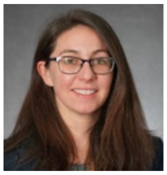
Operational Test and Evaluation (OT&E) for Rapid Software Development

HON Nickolas Guertin

Director, Operational Test and Evaluation
Washington, D.C.



Dr. Allison Goodman

Institute for Defense Analyses
Alexandria, VA

Abstract

The Department of Defense (DOD) is moving towards rapid and iterative software development and deployment. Operational testing strategies for programs utilizing iterative software development should include looking left for technical measures, looking at use in an operationally representative environment to evaluate mission accomplishment, and looking right for long-term suitability. Automated testing should inform operational testing and evaluation and is essential to software development. It alone, however, is not sufficient to determine a system's operational effectiveness, suitability, and survivability because those determinations incorporate the operationally representative use of the system by representative users. The operational test and evaluation community should build these concepts into planning now to position the DOD for the future.

Keywords: Iterative Software Development; Shift Left; Look Right; Human-Intensive Testing; Automated Testing

Introduction

Nearly every Department of Defense (DOD) acquisition program and weapon system today relies on software, and software itself is growing more complex to meet the capabilities needed for the most

The ITEA Journal of Test and Evaluation, (ISSN 1054-0929), is published four times each year by the International Test and Evaluation Association, 11350 Random Hills Rd, Suite 800, Fairfax, VA 22030 (www.itea.org)

challenging missions. The science of adequately, efficiently, and effectively testing software also is evolving, incorporating new tools and techniques to accommodate different software development methodologies, such as Agile and DevSecOps. Many test techniques inherent to these development methods focus on testing the software's technical and functional components. These techniques are essential parts of the test and development process, but on their own they are not sufficient to determine a system's operational effectiveness, suitability, and survivability.

The operational test and evaluation (OT&E) community can and should use testing data from across the development lifecycle, traditionally considered to include contractor and developmental test data, to inform evaluations of operational effectiveness and suitability: shift left for technical measures, test under operational conditions, and look right for long-term suitability. This paper describes how these areas integrate to support iterative operational evaluations that provide timely information to decision-makers. This paper also proposes next steps the OT&E community can take to implement this paradigm.

Operational Testing Strategies for Iterative Software Development

Operational testing of systems encompasses a broad view, centered on the operationally realistic use of the system in an end-to-end context to understand how that system interacts with other capabilities within a larger warfare scenario and mission. Software functionality and system integration should be tested every step of the way. Operational testing focuses on the interoperations of the software system, in the target environment, with the intended users. This focus builds the bigger picture for evaluating the success of both individual tasks and collective missions. Operational testing also assesses the software implementation process, which includes setting and refining tactics, techniques, and procedures (TTPs) for a new capability and training users on these TTPs. System roll-out processes are a critical part of this implementation and they can cause problems if not properly orchestrated.[1]

DOD mission systems are growing in their complexity of functionality and their depth of interaction with the environment, and the ability to test every attribute in an operational context has outpaced manual test methods. Therefore, operational evaluations require a balance of both automated testing and human-intensive testing. Automated testing can evaluate system functionality, assess inter- and intrasystem interactions, and identify defects in software functionality. Human-intensive testing includes, but is not limited to, assessing the software implementation process, training, usability, and workload.

