FY 2020 Annual Report

The United States military continues to be the strongest and most talented force on the planet. Our women and men in uniform – all volunteers – remain committed to the Constitution, preserving American freedom and prosperity, and supporting our allies. Their success in this most fundamental mission reflects their intelligence, bravery, and dedication to their fellow Americans. It also reflects the capabilities we place in their hands.

The operational and live-fire test and evaluation communities hold a most solemn responsibility: independently assessing those capabilities for effectiveness, suitability, survivability, and lethality in near-real-world combat conditions. Our evaluations determine whether a production-representative system does what it's supposed to, whether the warfighter can use it safely, and whether the warfighter can depend on it in combat.

DoD's operational and live-fire T&E have been sufficient to provide accurate information to decision makers in the department and on Capitol Hill, and to users – American warfighters, our national treasure. As global threats grow, however, with near-peer adversaries closing the capability gap, and the number and severity of potential attack vectors rapidly expanding, the very fundamentals of operational and live-fire T&E must be examined: Does the Defense Department have the right tools, infrastructure, processes, and people to properly evaluate the extraordinary technologies we plan to field next year and more than 10 years from now? Are we testing the right aspects of our systems and putting enough focus on the types of realistic threats and vulnerabilities our adversaries are likeliest to exploit? Is T&E prepared to adapt to global conditions in real time? How can testing streamline the acquisition process?

DOT&E is delving into these issues and, it appears, the time for significant change has arrived. As good as operational and live-fire test and evaluation are today, we must make them better – more effective, more efficient, more robust, and more flexible. We also must create a holistic set of capabilities and infrastructure to ensure that our newest branch, Space Force, can benefit from the same independent, rigorous assessments as its sister Services. Bringing T&E into the 21st century will require substantial investment, a different approach to acquiring expertise, and intragovernmental support of the live-fire and OT&E mission. That commitment of resources, time, and energy will pay enormous dividends for our women and men in uniform.

FY20 HIGHLIGHTS

Integrating Developmental and Operational Test & Evaluation

We are now 20 years into the 21st century but, in many ways, DoD acquisition functions appear to be stuck in the 20th century. Our processes are too slow. By the time many of our systems roll off the production line, the requirements against which they were designed are decades-old and no longer capture the threat or warfighter needs. With our near-peer adversaries rapidly gaining ground, and even getting ahead of us in certain areas, continuing along this path is dangerous!

To help make development and fielding more dynamic, in 2020 DOT&E and the developmental test (DT) community took the first steps to integrating DT and OT. DoD traditionally has executed test and evaluation in a segmented, sequential fashion. The strict DT-OT bifurcation is delaying getting weapons into the hands of the warfighter.

Test activities in key DoD programs, including the B-21, the VH-92A, the CH-53K, the MK-48 heavyweight and MK-54 lightweight torpedoes, submarine sonar systems, and many net-centric systems, are showing that the siloed, linear approach can be set aside – and that, by doing so, DoD can cut the time to field major weapon systems by as much as 40 percent. Developmental system configurations and conditions can yield OT-quality data for certain measures of effectiveness, suitability, and performance. Conducting incremental cyber assessments of each developmental system configuration, using the OT perspective, creates a cumulative body of evidence that enables more tailored and focused cybersecurity test events during initial OT&E (IOT&E).

A handful of guiding principles has emerged from these forays into DT-OT integration. Early DT-OT collaboration to shape DT plans is essential in order to maximize the opportunity for OT data collection during "dual-use" DT events. Similarly, the program must have a DOT&E-approved "early OT" concept prior to entering the engineering and manufacturing development phase. A collaborative, integrated-testing, data-scoring board, with program office, DT, and OT representatives, will approve each specific use of developmental and integrated test data for early OT reporting.

These process changes will not affect DOT&E's position as the sole independent source of authoritative OT&E data and findings. Dedicated IOT&E will still be necessary; not every OT requirement can be satisfied by early integrated test events. But, by gathering OT-type data and reporting it as soon as we know it, we can make testing more efficient and effective, and support better decision making.

F-35 and the Joint Simulation Environment

In FY20, F-35 testing crossed a major milestone, finishing planned open-air combat and electronic attack trials. Two IOT&E weapons test trials were scheduled for October 2020 and early calendar year 2021; a third weapons test included in the original test plan has been deferred to a later program phase.

