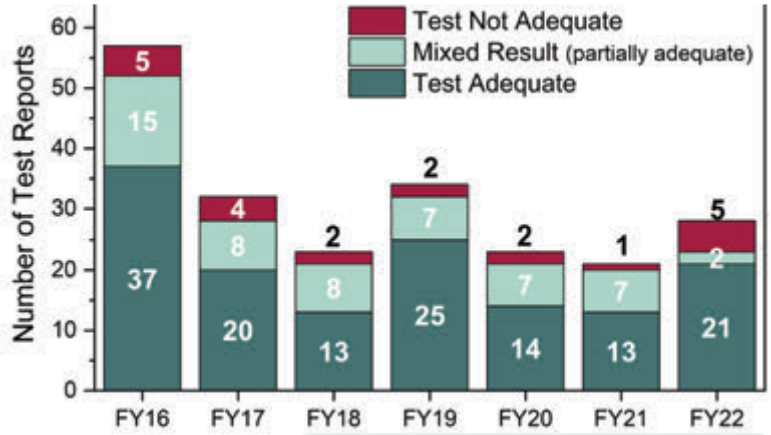

Figure 1. Test Adequacy Trends

Of the 28 reports, 27 noted at least one test limitation. Cyber survivability was the most common category of test limitation. Common cyber test limitations included testing that did not cover all cyber threat postures or attack vectors, such as supply chain compromise or outsider postures; failure to collect all data due to insufficient time or resources; lack of production-representative assets during early tests; and deferral of operational cyber testing to a later date. Other common test limitations included insufficient coverage of the threat environment or operational profiles, M&S deficiencies resulting from simplifying assumptions, inability to collect all required data due to test instrumentation limitations, and test range restrictions that prevented full employment of the system or threats.

» PERFORMANCE TRENDS

Effectiveness

In FY22, DOT&E evaluated 73 percent (11 of 15) of programs to be operationally effective. Since FY16, the fraction of programs assessed as operationally effective has ranged from 43 to 73 percent. DOT&E assessed four FY22 programs as not effective or having mixed effectiveness because of shortcomings when operating in particular environments, mission areas, or against specific threats. For example, one system was assessed as not effective because of performance deficiencies when employing the system at night. In another case, operational testing revealed that the units did not use the system as envisioned during developmental testing. Specifically, when used as a mobile command post, the system did not have enough secure beyond line of sight communication networks to support communication demands.

All 15 reports with an operational effectiveness assessment documented at least one problem with operational effectiveness. In several cases, operational testing of the full system of systems revealed important interoperability or integration deficiencies, such as a communication system that exceeded the bandwidth requirements of the tactical network it was operating on or a tracked vehicle that was not able to share target information with infantry target designators. Another common problem was human factors limitations that affected operator performance or unit effectiveness. In one example, the unit was not able to use the system effectively because of its complexity and the lack of user training prior to the test.

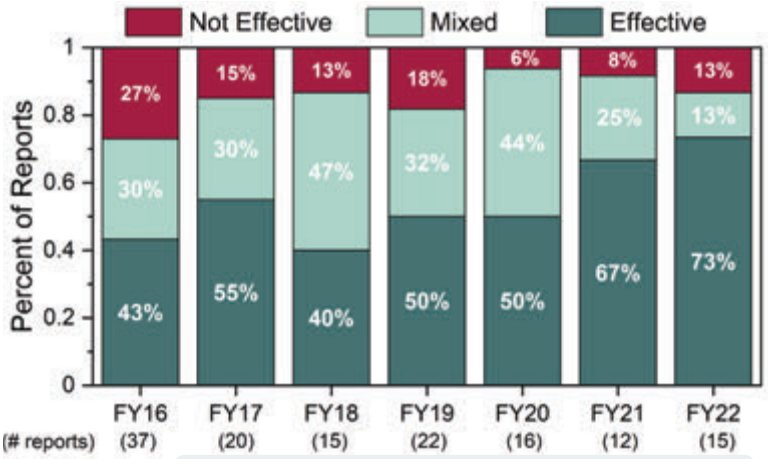

Figure 2. Operational Effectiveness Trends

Suitability

In FY22, DOT&E evaluated 38 percent (5 of 13) of the programs to be operationally suitable without any caveats. All six programs assessed as not suitable in FY22 had poor reliability resulting from a mix of

software and hardware failures. In three instances, availability and maintainability shortfalls also played a role. Of the 13 programs with a suitability assessment, 10 programs had Human System Integration (HSI) challenges. Similar to FY21, lack of adequate training or training resources continues to be the primary HSI deficiency.