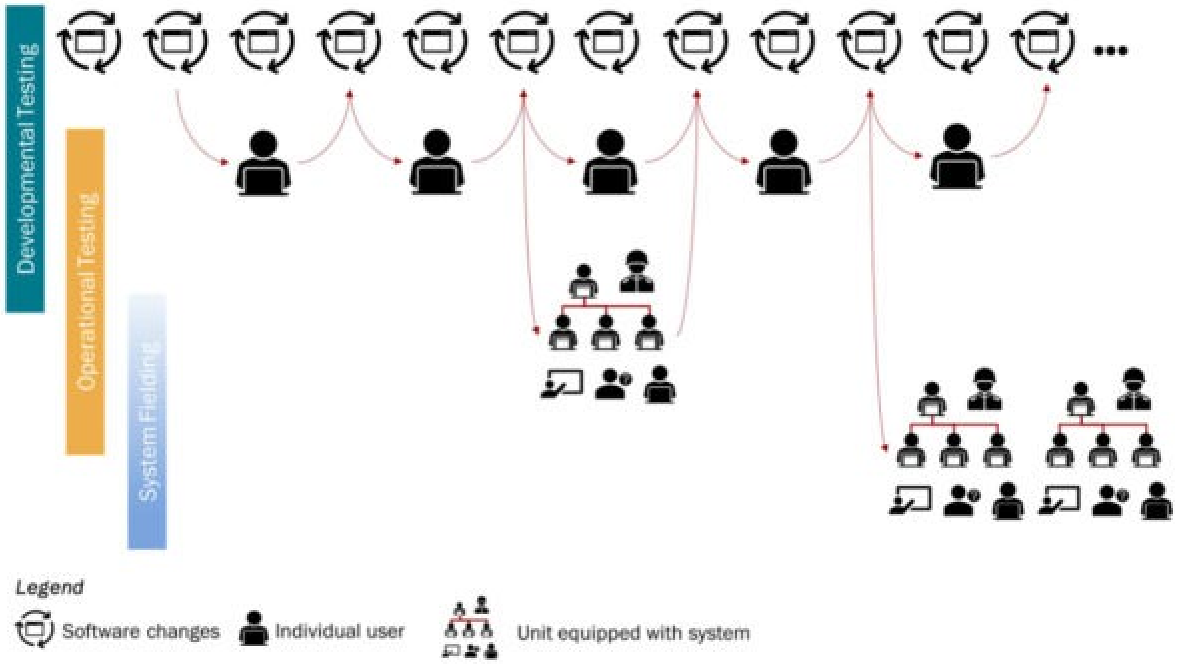
Relying on rapid, iterative software development and deployment to add new capability – which is fast becoming the DOD standard – means fielded systems will never be "complete."[2] Even when software perfectly fits the need, it must be maintained, for example, by patching vulnerabilities. The OT&E community needs to reexamine its tools and processes for operationally testing these kinds of changes to support deployment at scale.

The challenge for the OT&E community is to find the right balance between evaluating system updates and the need to quickly get those updates to the field. An operational testing strategy needs to be iterative and flexible to inform these updates as needed. Figure 1 shows a notional development cycle, presenting touchpoints with individual users and how these may align with the government's developmental and operational testing components. These touchpoints provide opportunities for the OT&E community to shift left and gather operational testing data earlier in the development cycle. After a capability is completed, it may be fielded to a subset of users, who receive representative training and system support in the context of a larger warfare scenario and mission. This limited fielding allows the

The ITEA Journal of Test and Evaluation, (ISSN 1054-0929), is published four times each year by the International Test and Evaluation Association, 11350 Random Hills Rd, Suite 800, Fairfax, VA 22030 (www.itea.org)

OT&E community to gather data in the operational context and perform incremental evaluations to support broader fielding decisions.

Figure 1. Operational testing should balance evaluations of the system with the rate of updating and the rate of deployment.



Shift Left for Early Operational Data

As the functional performance of software grows, the ability of humans to adequately execute all necessary testing shrinks immensely. Automated testing is the principal way the test and evaluation (T&E) community can achieve the depth and breadth of testing needed to identify and address defects. Early testing in an operational context can be used to explore mission threads. Together, this testing informs how a particular system interoperates across the joint warfighting domain, including with allies and partners.[3] For example, the System Architectural Virtual Integration (SAVI) study noted how early system testing reaped generous benefits by identifying and resolving architectural problems early in the development-test-deploy-operate cycle.[4] To be used throughout a capability's lifecycle – concept, coding, updates, and patching when necessary – automated testing needs to be designed into the software architecture and the testing infrastructure at the very beginning of software development.

