CONTENTS

OPERATIONAL TEST AGENCIES: ENHANCING OPERATIONAL REALISM

TECHNICAL ARTICLES

DEPARTMENTS

ON THE COVER: Insight into system performance must be available prior to system design reviews and acquisition milestones; otherwise, deficiencies are uncovered late in development when design changes are most costly. Identifying failure modes, capabilities, and limitations are steps toward early identification of problems. A further requirement is providing more realistic environments, operation, and operators earlier in the acquisition cycle. This issue addresses enhancing operational realism from several perspectives: the Operational Test Agency commanders, Office of the Secretary of Defense reinvigorating developmental test and evaluation; realistic threat representations; the role of experimentation, distributed virtual environments, and more. The goal is to make test and evaluation more timely, effective, and efficient.

The cover depicts T & E Challenges in the 21st Century and is provided courtesy of the William J.Hughes Technical Center of the Federal Aviation Administration.

■ ITEA Headquarters: 4400 Fair Lakes Court, Suite 104, Fairfax, Virginia 22033-3899; Tel: (703) 631-6220; Fax: (703) 631-6221, E-mail: *itea@itea.org;* Web site: *http://www.itea.org.*
■ ITEA is a not-for-profit international association founded in 1980 to further the development and exchange of technical information in the field of test and evaluation.
■ *The ITEA Journal* (ISSN 1054-0229) is published quarterly by the International Test and Evaluation Association at 4400 Lakes Court, Suite 104, Fairfax, Virginia 22033-3899. Single issue cover price for *The ITEA Journal* is $20. ITEA membership dues are $45 for individuals, $25 for full-time students, and $745 for corporations. Annual dues include a one-year subscription to *The ITEA Journal*. The annual subscription rate for libraries and other organizations providing timely reference material to groups is $60. All overseas mail (air mail or AOA) requires an additional $20. *The ITEA Journal* serves its readers as a forum for the presentation and discussion of issues related to test and evaluation. All articles reflect the individual views of the authors and not official points of view adopted by ITEA or the organizations with which the authors are affiliated.

POSTMASTER: Send address changes to: ITEA, 4400 Fair Lakes Court, Suite 104, Fairfax, Virginia 22033-3899.

President's Corner

ITEA Journal 2008; 29: 115

Over the years, ITEA has become recognized as a preeminent Test and Evaluation organization. The ITEA brand represents education, a forum for vetting critical issues that impact the future of T&E, and professional networking. To that end, representatives from the Indian Ministry of Defense Acquisition community have approached ITEA to provide T&E training courses to their T&E community. The education committee is working with our counterpart T&E professionals in India to tailor some of our offerings to meet their specific needs. This, too, expands the ITEA presence internationally and is only one example of how ITEA is working anew to emphasize the "I" in ITEA. I challenge each individual member to share your T&E expertise and lessons learned with the international community by submitting abstracts and papers to present at international workshops and symposiums. Also, the education committee is exploring the possibility of accrediting some of our short course offerings so that attendees can obtain continuing education credits when enrolled in an ITEA T&E course.

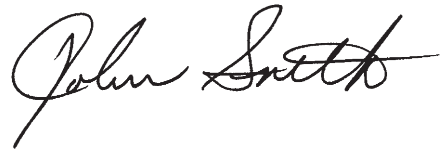
The Test and Evaluation community has over come many challenges. One of the most prevalent challenges today is how to create an operational representative environment for testing the enormously complex weapon systems that are being developed in support of the warfighter.

The operational test community has always strived to ensure that the environment used for its testing was operationally realistic. Traditionally operational tests include using live test assets and warfighters to operate the equipment during the test to create operational realism. Unfortunately, the cost of constructing an event that is primarily live is very costly and the schedule is affected by the availability of operators. The downside is that the operational environment created is service centric rather than joint. In contrast, contractor and developmental testing focus more on determining whether the system under test will operate as specified. This testing is conducted in a more controlled and repeatable environment. In addition the urban environment imposes more constraints on T&E as well. The added complexity of systems and operating environment makes it necessary to focus on testing systems in an operationally representative environment earlier in the test process. To address this problem Director Operational Test & Evaluation and Acquisition Technology & Logistics partnered to emphasize the need for early involvement and integration of contractor, developmental and operational testing. When the operational environment created includes a realistic threat representation, multiple services and coalition partners required to execute a task, we have the representation for a joint mission environment.

John Smith

Recently, the Director Operational Test & Evaluation chartered the Testing in a Joint Environment Roadmap Senior Steering Group(TSSG). The TSSG is the decision-making body for testing in a joint environment methods and infrastructure initiatives in lieu of an advisory group. The TSSG consists of a cross section of users and stakeholders from the acquisition community. This group will ensure that systems developed for the warfighter are tested and evaluated in an operationally realistic joint mission environment. I am sure that by now everyone has some familiarity with two of the testing in a joint environment roadmap initiatives; Joint Test Environment Methodology and Joint Mission Environment Test Capability. These initiatives provide the methods and processes as well as the infrastructure framework to execute a test in the joint mission environment. Integral to the implementation of testing in a joint environment are: a) what is the proper mix of Live Virtual and Constructive participants; and b) who validates that the design is an actual representation of the joint environment for a specific capability. The design and execution of the Joint Mission Environment is not as straight forward as one would believe. We all have heard the expression, "the devil is in the details." Well it's no different here. Each instantiation of the joint mission environment is different. While the network connectivity might be "persistent," each instantiation includes different players reacting differently thus affecting the logical and physical designs.

