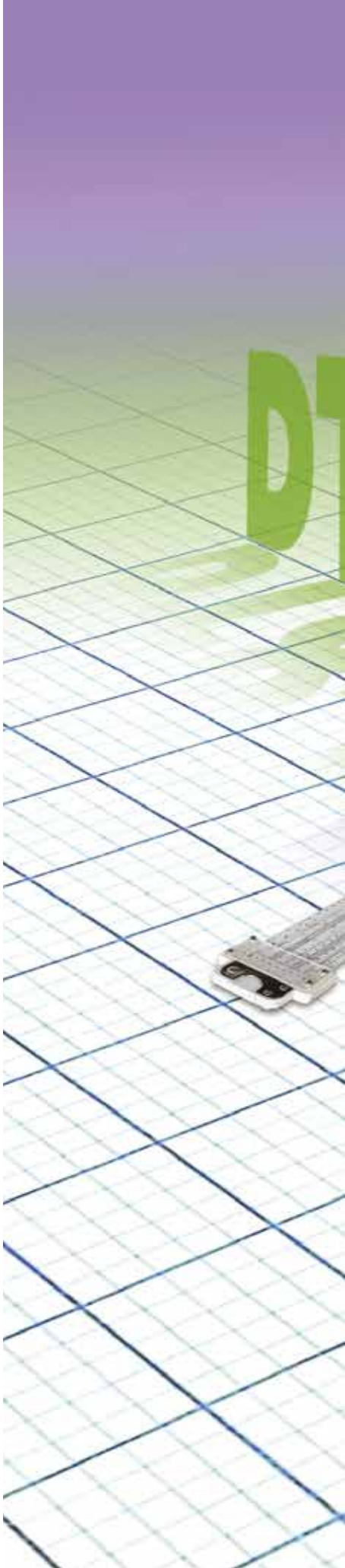
Test and Evaluation

Myths and Misconceptions

Steve Hutchison, Ph.D.



Test and Evaluation (T&E) is essential to successful system acquisition. For the last 43 years, the Office of the Secretary of Defense (OSD) has included various formations providing T&E oversight. Interested readers can review some of the history in the articles "The Original DT&E" and "What Happened to DT&E?" in the January–February 2014 and March–April 2014 issues, respectively, of the *Defense AT&L* magazine. Having been witness to just over a third of this history, I thought I would share some of the great myths and misconceptions about T&E that I have observed over the years. If we can dispel some of these myths, perhaps we can reduce the tension between testers and developers and get on with helping acquisition programs deliver capabilities more effectively and efficiently. After all, the Department of Defense (DoD) is not investing the nation's resources for programs to fail—our job as testers is to help programs succeed.

That actually might be one of the myths—that, because some testers are "independent," they actually are not supposed to "help" programs. I am going to take it on faith that most testers don't actually believe that; rather, even the most independent test organizations understand that it doesn't take a lot of talent to show up at the end of system development and point out the flaws. Instead, programs maximize their T&E Return on Investment (ROI) when their testers are engaged early, run meaningful tests and provide quick feedback to help move the program forward, not act as gatekeepers to block progress (the source of this idea is the book *Agile Testing: A Practical Guide for Testers and Agile Teams* by Lisa Crispin and Janet Gregory). The hard work of testing is not gatekeeping—it's providing constructive feedback. With that out of the way, I'll briefly count down my top five myths in T&E, and offer some thoughts on how to resolve them.

Hutchison *is director of test and evaluation for the Department of Homeland Security and previously served as the principal deputy for developmental test and evaluation in the Office of the Secretary of Defense.*


T&E
T&E
T&E
T&E

Myth No. 5: Only Operational T&E Matters

Many programs base their acquisition strategy on the belief that the only T&E that matters to decision makers is Operational Test and Evaluation (OT&E); after all, it's written in law—therefore, it must be the only T&E that matters. Title 10 USC §2399 "Operational test and evaluation of defense acquisition programs" stipulates that the Secretary of Defense may not permit Major Defense Acquisition Programs (MDAPs) to proceed beyond Low-Rate Initial Production (LRIP) until initial OT&E (or IOT&E) is completed and the Director of Operational Test and Evaluation (DOT&E) has submitted a report (commonly referred to as the "BLRIP report" [the B stands for "beyond"]), stating whether the operational test was adequate and the results confirm that the system is effective and suitable. Obviously, there is value in operational testing, particularly as the confirmatory activity stated above. However, the problem with this mandate is that it puts OT&E and the DOT&E in a gatekeeping role. Missing are the checks and balances prior to the start of production; in other words, feedback to programs is missing when it is needed most.