Operational evaluations should incorporate test data from development to evaluate the software's functionality. However, the testing infrastructure has to be operationally representative, with representative system interfaces and data flows, in order to achieve this. Operational testers need to be engaged when the testing architecture is being developed to determine how operationally representative it is.[5]

User involvement has become central to how software-reliant systems are developed and nurtured throughout the lifecycle. The rubric for implementing DevSecOps is replete with either continuous or periodic user involvement. Operational testers should also be part of these user touchpoints so they can

The ITEA Journal of Test and Evaluation, (ISSN 1054-0929), is published four times each year by the International Test and Evaluation Association, 11350 Random Hills Rd, Suite 800, Fairfax, VA 22030 (www.itea.org)

better understand user needs and concerns and gather data to support subsequent operational evaluations. In particular, these touchpoints may provide early information about human-system interaction (HSI) that will support evaluations of the capability at scale.

Test under Operational Conditions – the Human Element

Operational effectiveness evaluations must consider individual tasks in the system as well as the end-to-end mission those tasks collectively support. Large-scale software systems often have many different types of end users, all accessing the system to perform an assigned element of a larger end-to-end mission. Not only does each of type of end user perform different tasks in the system, the success of the overall mission also depends on the success of key individual tasks. Operational testing in a representative environment exercises these end-to-end processes to determine whether all users can accomplish their individual tasks, and how this individual accomplishment translates to overall mission accomplishment. For example, the initial operational test and evaluation of the Military Health System (MHS) GENESIS found that poorly defined workflows and user roles increased the time to complete daily tasks because the processes that each hospital and clinic had conflicted with the enterprise processes inherent to the system.[6]

Testing software suitability often requires a combination of human-intensive and automated testing since evaluations of overall capability need to address all aspects of software functionality. HSI evaluations capture the human experience in operating the system, including system usability and user workload. User perceptions are often context driven, and therefore HSI is best evaluated under representative conditions. For example, in a study of a new threat detection technology for helicopters, operational testing demonstrated that the new technology aided mission success. However, under certain conditions, the technology created a higher workload for operators than when the technology was absent.[7]

Representative use of the system enables the evaluation of training and TTPs from all user perspectives. It further enables discovery of any faults in the software implementation process, from change management to training to daily use. For example, a limited user test of the Distributed Common Ground System – Army Capability Drop 1 showed that users rated the system as user friendly, but the unit lacked mature TTPs to integrate system capabilities into its mission. Further, insufficient collective training time prevented units from developing standard operating procedures, which may have mitigated the TTP immaturity.[8] Additionally, during operational testing of the Defense Enterprise Accounting and Management System, users stated that the training did not prepare them for site-specific workflows.[9]

Look Right for Long-Term Suitability

The OT&E community should also look right when testing systems built to change, particularly when evaluating the long-term suitability of those systems. During development stages, testers must expand their focus to consider the ease and feasibility of repeatedly – and perhaps frequently – evolving a system, both technologically and in terms of user support. Updating training, documentation, and other aspects of user support is critical to establishing the system's suitability for use in combat by the typical operator.

It is neither feasible nor desirable to perform independent operational testing on every software update. As new capabilities are implemented, the OT&E community should assess, on an ongoing basis, the

The ITEA Journal of Test and Evaluation, (ISSN 1054-0929), is published four times each year by the International Test and Evaluation Association, 11350 Random Hills Rd, Suite 800, Fairfax, VA 22030 (www.itea.org)