I am sure that as you read this edition of the journal you will be enlightened by the articles and papers from the testers who have given genuine thought to the challenges of testing in an operationally realistic joint mission environment.



ITEA Journal 2008; 29: 117–119

FAA Test Challenges in the 21st Century

John Wiley
Federal Aviation Administration,
Atlantic City International Airport, New Jersey

Asymmetric Warfare, in very simple language, is war between an extremely strong opponent and a weaker, less capable one. On September 11, 2001, asymmetric warfare touched America when our commercial aircraft were used as weapons of mass destruction against innocent individuals. This one day in history changed the way the Federal Aviation Administration (FAA) looks at its future role in maintaining safeguards in the air traffic control system. The loss of life was no longer by human error, equipment failure, or an act of God; the rules of engagement had changed. How do we strengthen the ability to identify the risk to operational usability and safety for our users?

John Wiley

The mission of the FAA is to provide the safest, most efficient aerospace system in the world. Our vision is to improve the safety and efficiency of aviation, while being responsive to our customers and accountable to the public. The impact in our economy by the civil aviation contribution to the U.S. Gross Product for Jobs is appropriately 10.1 million jobs a year. The FAA is a highly visible and media-driven organization. The FAA is held to a very high standard when it comes to loss of life, and only zero percent is acceptable to the public.

Today's air traffic control system is operational 24/7 and is a ground-based, human-centric, relatively unautomated system, which still uses single-channel voice control in an aging infrastructure (youngest en route facility is 43 years old). The volume of air travel/transportation is expected to grow by 70 percent in the next 10 years. Our current National Airspace System (NAS) and supporting infrastructure were simply not designed to handle the voluminous increases anticipated. Advances in technologies (e.g., unmanned aerial vehicles) will mandate changes to the NAS. The current NAS must be redesigned from the ground up.

The Joint Program Development Office (JPDO) via Public Law and Executive Order, consisting of the Department Of Transportation, FAA, National Aeronautics and Space Administration (NASA), Department of Health Services, Department Of Defense, and Department Of Commerce was established to collectively support development of the Next Generation Air Transportation System (NextGen) with one mission in mind, redesign of the NAS. The NextGen key capabilities must meet the existing system's capabilities plus handle the anticipated voluminous increase in air travel in the next 10 years and beyond. The NextGen will move people and goods expeditiously from curb-to-curb while ensuring protection from foreign and domestic threats.

NextGen challenges are—while operating the current system 24/7—to design, develop, and implement multiple systems simultaneously; to redefine the roles of controllers, pilots, and dispatchers etc.; to move to an aircraft-centric design; to address the aging facilities issue; and to bring this new concept together on time and within budget, **safely**.

The William J. Hughes Technical Center, with five decades of air traffic management test and evaluation achievements, is the nation's leading aviation and air traffic management Federal Laboratory. The Technical Center is a fully integrated network of specialized labs. These are the only laboratories in the world where it is possible to simulate the end-to-end NAS in a nonoperational environment. The Technical Center also has the recognized aviation expertise in automation, communications, navigation, and surveillance needed to support this monumental task. This is the facility where the NextGen capabilities will be integrated and tested. One of the major challenges is the evolution (commonly called "transformation") from the NAS of today to NextGen of the future.

NextGen key capabilities will be supplied through performance-based services and will include: weather assimilated into the decision making at all levels; layered, adaptive security; and position, navigation, and