A substantial amount of testing remains, and it cannot be executed until the Joint Simulation Environment (JSE) is ready. The JSE is a man-in-the-loop, software-in-the-loop mission simulator that will provide the only venue, other than actual combat, to test the F-35 against modern threats in realistic densities and mission scenarios. Development of the JSE is now more than three years behind schedule. In late fall 2020, the Joint Program Office projected that completion of the 64 mission trials planned for the JSE would slip to mid or late calendar year 2021.

The data to be gathered via the JSE are essential to test adequacy. DOT&E cannot write the statutorily required beyond low-rate initial production report until the 64 JSE trials have been completed and the data analyzed.

Once the JSE is fully functional and IOT&E finished, the JPO will need to focus on ensuring that it remains verified, validated, and accredited (VV&A) for the rapid software cycle planned for future blocks of the F-35. The continuous capability development and delivery model will produce a new software release every six months. As currently constructed, test plans do not appear to collect enough open-air flight data to conduct sufficient VV&A for Block 4 capabilities.

Longer term, DoD must explore maximizing our investment in JSE by adding other current and future air platforms, and by expanding its simulations to cover space and cyberspace. As with the F-35, DoD largely cannot test space assets or weapons system cybersecurity live or in operationally representative conditions. JSE's high-fidelity environment potentially could provide a venue to assess these critical operational capabilities against realistic threats.

PREPARING LIVE-FIRE AND OPERATIONAL TEST & EVALUATION FOR THE NEXT DECADE AND BEYOND

The next 10 years may prove to be the most challenging period in the history of live-fire and operational test and evaluation. The capabilities of near-peer competitors are advancing at breakneck speed. Many systems in our acquisition pipeline comprise technology never before fielded. The creation of Space Force brings to the forefront an increasingly crowded and contested domain. And the potential for harm, and even mission failure, as a result of cybersecurity failures continues to grow.

Are DoD Ranges Ready for the Future?

At the end of FY20, DOT&E engaged the National Academies of Sciences, Engineering, and Medicine (NASEM) to conduct an independent, objective, peer-reviewed study of the DoD test and training ranges used for live-fire and operational test and evaluation. The two-part study will assess the adequacy of ranges and associated infrastructure in the 2025-2035 timeframe to support DOT&E's statutory mission to establish a system's operational effectiveness, suitability, survivability, and lethality.

The first tranche of the study will examine test and training ranges' physical suitability, to include capacity / throughput, condition of infrastructure, security, and encroachment challenges; and their technical suitability, which includes instrumentation, cyber and analytic tools / algorithms, and modeling and simulation capabilities. NASEM will release an unclassified report on these areas to the public in summer or early fall 2021.

Concurrently, a second NASEM team will examine ranges' operational suitability. This includes threat and threat countermeasure replication and representation, which are crucial to both testing and training; capacity for advanced weapons; spectrum management; and infrastructure cybersecurity. The assessment of advanced weapons and threats will cover, but not necessarily be limited to, directed-energy weapons, hypersonic systems, 6th generation aircraft, autonomous systems, artificial intelligence, space systems and threats, and advanced active electronic warfare / cyber capabilities. The final report will be classified but available to DoD and the Congress.

Both reports will present conclusions regarding whether DoD test and training ranges can fulfill our anticipated needs. Importantly, each will also offer actionable recommendations.

The T&E Resources section of this report already notes multiple existing shortfalls. And, after three years of visiting our ranges and test facilities, I can offer this admittedly unscientific observation: The majority of our ranges were built around World War II (planes still fly over the same terrain at Eglin Air Force Base that the Doolittle Raiders used to train for their famous 1942 Japan raids); most were updated at the height of the Cold War in the 1980s; but little has been done since then. I anticipate that NASEM will independently determine the same and I strongly encourage DoD planners and programmers, as well as Capitol Hill, to start thinking now about how to make capabilities and infrastructure match our warfighters' and testers' needs.

Space Test & Training

Since last year's report, Space Force has made great strides in standing up our newest cadre of warfighters. In November 2020, Gen. John Raymond, Space Force commanding general, assigned Space Training and Readiness Command (STARCOM) responsibility for operational test and evaluation. DOT&E looks forward to collaborating with STARCOM as it grows and begins to crystallize OT&E and training processes and plans.

The creation of STARCOM comes at a pivotal moment. The likelihood that the next fight will occur in space and cyberspace before it goes kinetic is high. And, over the Future Years Defense Plan, DoD intends to spend nearly $100 Billion on space assets. Yet, the department has no operationally realistic way of testing space-based systems. Currently, DoD expects to spend less than 1 percent of space program acquisition funding on test infrastructure. DoD would be wise to invest significantly more than that to develop a National Space Test and Training Range (NSTTR).