Once a program has formally entered the acquisition process, I would argue that the most important decision in the program life cycle is the decision to begin production. Program managers need to have it right at production start because, once the decision is made to begin production, designs are essentially locked and production fixtures set. If programs have not discovered and corrected design problems or key failure modes earlier, those problems will almost certainly become the warfighter's problems, because it will cost too much to correct them, and the tyranny of the urgent will demand that the capability get to the field. Permitting development problems to become the warfighter's problems is the real definition of acquisition malpractice. Thus, if you accept the premise that the most important decision is entry into production, then the T&E that matters most must inform that decision.

In the DoD process shown in Figure 1, the decision to begin production typically is made to authorize LRIP at Milestone C. Since 10 U.S.C. §2399 requires IOT&E to inform the full-rate production decision, acquisition decision authorities must rely on Developmental Test and Evaluation (DT&E) to inform the Milestone C decision. If programs get it right at production start, then OT&E will be that confirmatory activity described above rather than a discovery activity that tarnishes most operational test outcomes today.

There are a couple corollaries to this myth. They include:

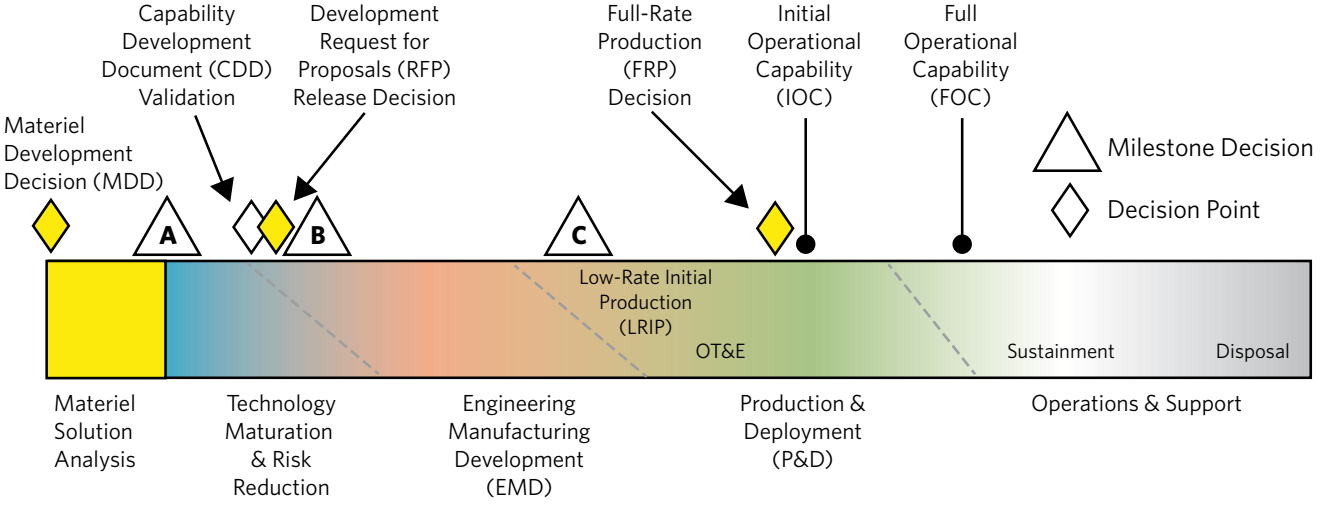
Corollary 1: DT&E is technical testing.
Corollary 2: Users aren't involved in DT&E.

These are the leading contenders for what I would call "T&E malpractice" and the reason so many programs discover problems during OT&E; hence the rallying cry to "shift left!" DT&E should never be considered just technical testing. Sad to say though, this is not myth. The *Glossary of Defense Acquisition Terms*, 15th Edition, December 2012, defines DT&E as:

> Any engineering-type test used to verify status of technical progress, verify that design risks are minimized, substantiate achievement of contract technical performance, and certify readiness for initial operational testing (see the full definition online at https://dap.dau.mil/glossary/).