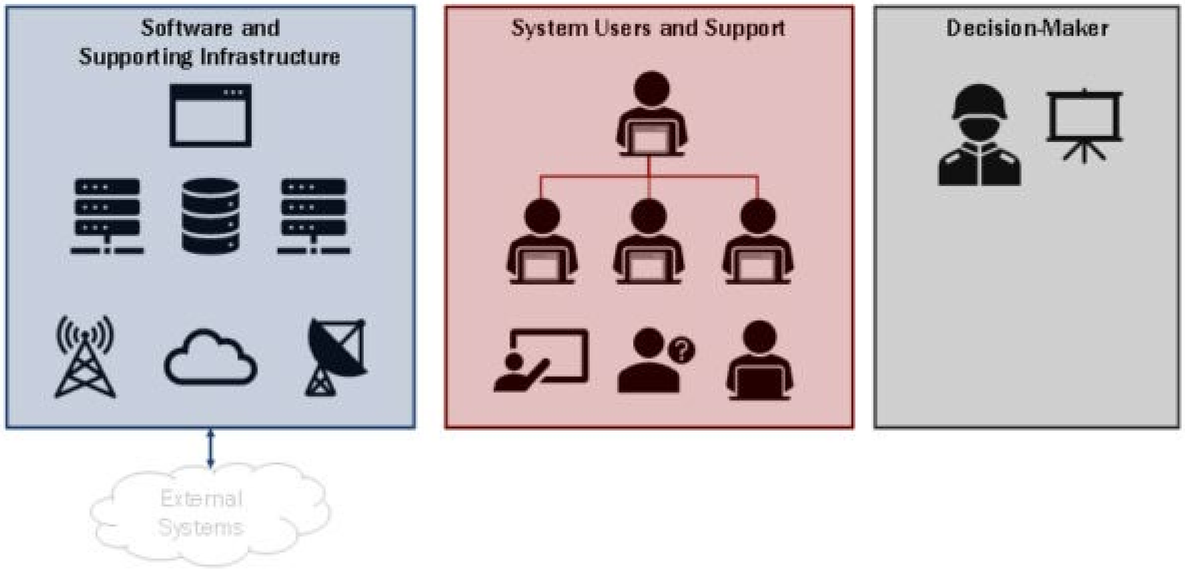
degree of risk the capabilities pose to mission accomplishment and then appropriately scope the operational testing and evaluation of those capabilities. Such risk assessments should be informed by data from continual, automated testing throughout system development and data from operational testing of previous system versions. Additionally, automated monitoring and collection of key metrics, such as operational availability and system support, can provide valuable information for risk assessments and operational evaluations of systems that already have been fielded.

What Should We Do Now?

As discussed throughout this article, automated testing does not obviate operational testing; the different types of tests are used to evaluate different aspects of a system. Automated testing and developmental testing should inform operational testing; operational testing should evaluate how the system supports mission accomplishment in the end-to-end mission context.

One of the five pillars of the June 2022 DOT&E Strategy Update is to "test the way we fight." Operational evaluations consider more than just software and its supporting infrastructure. Operational evaluations consider the holistic system – including the system's users, the technical support available to those users, and the decision-making the system enables (Figure 2). To achieve this holistic evaluation, independent operational testing must include operationally representative users completing end-to-end missions. Doing so allows testers to evaluate both the software's mission functionality and the support needed to operate the software. The OT&E community should look for opportunities to collect data for OT&E during culminating exercises from training periods or early system fielding, as feasible.[10]

Figure 2. OT&E includes the holistic system, not just the software and supporting infrastructure.



As operational testing dives deeper into the integration of software and users, the OT&E community must improve its methods for evaluating operational suitability. Further, operational testing strategies should also evaluate how users learn about and are trained on changes to the software and the associated processes. Operational testing should continue to incorporate validated survey instruments as well as other qualitative and quantitative measures[9] to assess HSI. The OT&E community should also