To be operationally representative, the NSTTR threat array must include cyber, directed-energy, kinetic, and electronic-warfare threats, as well as natural hazards. This multi-layered capability would be multifunctional, as well, supporting development and validation of space-based warfighting tactics, techniques, and procedures, development of multi-domain operating concepts, and more effective warfighter training.

Space systems present a significant challenge. They form our data, command and control, intelligence, surveillance, and reconnaissance highways, and they constitute an unprecedented attack surface. Each component of the space system architecture is vulnerable to cyber threats: the orbital segment (the spacecraft itself), the terrestrial segment (ground equipment required to operate the spacecraft), and the link segment (which transmits data between and among the orbital and terrestrial segments using electromagnetic spectrum). All three of these elements must be demonstrably cyber-secure, and the testing community must have the right talent and tools to assess them properly.

Cybersecurity

Space-based platforms' need for stringent cybersecurity is emblematic of DoD as a whole. Nearly every warfighting and business capability is now software-defined. Simply put, the system – plane, ship, vehicle, radio, operations center, missile, satellite, health records management – doesn't work if the software doesn't work. We are likelier to upgrade a system by installing new software than by replacing hardware. Yet, cybersecurity often is treated as an add-on feature, rather than being "baked in"; and our ability to assess and protect software, though improving, is not keeping pace with our reliance on it or our adversaries' ability to compromise it. In FY20, 62 percent of test plans noted cybersecurity testing limitations. Over the last several years, cybersecurity flaws have been the most common reason DOT&E determined a system to be not completely survivable.

Every program manager and every tester must be able to answer the same basic questions: How good is our software, and how do we know? How do we know our systems are secure? How do we know when we are being hacked, or when something anomalous has occurred in our software? How do we test to ensure that we minimize the maximum regret? And, with deference to the taxpayer, how much will the software cost over its lifetime, including updates and continuous testing of those updates?

Some aspects of cybersecurity OT&E are improving. Operational test agencies have broadened and made more rigorous the testing of systems that rely on Internet Protocol. More organizations are requesting assistance from DOT&E's Cybersecurity Assessment Program, which focuses on defense against advanced threats. And, the T&E community is strengthening cybersecurity testing processes; new guidance should be released in FY21.

Significant cybersecurity T&E gaps remain, however. Tools and techniques necessary to test specialized protocols, such as those in industrial control systems, tactical data links, and aircraft transponders, are not adequate. DOT&E is growing capabilities to execute threat-realistic cyber assessments against these technologies. In addition, DoD must ramp up realistic T&E of offensive cyberspace operations capabilities and procedures to give commanders confidence in their availability and efficacy. Test and evaluation of the junction between cyber and electromagnetic spectrum operations, and the burgeoning threat vector of cloud-based computing, must be augmented, as well.

More fundamental, though, is DoD systems' inability to self-monitor continuously for anomalies: The user doesn't know the health of her system's software. The plethora of gauges in today's cockpits tells the pilot almost everything she needs to know regarding the status of her aircraft. The one parameter into which she has no insight is the plane's software – and she likely won't know until something catastrophic occurs.

With software driving nearly everything we place in the warfighter's hands, this information shortfall is no longer tenable. Red-teaming and cybersecurity vulnerability penetration assessments are good but the software "surface" is too large and the pace of operations too fast for humans to keep up. The warfighter needs a 24/7, automated, autonomous software monitoring and testing capability that alerts her to defects, malware, hacking, and other types of compromise and failure.

People Are the Key

With its dependence on software, the department faces a breathtaking human-capital requirement. Development of cutting-edge cybersecurity testing tools and processes, and preventative diagnostics; in-depth understanding of emerging adversary techniques and capabilities; and the innovation necessary to adopt, test, and manage systems fueled by artificial intelligence and machine-learning demand a skillset that does not exist in DoD today. And, the department cannot build it internally: DoD will always be outbid in salary and geographic and workplace flexibility by the private sector. We therefore must apply a different model to get the people we need in the information technology, software, and cybersecurity spheres.

To tap the necessary creativity, intellect, and deep domain expertise, DoD should establish a federated university-affiliated research center (UARC). Similar to the university consortium for applied hypersonics that the Deputy Undersecretary of Defense (Research and Engineering) launched in October, a cyber UARC would give DoD access to the top tier of academia and their supporting partners in the commercial world. Instead of a full-time, static, in-house workforce, DoD would reach into the UARC as needed. This talent pool, which already is breaking boundaries in software, IT, and cybersecurity, is the only means to keeping DoD and our warfighters ahead of our adversaries.