If the developmental tester focuses only on assessing technical performance specified in the contract, programs will completely miss the sense of whether the capability could satisfy user needs in performing the mission. If, however, DT&E has a mission context, not only will programs and decision makers understand the technical issues, they also will obtain user feedback that is essential early in the life cycle, when there is time to adjust course if necessary. Mission context does not mean program managers have to shift the IOT&E to the left,

Figure 1. DoD Acquisition Life Cycle (Source: Interim DoD Instruction 5000.02)



but user involvement should be a DT&E priority. Using the Operational Test Agencies (OTAs) to help design and conduct mission-relevant developmental tests with typical operators would be a really good DT&E strategy. Ultimately, DT&E must employ the right resources to provide confidence in the decision to enter production.

Myth No. 4: Cybersecurity T&E Is Someone Else's Responsibility

I was an operator once, a boots-on-the-ground infantryman. My radio was perhaps the most valuable weapon in my arsenal; with it, I could change the terms of the current fight and the next engagement. Keeping my communications secure, and therefore keeping my mission parameters secure, was my responsibility. Technology has far exceeded the capability of those old radio days, but one thing remains unchanged: Security is an operator's responsibility. In the (dare I say it) "unfamiliar" cyberspace domain, providing "good" cybersecurity may well be today's most challenging development task. As testers, we put ourselves in the operator's boots to answer the "so what" question. So, when it comes to cybersecurity, why do we (sometimes) leave that part of the "so what" question for someone else to answer? It's an artifact of security processes that have become very specialized over the decades.

The assignment of cybersecurity responsibilities outside mainstream requirements and acquisition channels, not to mention outside the operator's channels, has many downstream impacts.

Beginning in the 1970s, DoD managed the acquisition of information technologies and their security requirements separately from the mainstream Defense Acquisition System and requirements processes. For example, the first DoD Directive (DoDD) 5000 formalized the acquisition process back in July 1971, but in October 1978 the Department issued DoDD 7920.1, Life Cycle Management of Automated Information Systems (AIS), and managed information technology under this separate acquisition process until eventually merging it with the DoDD 5000 in 1996. Security requirements appeared even earlier with the 1972 DoDD 5200.28 Security Requirements for Automatic Data Processing (ADP) Systems, reissued in 1988 as Security Requirements for Automated Information Systems (AISs), eventually becoming today's DoD 8500 series on Cybersecurity and the Risk Management Framework. These directives introduced another decision maker—the Designated Approving Authority (DAA)—with assigned responsibilities, many of which are still in use today. For example, the 1988 directive stated: "The accreditation of an AIS shall be supported by a certification plan, a risk analysis of the AIS in its operational environment, an evaluation of the security safeguards and a certification report, all approved by the DAA." In today's "risk management framework," the DAA is called an Authorizing Official (AO), and the AO retains responsibility for information security and approves the system authority to operate. To assist with these functions, the AO designates a Security Controls Assessor (SCA) to perform the checks of security controls. The SCA typically is not one of the program's DT&E or OT&E organizations.

The assignment of cybersecurity responsibilities outside mainstream requirements and acquisition channels, not to mention outside the operator's channels, has many downstream impacts. Since the modus operandi in the T&E community is to test to requirements, when cybersecurity considerations are absent from operational requirements documents they likely also will be absent in the T&E Master Plan (TEMP), DT&E and OT&E event test plans, and the test reports. The downstream effect is that the "cyber so what" question may not be adequately answered at critical acquisition decision points.

Cybersecurity is an operator's responsibility; therefore, it is incumbent on the T&E community to answer the "cyber so what" question: Does this new capability operate securely in the cyberspace domain? Our challenge is to fully integrate cybersecurity into our test processes to help programs identify risks, minimize the attack surface and reduce kill chain effects to improve resilience. Cybersecurity should be integrated into every test activity and inform acquisition decision making. In the summer of 2013, the Deputy Assistant Secretary of Defense (DASD) for DT&E and the DOT&E offices collaborated to produce a set of procedures for cybersecurity T&E that would go a long way toward helping testers develop and execute such plans and help programs close the gap between authorities to operate and operating securely.

Myth No 3: OTAs Can't Do DT&E

OTAs have often told me that they can't do DT&E (as in "not permitted" to do DT&E as opposed to lacking competence to perform DT&E). I'm not sure how this myth came to be, but unless the Component T&E regulations actually prohibit the OTAs from conducting DT&E, then it simply remains a myth that OTAs can't do DT&E.