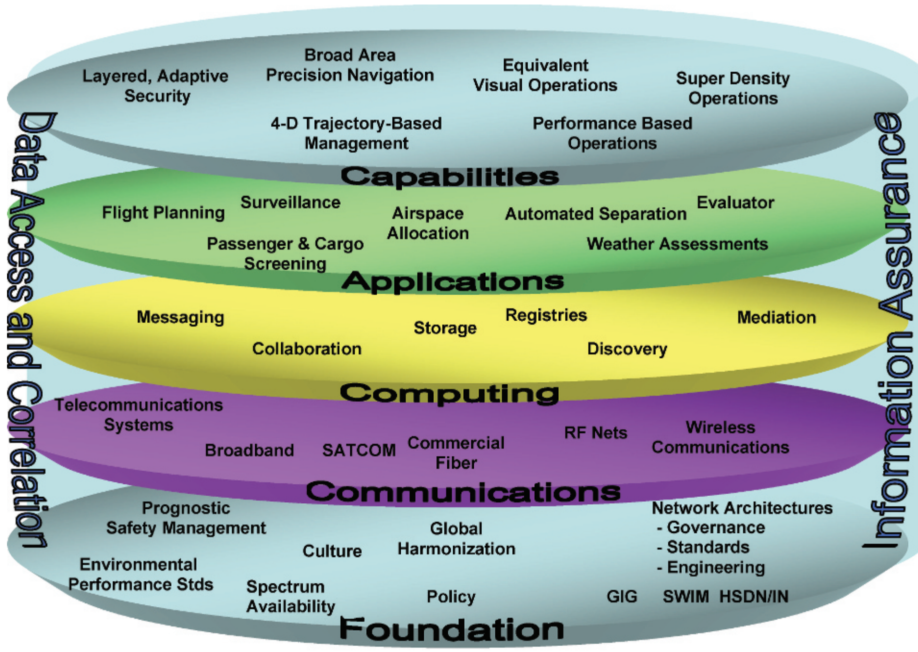


Figure 1. Data Access and Correlation, and Information Assurance

time services with trajectory-based aircraft operations. Equivalent visual operations with improved information availability will allow aircraft to operate without regard to visibility. The access to high precision position, navigation, and timing services will enhance the efficiency of airport arrivals and departures, plus enable more predictable and efficient operations regardless of meteorologic conditions. Key factors for moving to super density operations are the use of the required navigation performance (RNP) operations, the mitigation of wake vortex constraints, the improvement of runway incursion prevention algorithms, the distribution of taxi instructions before landing, and the use of aircraft as sensors in the system. NextGen will require the use of network-enabled information technologies including Network-Enabled Operations, Network-Enabled Infrastructure, and NextGen Network-Enabled Weather.

As displayed in *Figure 1*, the NextGen has five layers (Foundation, Communications, Computing, Applications, and Capabilities), and within each layer there are multiple functions. These functions are a major part of the NextGen success, which will assist the FAA in making flight safe and secure for our future travelers.

The FAA strategy is to focus on system life cycle, therefore strengthening test and evaluation processes with a focus on problem prevention, detection, and resolution as early in the life cycle as possible. To refine the new integrated procedural test practices and explore new theory of test for highly flexible and complex system of systems, the FAA has joined with academia, industry, and other government agencies. The FAA has started to conduct resource gap analysis on people and infrastructure to meet these new challenges and has started to acquire the needed proficiencies in new technologies. The questions that Testing and Evaluation (T&E) must answer are as follows: (1) Does the system meet the specified requirements? (2) Does the system fulfill its intended mission in its operational environment? (3) Is the system safe and secure? If the answers to these questions are that the risks are acceptable and therefore the system meets the desired mission results, we in the T&E organization are successful. ❑

To learn more about the FAA's T&E Program, as well as other advancements in T&E, attend the 2008 Annual ITEA International Symposium, November 10–13 Atlantic City Convention Center in New Jersey.

JOHN WILEY is the manager of the Technical Strategies and Integration organization at the Federal Aviation Administration's William J. Hughes Technical Center. He is responsible for the governance, integration, transition, and strategic management of the Technical Center's products, services, and initiatives. The Technical Center is one of the world's leading engineering, research, development, and testing facilities for nearly every aspect of air transportation systems.

Mr. Wiley has over 30 years of technical expertise in the aviation industry, 21 of those years in management. He served on the FAA Air Traffic Organization's transition team in Washington, D.C., where he provided strategy and

direction in defining, communicating, and implementing a broad spectrum of new agency management initiatives.

Mr. Wiley is a member of the International Test and Evaluation Association (ITEA) Board of Directors and is also a member of the Rowan University Electrical and Computer Engineering Industrial Advisory Board. He is an active member of the Air Traffic Control Association, former president of the ITEA South Jersey chapter, and former member of the Human Factors Society of South Jersey. He holds a bachelor's degree in electrical engineering from Drexel University and has completed graduate courses in artificial intelligence, management, strategic planning, and project management.

ITEA Journal 2008; 29: 121–124

Enhancing Operational Realism in Test & Evaluation

Ernest Seglie, Ph.D.
Office of the Director,
Operational Test & Evaluation, Washington, D. C.

The Test & Evaluation (T&E) community likes to talk about realistic operational testing. Today, weapons systems are being rushed overseas to face the real test in real — not realistic — operations. Our nation is at war and T&E must do all it can in support. Rapid acquisition is a worthy response, and finding failure modes early is an obvious way to avoid delay when problems are discovered, avoid costly rework for the fix, or preclude fielding systems with problems. There are examples:

- The Stryker Mobile Gun System's coaxially mounted 7.62-mm machine gun. In predeployment testing, the system appeared to function satisfactorily. Yet, when handed off to troops and fired in new equipment training, the machine gun's bore-sight failed.
- Two air weapons systems: Small Diameter Bomb, and Joint Air-to-Surface Standoff Missile. Both deployed systems were grounded. The weapons systems were under-designed, not tested for realistic operational tempo, and experienced unacceptable reliability and related operational performance failures.
- The MH-60S Armed Helicopter Weapons System. Hellfire missiles hanging from external launchers — forward of and level with the open cabin doors — created a hazard for the helicopter's GAU 21 gunner. In a post OT training incident, blast from a Hellfire broke pins securing the gun and forced the gun barrel around, causing it to pierce the fuselage. If the gunner had been in contact with the weapon, results could have been catastrophic.