COVID-19: IMPACTS AND OPPORTUNITIES

Live-fire and operational T&E are critical elements of DoD's acquisition process. The T&E community does whatever it takes to ensure that the equipment the department intends to field has been thoroughly assessed and its performance is understood. COVID-19, however, made this year as challenging for testers as it was for the rest of the country.

To protect the health of our personnel and their families, DOT&E followed national guidelines and significantly restricted travel from the middle of March through the end of the fiscal year. Action officers participated only in events deemed mission-essential by the Services, such as CVN 78, CH-53K, F-35, KC-46, and Amphibious Combat Vehicle testing. DOT&E's primary federally funded research and development center, the Institute for Defense Analyses (IDA), similarly limited travel to tests, as well as other office-based support, in order to safeguard its employees.

These constraints, and changes the Services and agencies instituted in response to the pandemic, affected T&E for one-third of programs under oversight. Certain tests and associated activities were postponed; others went forward with a reduced scope or number of events. In some cases, the DOT&E action officer and/or IDA analysts could not attend a test or preparatory event in person. Just over 20 percent of events scheduled for support from the Center for Countermeasures slid from the third to the fourth quarter of FY20, another 20 percent were postponed until FY21, and two were canceled.

The number of DOT&E cyber events and assessments for efforts not under oversight also was substantially smaller due to cancellations and postponements by sponsors. The notable exception was DOT&E's persistent cyber operations team, which logged its highest demand and operating tempo ever, driven particularly by combatant command requests to help facilitate operations by off-site personnel.

A chart highlighting which programs were impacted can be found on page 9, and details of COVID-related changes are included in individual program articles.

The unexpected and sudden halt to normal business revealed a substantial gap in DoD's T&E capabilities. While sectors of the commercial world were able to quickly resume their production monitoring and acceptance testing via telepresence technologies, the Services, agencies, and DOT&E largely were not prepared to adapt to COVID reality. Without question, the live-fire and operational T&E communities need that flexibility. Even in the face of a global pandemic, national defense cannot stop, and that includes the test and evaluation on which our decision makers and warfighters rely.

With that in mind, last summer DOT&E began to explore the feasibility of remote participation in live-fire and operational T&E. Given the wide variety of systems under oversight, the worldwide distribution of test events, and the classification of the information they generate, we envision a remote presence suite that includes: a high-capacity, reliable, and very secure transport layer; extremely high-definition, real-time video, often of multiple locations; two-way, live audio; possibly capabilities that replicate other human senses; and multiple collaboration tools.

Many of these technologies already exist and now is the time to determine how well they work in the live-fire and operational T&E context. The first logical target for a remote / telepresence operational test is an IT system. Much of the OT data for IT systems already is collected remotely. Where beneficial, live screen and data sharing, and real-time video and audio that allow the evaluator to observe users in action and to speak to them, potentially would be enough to complete the toolkit. Live-fire testing of major platforms, such as tanks, also is an immediate candidate for remote presence. Again, secure live video and audio would be required. In addition, a small, remote-controlled device that crawls over, under, and inside to examine damage, perhaps paired with a virtual or augmented reality system, potentially could be used to replicate the in-person experience.

It's safe to say that remote T&E won't be possible for every type of event or every type of system, but we must launch proofs of concept now to start building this critical capability. While the end of the COVID pandemic may be in sight, DoD cannot forego this opportunity to prepare for the next existential crisis; continuity-of-operations capacity must be at the top of the department's objectives. And there will be a bonus: Remote presence will improve general efficiency and efficacy, as well.

Remote T&E will require potentially large technology and infrastructure investments across the entire department. Protecting the integrity of live-fire and operational T&E, the health of our personnel, and national security will be money well-spent.

THE ROAD AHEAD

Serving my country and my sisters and brothers in arms as the Director, Operational Test & Evaluation has been a tremendous honor, and one I did not take lightly. Warfighters rely on the test community to stand as unbiased, independent arbiters of system quality and performance. Our work allows them to adhere to the third imperative of combat: Believe in your equipment and weapons.

The women and men of DOT&E have fulfilled this duty exceptionally well over my three years in office. For their success to continue, as the volume of ever-more complex systems in the acquisition pipeline grows, the department must provide live-fire and operational T&E resources that match the mission. DOT&E will continue to explore ways to augment efficacy and efficiency. With the right support from our partners throughout the Defense Department and in Congress, DoD's live-fire and operational test and evaluation communities will keep America safe and strong.

Robert F. Behler
Director