The idea may have originated as an extension of statutory language limiting DOT&E involvement in DT&E. Specifically, 10 U.S.C. §139 (d) states that the DOT&E "may not be assigned any responsibility for developmental test and evaluation, other than the provision of advice to officials responsible for such testing." Component acquisition authorities may simply be extending this limitation to their OTAs, perhaps to protect their independence—the idea being that, if an OTA is involved in DT&E, it is not independent. That's just absurd. Independence seeks to ensure that an agent separate from the developer and user perform the test and evaluation; it has nothing to do with when the tester is involved or the type of testing performed.

Guidance on independence appeared in May 1976 with the issuance of Office of Management and Budget (OMB) Circular A-109, *Major System Acquisitions.* The A-109 established policy that federal agencies acquiring major systems should "provide strong checks and balances by ensuring adequate system test and evaluation" and "conduct such tests and evaluation independent, where practicable, of developer and user." The A-109 did not make a distinction between DT&E and OT&E; it made a distinction between tester, user and developer.

To its credit, the DoD had embarked on this course several years earlier. The July 1970 Report of the Blue Ribbon Defense Panel (BRDP) had some very critical findings on OT&E and highlighted the lack of OT&E oversight in OSD as a "glaring deficiency." Deputy Secretary of Defense David Packard responded by tasking the DoD's chief acquisition official, the Director of Defense Research and Engineering, to establish a Deputy Director for Test and Evaluation, who would have "across-the-board responsibilities for OSD in test and evaluation matters." More than a decade later, however, Congress found the reporting relationship between the test overseer and chief acquisition official to be unsatisfactory and created the office of the DOT&E (Public Law 98-94, September 1983), independent of officials in the acquisition decision-making chain.

There have since been two T&E camps in OSD: operational testers under the DOT&E and developmental testers under the chief acquisition official. Unfortunately, though, considering the relative proportion of DT versus OT during a program life cycle, OSD resources for these offices have shifted significantly out of balance and today are almost exactly opposite of where they need to be, and the DOT&E oversees an acquisition portfolio almost twice as large as DoD's chief acquisition official. There are 310 programs under DOT&E oversight; the Under Secretary of Defense for Acquisition, Technology, and Logistics MDAPs/major automated information systems (MAIS)/Special Interest list includes 150 programs.

In the wake of the BRDP recommendations, the DoD has focused almost singular emphasis on OT&E (more reason there is Myth No. 5), and DT&E oversight became the glaring deficiency. The Weapons Systems Acquisition Reform Act (WSARA) (PL111-23) of 2009 directed the DoD to establish the office of what is now the DASD (DT&E), and more legislation followed to bring more attention to DT&E. For example, the National Defense Authorization Act for Fiscal Year (FY) 2012 (PL112-81) requires that each MDAP be supported by "a governmental test agency, serving as lead developmental test and evaluation organization"—in other words, a "DTA." Thus, OSD has a DOT&E and a DASD(DT&E), and programs have an OTA and a DTA, not to mention the SCA.

In the 21st century, we generally know how to build the machinery that makes things go (or go "bang"); our challenges arise when we connect them to a network.

An alternative and perhaps more efficient approach might have been to revise the statute already in place (i.e., 10 U.S.C. §139) and remove the arbitrary boundary to DT&E, establishing an office whose function is to provide independent T&E oversight throughout the life cycle. Likewise, additional efficiencies can be gained, including actually achieving the elusive "early involvement," by having the OTAs engaged throughout the life cycle as a program's independent test agent (ITA versus OTA). As this is entirely consistent with the independence requirement of the A-109, it would improve synchronization of the overall T&E effort, bring needed mission context into early testing and may produce the downstream benefit of reducing the scope of testing later. The Army Test and Evaluation Command, for example, already serves as both OTA and DTA.

Myth No. 2: Effectiveness and Suitability Completely Describe Today's Systems

Having worked in information technology T&E for most my testing career, I have a particular bias for the terms "effective and suitable" used to evaluate systems and inform system acquisition decisions, and it goes something like this: In the 21st century, we generally know how to build the machinery that makes things go (or go "bang"); our challenges arise when we connect them to a network. Interoperability and cybersecurity are today's chief concerns. I see effectiveness and suitability as industrial-age bins into which we try to stuff information-age issues. I have read countless evaluation plans and test

reports, none of which has a compelling structure where interoperability and cybersecurity fit into the evaluation of effectiveness and suitability; some of them, in fact, do not even address these issues and rely instead on certification agents (i.e., the Joint Interoperability Test Command and SCA) to assess them. More disconcerting, however, is that, because we are obliged to report in terms of effectiveness and suitability, interoperability and cybersecurity are rarely discussed during acquisition decision events.