Ernest Seglie

In the above examples, more realistic test environments and operations would have enabled discovery of the problems. The need to find problems earlier is recognized inside and outside the Department of Defense (DoD). DoD promulgated T&E policy changes in a December 22, 2007, memorandum (reprinted in the appendix following this article and available at https://akss.dau.mil/Documents/Policy/TE-Policy-Memo-Dec-2007.pdf). The goal of this policy is "early identification of technical, operational, and system deficiencies, so that appropriate and timely corrective actions can be developed prior to fielding the system." The leverage of finding problems early in system design is obvious, but unfortunately we often discover problems late, even after production commitments have been made. In this issue Darlene Mosser–Kerner explains the new policy and its impact on revitalizing T&E.

The need to find problems earlier is recognized outside the Department of Defense. The basis for the new T&E policy is a DoD Report which cites recommendations by the Government Accountability Office, Defense Acquisition Performance Assessment, and National Research Council. The report recommended that DoD: lessen dependence on testing late in development; consolidate Developmental Testing (DT) and Operational Testing (OT); and, require DT to have an operational perspective. In all instances, the implied goal is early identification of operational failure modes and system deficiencies.

At the same time, Dr. Charles McQueary, Director for Operational Test and Evaluation (DOT&E), established a priority: "Enhance operational realism in early tests, including DT." ITEA should be praised for selecting this theme for *The ITEA Journal*.

The DOT&E desired end-state is, "Sufficient operational insights gained prior to design reviews and acquisition decision points to influence system design and reduce surprises in operational test." As Dr. McQueary stated in his guest editorial in the September 2007, ITEA Journal: "OT&E should be a time of confirmation, not discovery." The focus of employing operational realism in early tests is on designing the system to operate effectively in the environment (with threat conditions) and with the system operators and maintainers anticipated when the system is deployed.

Operational realism may be easiest to imagine when thinking about system level tests — when the system or a prototype is available. But operational realism has a role very much earlier during design, when components, subsystems, and operational procedures are chosen.

What does it mean to include operational realism in testing early? Enhanced operational realism should be part of demonstrating technology readiness. Often, systems enter design and development at Milestone B (MS B) with insufficient technology maturity. Public law (PL 109–163, Section 801) now requires the Under Secretary of Defense for Acquisition, Technology and Logistics (USD(AT&L)) to certify technology has been demonstrated in a relevant environment. This certification could add discipline to acquisition guidance calling for demonstrating Technology Readiness Level 6 prior to MS B, i.e., demonstrating the technology in an operationally relevant environment before MS B. Testers should be involved in the technology demonstration, and influence the characterization of the "relevant environment" so that it is both operationally relevant and consistent with subsequent testing.

Operational realism in early testing should mean that components are tested in a relevant environment. For example, the vibrational environment on the center pylon of the F-15 necessitated major redesign of the original AMRAAM missile because no one characterized that acoustic environment and tested the components and system in those conditions. Typically, designers assume the systems with which they interface are characterized by the design specifications of the item in development. Often the truth in the field is different. For example, electrical power is often "dirtier" (noisier) than specified. Testing system components in a realistic environment — under realistic stress — can save substantial system development time. Dr. Cliff Duncan, the second Director for OT&E, once told me that with avionics he found the first places to look for trouble were in the behavior of the power supply and the connector. Shouldn't that be the first place we test components too?

Enhanced operational realism should be part of determining whether commercial-off-the-shelf (COTS) technology, software, or system components will actually perform as needed when in real operations. COTS functionality can be highly desirable but a key question is: does COTS performance hold up in operations? Obviously, the military environment can be more demanding. System designers should step up to the challenge of assessing COTS in the operational environment. Testers are obligated to identify COTS risks by testing it, early, in the proper environment.

Enhanced operational realism means we need to employ real operators to operate the system. We have had cases where equipment that was effective when used by other nations, was not so effective when our forces used it. Part of the reason was that another country could man the system with highly educated soldiers who understood much of the basic chemistry that impacts system operation. You may have heard stories of systems failing because the operator used them inappropriately. The first operational test of the SINCGARS radio had to be stopped. Soldiers when ordered to move the radio grabbed it by the antenna — which broke — leading to a mission failure. The radio did not have a handle. The point is, include those operators early so we will understand differences between the systems operated by system developers and warfighter operators.

Enhanced operational realism also means that all the interactions and interfaces that have to work for a mission to be successful are checked before the design is finalized. More general than connectors and power supplies, the age of net-centric operations and service oriented architectures requires sharing data and coordinating activity of separately developed services. If "n" services must work for mission success, and each service has a probability of success "x," then — to a first approximation — the mission will have a probability of success of only x^n. For example, if you have six services that have to work, and you only want 0.8 chance of success, then each service has to exceed 0.96 probability of success! That does not consider the case when the services are correlated, but they will be. Service oriented architectures will require more unit level testing, much more regression testing, and lots of end-to-end testing. It will be essential in testing a new service that simulated inputs from other services be realistic.