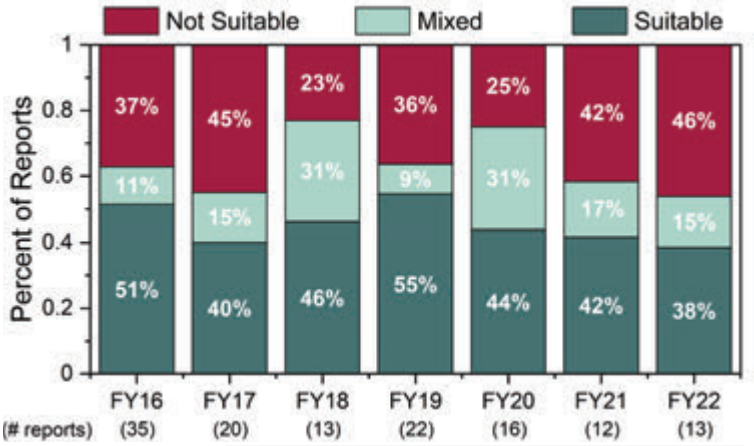

Figure 3. Operational Suitability Trends

Survivability

DOT&E evaluated 23 percent (3 of 13) of the programs to be survivable without any caveats in FY22. Similar to FY21, survivability against cyber threats was the most common problem followed by survivability against kinetic threats.

Common cyber survivability issues included unencrypted software, hardware, or network traffic; lack of safeguards to limit access to serial, USB, or Ethernet ports; and use of a 1553 data bus without encryption or authentication. In most cases, operators were the primary cyber defenders of the system and the system lacked the capability to detect, monitor, or notify the operator of a potential cyber attacks. Eight systems had vulnerabilities to specific kinetic threats unique to the system designs.

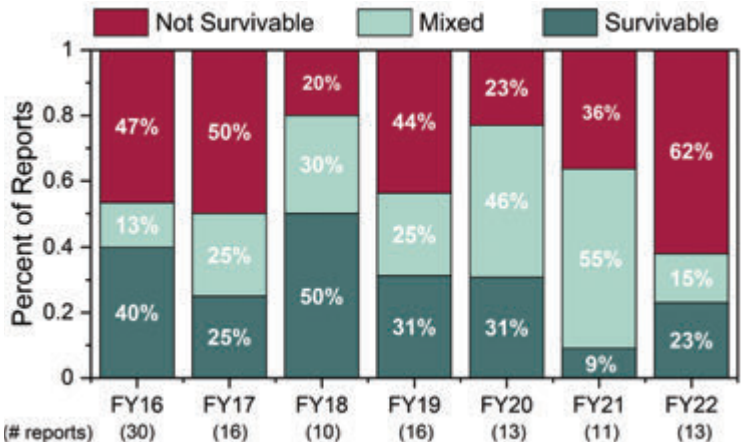

Figure 4. SurvivabilityTrends

RECOMMENDATIONS

The following recommendations are expected to better posture a program for success during operational and live fire testing:

- Integrate test planning and execution across the T&E community to increase efficiency and discover problems early by requiring demonstration of operationally relevant, mission-level goals during early testing, instead of focusing solely on specification compliance.
- Conduct operational testing that supports an assessment of the full system of systems across the relevant set of missions and operating conditions.
- Follow best practices early in the acquisition phases of a program to avoid common cyber vulnerabilities and build systems that are capable of detecting, monitoring, and notifying operators of cyber attacks.
- Establish a reliability growth process that is supported by system engineering efforts and contractual requirements.
- Refine and validate training manuals and other training resources prior to operational testing and allocate more time for operator and collective unit training.
- Develop robust and independent V&V for all M&S to be used in T&E.