What about that other bin: survivability? Is cybersecurity part of survivability? In short, survivability is another industrial-age bin that also has a basis in law. First written in Public Law 99-500 in October 1986 (now 10 U.S.C. §2366), realistic survivability testing places "... primary emphasis on testing vulnerability with respect to potential user casualties ..." and is required for "covered systems," which include vehicles, weapon platforms or conventional weapon systems when they have "... features designed to provide some degree of protection to users in combat."

In other words, if the system has features designed to protect the human, it has to be tested to ensure it protects the human. Survivability is about saving lives, not saving data—so cybersecurity is not a good fit in the survivability bin.

When the terms effectiveness, suitability and survivability were written into laws back in the 1980s, the DoD was acquiring information technologies through a separate acquisition process with separate security procedures (see discussion of Myth No. 4), and it is unlikely that anyone foresaw the challenges associated with today's network-enabled technologies. Interoperability and cybersecurity are the developmental challenges that concern me most today, and subordinating them within the effectiveness and suitability model marginalizes their importance and reduces their exposure to decision makers. So let's compromise for today's network-enabled systems: Let us evaluate them based on effectiveness, suitability, interoperability and cybersecurity.

Finally, my No. 1 myth in T&E is:

Myth No. 1: The Purpose of DT&E Is To Get Ready for OT&E

This is what happens when developers, testers and decision makers believe Myth No. 5. Except it's not a myth; it's doctrine written in the DAU *Glossary* (quoted above): "... to certify readiness for initial operational testing." Just like the terms "effectiveness" and "suitability," this is an outdated idea that stuck, and most of our acquisition leaders, program managers and testers describe DT&E in these terms today. At one point, the DASD(DT&E) office even published an "assessment of operational test readiness (AOTR)" and briefed the assessment at operational test readiness reviews. The AOTR had a lot of good information; in fact, it was a very a good predictor of the test outcome, but it was too late to help programs positively affect the outcome. We had to change the value proposition for the DASD(DT&E) office, and change the paradigm of conducting DT to determine readiness for OT. To help programs improve outcomes, we had to shift left and provide the DT&E assessment at the point when the program could act on the information provided—prior to starting production. All tests inform production decisions—build-it or fix-it decisions—and acquisition decisions. The purpose of DT&E is to help programs set the conditions for entry into production.

Figure 1 positions OT&E in accordance with statute to bring data to inform the Full-Rate Production decision. DT&E brings data to inform all the other decisions programs make but with particular emphasis on ensuring readiness to begin production at Milestone C. Ultimately though, this type of DT&E-OT&E "stovepiping" or bureaucratic separation is inherently inefficient. The more effective strategy is to combine what we now think of as DT&E, OT&E, interoperability and cybersecurity testing into an integrated test approach to maximize the ROI of every test activity throughout the life cycle. To help programs reduce discovery of deficiencies late in the life cycle, testers must develop a comprehensive evaluation framework and then formulate a logical sequence of integrated test activities to collect the data needed to answer the so-what questions before commitment to production. When properly planned and executed, integrated testing will enable improved acquisition outcomes.

Summary

We've learned some very important lessons over the last 43 years, and as a result, we do a lot of things very well in T&E. However, we should always look for ways to improve our support to programs and decision makers, and there are a few myths and misconceptions we need to dispel. Program managers understand that T&E is essential to helping move development forward; they are not looking for us to be gatekeepers. There are enough gatekeepers as it is. Rather, program managers look for the T&E community to be engaged throughout the life cycle, to treat every test activity as a shared resource and to provide feedback. However, to maximize their testing ROI, programs must weight the T&E effort early—shift left—to set the conditions for a successful acquisition outcome. We need to work with programs to help them shift left, and bring the same kind of post-LRIP OT&E rigor that we have developed over the years into an integrated T&E approach—and, for today's network-enabled technologies, include tests to help programs deliver not just effective and suitable capabilities but interoperable capabilities that operate securely in the cyber domain. We must also be draconian stewards of the nation's resources and ensure tests support decisions that drive development forward. The paradigm of doing DT&E to get ready for OT&E has had its day, and that day is past. The future of T&E is to be an integrated, life-cycle activity that informs acquisition decisions. And, while independent, we also are a partner because we share the goal of ensuring that development problems do not become the warfighter's problems.

The author can be contacted at **steven.hutchison@hq.dhs.gov**.