So, it is important for testers to enhance operational realism in T&E in every way they can imagine. Operational realism in early T&E can improve the chances of success in rapid fielding and in OT. It is a way we can better support our deployed forces. ❑

Dr. Ernest Seglie is science advisor, Office of the DOT&E, the Pentagon, Washington, D. C. He provides scientific and technical guidance on the overall approach to DoD evaluation of the operational effectiveness and suitability of major DoD weapon systems; provides technical review of test reports; and serves as chief technical advisor to the DOT&E.

OFFICE OF THE SECRETARY OF DEFENSE
WASHINGTION, DC 20301-1000

DEC 2 2 2007

MEMORANDUM FOR: SEE DISTRIBUTION

SUBJECT: Test and Evaluation Policy Revisions

The fundamental purpose of test and evaluation is to provide knowledge to assist in managing the risks involved in developing, producing, operating, and sustaining systems and capabilities.

T&E measures progress in both system and capability development. T&E provides knowledge of system capabilities and limitations to the acquisition community for use in improving the system performance, and the user community for optimizing system use in operations. T&E expertise must be brought to bear at the beginning of the system life cycle to provide earlier learning about the strengths and weaknesses of the system under development. The goal is early identification of technical, operational, and system deficiencies, so that appropriate and timely corrective actions can be developed prior to fielding the system. Consequently, to achieve this goal we have decided to immediately implement the following policies:

- Developmental and operational test activities shall be integrated and seamless throughout the system life cycle. As technology, software, and threats change, follow-on T&E should be used to assess current mission performance and inform operational users' during the development of new capability requirements.

- Evaluations shall include a comparison with current mission capabilities using existing data, so that measurable improvements can be determined. If such evaluation is considered cost prohibitive the Service Component shall propose an alternative evaluation strategy.

- T&E should assess improvements to mission capability and operational support based on user needs and should be reported in terms of operational significance to the user. Consequently, evaluations shall be conducted in the mission context expected at time of fielding, as described in the user's capability document, and consider any new validated threat environments that will alter operational effectiveness.

- To maximize the efficiency of the T&E process and more effectively integrate developmental and operational T&E, evaluations shall take into account all available and relevant data and information from contractor and government sources.

- Operational evaluators will continue to fulfill their statutory roles in providing assessments of operational effectiveness, operational suitability, and survivability to the MDA. In addition, program managers shall report the results of completed developmental testing to the milestone decision authority at milestones B and C. The report shall identify strengths and weaknesses in meeting the warfighters' documented needs based on developmental evaluations. The operational evaluators assessment will be provided to the MDA at the full rate production review.

- To realize the benefits of modeling and simulation, T&E will be conducted in a continuum of live, virtual, and constructive system and operational environments.

These policies will be incorporated in the next revision to DoDI 5000.2.

Dr. Charles E. McQueary
Director, Operational Test & Evaluation

John J. Young, Jr.
Under Secretary of Defense
for Acquisition, Technology & Logistics

ITEA Journal 2008; 29: 125–127

Realistic Threat Ground Vehicle Targets for Test and Evaluation

Jeffrey S. Sanders
Trideum Corp., Huntsville, Alabama

Modern weapon systems are continually pushing the envelope in sensor resolution and operator battlefield awareness. Proper stimulation of these sensors is critical in the Test & Evaluation (T&E) phase to ensure that the weapon systems under development perform acceptably when they go into battle with the warfighter. Since the whole purpose of a weapon system is to detect, identify, engage, and destroy a threat, realistic threat targets are critical to the world of T&E and threat ground vehicle targets are the ones where the rubber meets the road (literally) on modern battlefields. Even with the current focus on the danger of improvised explosive devices, threat ground vehicles such as main-battle tanks, infantry fighting vehicles, and armored personnel carriers are still recognized as the primary targets of interest for U.S. ground vehicle development programs.

Unfortunately, threat ground vehicle target technology for T&E has simply not made the advances that sensor systems and platforms have made over the last few decades. For many T&E scenarios of the recent past, the only way to get an accredited target for a test to support an acquisition milestone decision was to use actual threat vehicles with all of their associated costs, long acquisition lead times, and logistical burdens. There have been cases where surrogate targets have been successfully brought into play, but due to time constraints, there are also instances where surrogate target development efforts may not be completed in a timely manner to meet identified T&E requirements. This situation places test designers for the system under test in the classic quandary of having to meet high target signature fidelity requirements with limited budgetary resources.

The signature fidelity history of ground targets for T&E is extremely bimodal. Targets are usually either something fabricated on a test range or they're actual threat vehicles. The first option provides for a low-cost approach to providing targets for testing with varying (usually low) levels of signature fidelity. The second approach using real vehicles provides testers with high confidence that the threat representations are valid but at a significant, and often prohibitive, cost. However, there have been recent advances in low-cost target surrogates for T&E that have been successfully validated and accredited for use by certain programs. These targets can bridge the gap between nonrealistic, "whatever's handy on the range" targets and actual threat vehicles by providing test-specific signature elements to meet validation and accreditation requirements and avoid the need to use real threat vehicles.