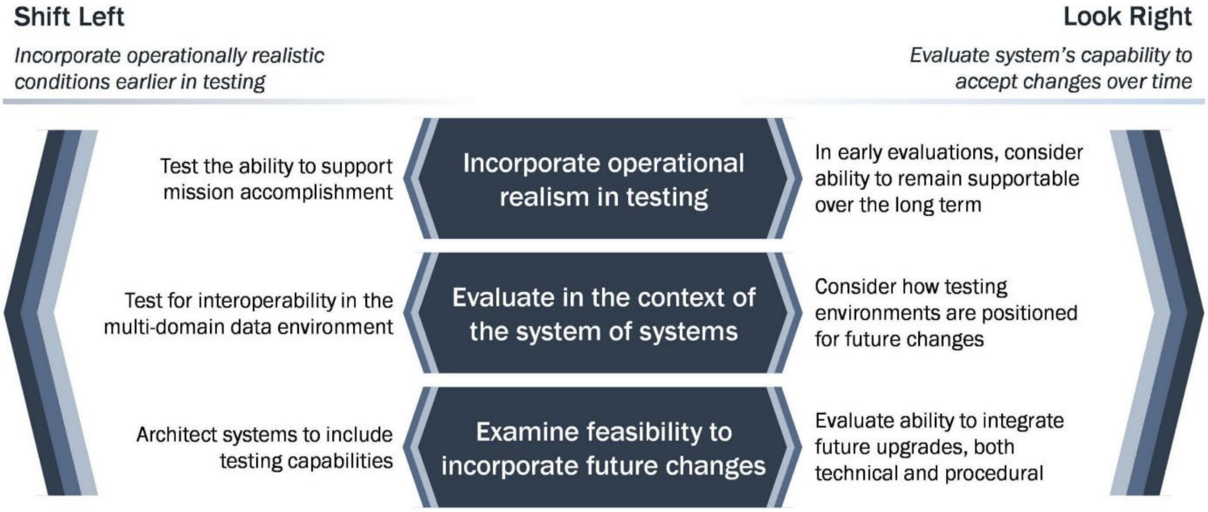
The ITEA Journal of Test and Evaluation, (ISSN 1054-0929), is published four times each year by the International Test and Evaluation Association, 11350 Random Hills Rd, Suite 800, Fairfax, VA 22030 (www.itea.org)

build robust training evaluations that include validated survey instruments, such as the Operational Assessment of Training Scale.[11] In addition, testers should evaluate the implementation of TTPs that incorporate the new software to determine whether they are sufficient to support continued operations, both in the near term and as the software continues to be updated.

How Should We Position for the Future?

The T&E community should start building the shift-left and look-right concepts into planning now to begin positioning the DOD for the future (Figure 3). The sooner we begin to look right and consider future possibilities, the better we can characterize a system's potential.

Figure 3. T&E should shift left to incorporate operationally realistic conditions earlier in testing and look right to evaluate how systems are positioned for the future.


Shift Left
Incorporate operationally realistic conditions earlier in testing
Look Right
Evaluate system's capability to accept changes over time
Test the ability to support mission accomplishment
Incorporate operational realism in testing
In early evaluations, consider ability to remain supportable over the long term
Test for interoperability in the multi-domain data environment
Evaluate in the context of the system of systems
Consider how testing environments are positioned for future changes
Architect systems to include testing capabilities
Examine feasibility to incorporate future changes
Evaluate ability to integrate future upgrades, both technical and procedural

The T&E community, including operational and developmental testers, should work with designers and engineers to incorporate testing capabilities, including automated tools and processes, into a software's baseline architecture to achieve the shift-left and look-right concepts for testing. Likewise, designers and engineers should consult the OT&E community about automated testing requirements to understand how these tools could support operational evaluations and what gaps will remain for operational testing.

Software-intensive systems do not operate in isolation; operational evaluations must assess these systems in the context of the data environment in which they will function. To do so, the DOD needs to develop and maintain a data environment and a platform for testing interoperability under operationally realistic conditions. This environment and platform will allow the DOD to evaluate how changes in one system may affect the complex system of systems.

The ITEA Journal of Test and Evaluation, (ISSN 1054-0929), is published four times each year by the International Test and Evaluation Association, 11350 Random Hills Rd, Suite 800, Fairfax, VA 22030 (www.itea.org)

As we shift testing left, we need to consider how to use early testing to examine challenges with both functionality and suitability. To look right, we need to go one step further and determine how our systems are positioned for the future and how to build the testing capabilities for these future needs.

References

[1] Simon Cleveland and Timothy J. Ellis, "Orchestrating End-User Perspectives in the Software Release Process: An Integrated Release Management Framework," *Advances in Human-Computer Interaction*, 2014 (November 16, 2014).