The types of targets needed for a test depend primarily upon the geometry of the sensor/shooter/target engagement and cost savings can be realized by understanding target signature requirements and designing targets to satisfy specific test needs. *Figure 1* shows the ground target signature fidelity/cost continuum and demonstrates the constantly increasing cost associated with increasing signature fidelity. Starting on the left of the figure are common target silhouettes as specified by Army field manual FM-17-12-1 for use in gunner training. These targets are inexpensive, easily fabricated as-needed at gunnery ranges, but offer almost nothing in the way of threat representative visual and thermal infrared (IR) signatures. Until recently, these were the primary option for weapon platform sensor stimulation on gunnery ranges.

On the far right of the figure are actual threat vehicles which only need to be certified as actual threats to be accredited for T&E use. In between these two extremes are targets of varying complexity, some of which have been validated for T&E. The 2.5-dimensional (2.5-D) target surrogates shown were designed for use by the Expeditionary Fighting Vehicle (EFV) program during gunnery Operational Testing (OT). These targets are limited to use on target lifters on gunnery ranges but they are capable of presenting very realistic visual and thermal infrared (IR) signatures over a limited range of angles. Thanks to their inherent realism, they were able to be validated for use in T&E and represented an over $10 million cost avoidance for the EFV gunnery OT. More complex, multi-aspect target engagements require full three-dimensional (3-D) surrogate targets and there are examples of these in DoD inventory with IR and radio

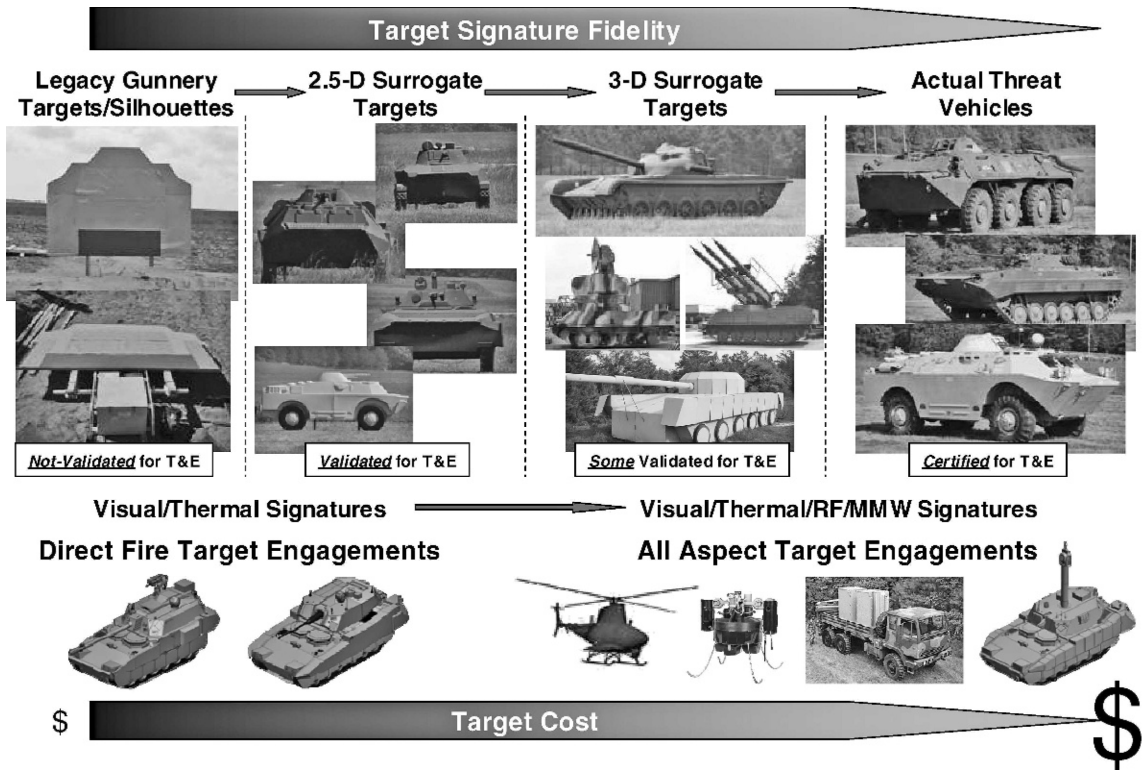


Figure 1. Ground target signature fidelity/cost continuum

frequency/millimeter wave (RF/MMW) signatures that have been validated for use in T&E activities.

The 2.5-D targets shown in *Figure 1* are known as the Threat Vehicle Surrogate Targets (TVSTs). These targets were developed by the U.S. Army's Targets Management Office (TMO) for the Marine Corp Operational Test and Evaluation Activity (MCOTEA) to meet the specific needs of the EFV gunnery OT. These targets were designed in collaboration with the National Ground Intelligence Center (NGIC) and were developed utilizing NGIC threat data and 3-D geometry and IR signature models. Getting NGIC involved in the early stages of the TVST design process greatly helped with the validation of the target's signatures, as NGIC has to concur on the final designs as threat representative. Even with the additional cost of satisfying the DA PAM 73-1 Threat Simulator Validation Process, the entire TVST design, validation, and deployment process only cost on the order of $1 million which was more than an order of magnitude less than the estimated $12 million plus cost of meeting EFV gunnery OT requirements with actual threats. The TVSTs were made from vacuum-formed ABS plastic and the target shells were fabricated by a U.S. Navy facility at Patuxent River Naval Air Warfare Center. These shells were then attached to wooden frames and from there to the mechanical gunnery range lifters. If a particular test required targets with IR signatures, a thermal signature kit could be added to the interior of the target shell to provide a validated IR signature. A total of six types of TVSTs were created and *Figure 2* shows two of them, the BTR-70 frontal target and the BRDM-2 starboard flank target.

The upper half of *Figure 2* shows an actual BTR-70 and a BTR-70 frontal TVST photographed from a distance during the TVST signature validation test. Both targets are facing the same direction and the photographs were taken within a minute of each other. As can be seen in the pair of photographs, a large piece of vacuum-formed plastic can interact with natural sources of illumination and create a very convincing threat vehicle visual signature for a limited range of viewing angles to meet gunnery OT test requirements.

The bottom half of *Figure 2* shows a thermal image pair of an actual BRDM-2 and the BRDM-2 starboard flank TVST. Once again, it is immediately apparent that a properly designed target surrogate can create a realistic threat signature. The TVST IR signature capabilities included the different vehicle operational states of cold, idling, and exercised, although only the exercised IR signature case was used in EFV gunnery OT.

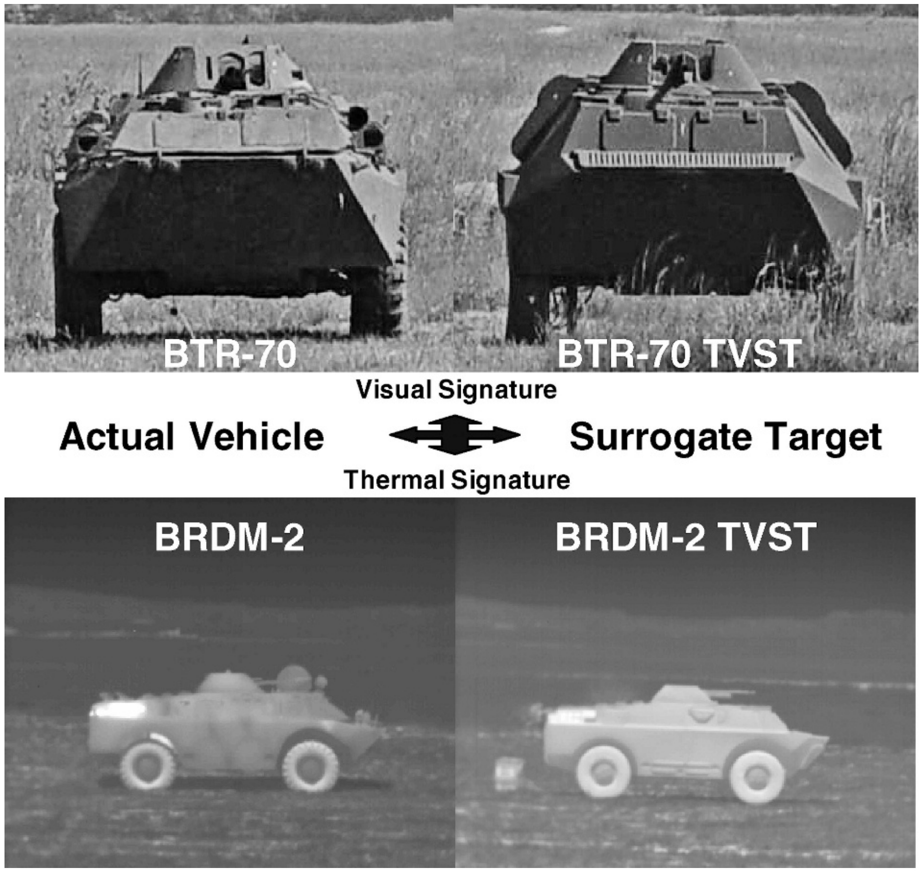


Figure 2. TVST 2.5-D visual and thermal signature comparison to actual threat vehicles

The TVSTs represent a highly successful target surrogate program that was able to help get a major developmental weapon system past an acquisition milestone. However, the TVSTs are not without their limitations and other developmental platforms will require different target technologies in order to create realistic ground vehicle target surrogates. Of particular concern is the looming transition to multispectral and hyper-spectral sensors that will have an entirely new set of requirements unaddressed by the TVST program. Nevertheless, the TVST was more than sufficient to meet the intended user's test requirements at the time and is still being used for various scenarios.