[2] Nikolas H. Guertin, Douglas C. Schmidt, and William Scherlis, "Capability Composition and Data Interoperability to Achieve More Effective Results than DoD System-of-Systems Strategies," Acquisition Research Symposium, April 2018.

[3] Department of Defense, "Fact Sheet: 2022 National Defense Strategy," 2022, https://media.defense.gov/2022/Mar/28/2002964702/-1/-1/1/NDS-FACT-SHEET.PDF.

[4] Peter H. Feiler, Jörgen Hansson, Dionisio de Niz, and Lutz Wrage, *System Architecture Virtual Integration: An Industry Case Study* (Pittsburgh, PA: Carnegie Mellon University, November 2009).

[5] Nikolas H. Guertin and Gordon Hunt, "Transformation of Test and Evaluation: The Natural Consequences of Model-Based Engineering and Modular Open Systems Architecture," Acquisition Research Symposium, April 2017.

[6] Director, Operational Test and Evaluation, "DOD Healthcare Management System Modernization," in *FY 2018 Annual Report* (Washington, DC: DOT&E, December 2018), 19-22.

[7] Heather Wojton, Chad Bieber, and Daniel Porter, *A Multi-Method Approach to Evaluating Human-System Interactions during Operational Testing* (Alexandria, VA: Institute for Defense Analyses, 2017).

[8] Director, Operational Test and Evaluation, "Distributed Common Ground System – Army (DCGS-A)," in *FY 2019 Annual Report* (Washington, DC: DOT&E, December 2019), 77–78.

[9] Director, Operational Test and Evaluation, "Defense Enterprise Accounting and Management System (DEAMS)," in *FY 2018 Annual Report* (Washington, DC: DOT&E, December 2018), 179–180.

[10] Director, Operational Test and Evaluation and Undersecretary of Defense for Research and Engineering, "Software Acquisition," chap. 5 in *T&E Enterprise Guidebook* (DOT&E and USD(R&E), Washington, DC: 2022).

[11] Director, Operational Test and Evaluation, *Validated Scales Repository* (Washington, DC: DOT&E, 2020).

The ITEA Journal of Test and Evaluation, (ISSN 1054-0929), is published four times each year by the International Test and Evaluation Association, 11350 Random Hills Rd, Suite 800, Fairfax, VA 22030 (www.itea.org)

Author Biographies

Nickolas H. Guertin was sworn in as Director, Operational Test and Evaluation on December 20, 2021. A Presidential appointee confirmed by the United States Senate, he serves as the senior advisor to the Secretary of Defense on operational and live fire test and evaluation of Department of Defense weapon systems. Mr. Guertin has an extensive four-decade combined military and civilian career in submarine operations, ship construction and maintenance, development and testing of weapons, sensors, combat management products including the improvement of systems engineering, and defense acquisition. Most recently, he has performed applied research for government and academia in software-reliant and cyber-physical systems at Carnegie Mellon University's Software Engineering Institute. Over his career, he has been in leadership of organizational transformation, improving competition, application of modular open system approaches, as well as prototyping and experimentation. He has also researched and published extensively on software-reliant system design, testing and acquisition. He received a BS in Mechanical Engineering from the University of Washington and an MBA from Bryant University. He is a retired Navy Reserve Engineering Duty Officer, was Defense Acquisition Workforce Improvement Act (DAWIA) certified in Program Management and Engineering, and is also a registered Professional Engineer (Mechanical).

Allison Goodman, Ph.D., is a Research Staff Member at the Institute for Defense Analyses. She provides expertise in the operational test and evaluation of large-scale software systems within the DOD. Dr. Goodman received her Ph.D. in biomedical engineering from Virginia Tech and received a B.S. in mechanical engineering from Rowan University.

The ITEA Journal of Test and Evaluation, (ISSN 1054-0929), is published four times each year by the International Test and Evaluation Association, 11350 Random Hills Rd, Suite 800, Fairfax, VA 22030 (www.itea.org)