In summary, ground vehicle surrogate targets have the potential to represent a huge cost avoidance for T&E by replacing actual threat vehicles for specific tests. Real threat vehicles will never be completely replaced in the T&E world but the TVST program is an example of where they could be replaced with a much less expensive alternative to meet a specific test need. ❑

Jeffrey S. Sanders is a principal engineer at Trideum Corp. in Huntsville, Alabama. He supports the TMO in the area of target surrogate visual and IR signature design and threat target signature validation and certification. He has also managed the IR signature modeling component of NGIC's Simulated Infrared Earth Environment Lab (SIREEL) program. Dr. Sanders received a BSEE, MSEE, and a Ph.D. in engineering from Memphis State University in 1987, 1990, and 1994, respectively. E-mail: jsanders@trideum.com

Acknowledgments

The TVST development program was funded by the Central Test and Evaluation Investment Program (CTEIP) as a Resource Enhancement Project (REP). The TVST program director was Robbin Finley of the Targets Management Office, (Robbin.Finley@us.army.mil) and the initial TVST user was Ken Lardie of EFV MCOTEA, (ken.lardie@efvmcotea-usmc.org).

Featured Capability

ITEA Journal 2008; 29: 129–131

The Air Force Distributed Mission Operations Center — Over 25 Years of History Supporting the Warfighter Through Test and Evaluation

Michael J. Mora

Scientific Research Corporation (SRC), Huntsville, Alabama

The U.S. Air Force Distributed Mission Operations Center (DMOC) facility *(Figure 1)*, located on Kirtland Air Force Base, New Mexico and run by the 705th Combat Training Squadron (705 CTS) has a long history supporting Joint Test and Evaluation (JT&E). The roots of the DMOC can be traced back to the early 1980s when it was known as the Identification Friend Foe Neutral (IFFN) facility. The IFFN Joint Task Force (known today as a JT&E) was charged with the responsibility of evaluating military identification capabilities in a World War III-type scenario over Europe. In the years following its inception, a very reliable and technologically advanced large-scale, real-time modeling and simulation (M&S) capability was developed. The M&S capability consisted of several interactive virtual simulators, including Air Force Modular Control Equipment (MCE) and F-15C fighter jets; U.S. Army Hawk, Patriot, and Mixed Brigade; and the North Atlantic Treaty Organization's GEADGE Control and Reporting Center and Airborne Warning and Control System (AWACS). The heart of the simulation architecture, the Master Simulation, generated hundreds of constructive scripted, semi-intelligent, or keyboard driven players and supported environmental modeling of various jamming and weather situations. IFFN also provided simulated voice and data link communications representative of the tactical communications available to the warfighter at the time. The IFFN facility was the first of its kind, providing M&S support for Department of Defense (DoD) testing, training, research, and development and building a solid foundation for the future of M&S.

At the completion of the IFFN in 1989, Air Force leadership recognized the great potential for leveraging the M&S infrastructure. In 1990, the facility was renamed the Theater Air Command and Control Simulation Facility (TACCSF) and opened for use by other DoD programs. Some of the first to take advantage of the capabilities at the TACCSF were joint tests. The Joint Air Defense Operations/Joint Engagement Zone Joint Test used the TACCSF extensively to conduct its testing and to develop the tactics, techniques, and procedures that are being used by warfighters today. The Air Force Operational Test and Evaluation Center (AFOTEC) headquarters, also located at Kirtland AFB, utilized the TACCSF to conduct a series of survivability and employment option tests for one of today's most critical air-to-ground surveillance platforms, the Joint Surveillance Target Attack Radar System (JSTARS). The results of these tests were extremely useful in combat operations during Desert Storm. The 1990s represented a period of significant growth for the TACCSF as new virtual simulators like the Cobra Ball, the Expert Missile Tracker, and the Airborne Laser were integrated into the architecture and the capability to distribute simulation outside of the facility with other simulators was developed using Distributed Interactive Simulation Protocol Data Units. In fact, the distributed nature of M&S matured significantly during this time and the TACCSF became the hub of the Air Force distributed simulation capability and the largest real-time, tactical-level distributed simulation facility in the world. With distributed connections to several additional simulation facilities and development of an ever expanding virtual battlefield, the test community quickly realized the potential to expand their use of M&S. For example, TACCSF was used by the Office of the Secretary of Defense sponsored Joint Advanced Distributed Simulation Joint Feasibility Study to address issues, principles, procedures, and practices for the increased use of distributed simulation support to both developmental and operational tests and evaluations. TACCSF was also selected as the M&S hub for several JT&E organizations including Joint Theater Missile Defense Attack Operations, Joint Combat Search and Rescue, and Joint Cruise Missile Defense. The test community was quickly ramping up use of M&S as technology improved and budgets for live testing decreased.

In the year 2000, the TACCSF was relocated to a new $14M facility and another decade of significant growth began. The local simulators and environment

Figure 1. The Distributed Mission Operations Center facility at Kirtland Air Force Base, New Mexico

generators were upgraded with new technology and several new simulators were acquired including a JSTARS simulator, a new AWACS simulator, and two real world MCE Operations Modules (OMs) with front end simulation interfaces for sensors, data links, and voice communications *(Figure 2)*. The network connectivity to other distributed simulation facilities expanded to include over 20 sites utilizing numerous DoD networks, and DMOC engineers began developing a Multi-Level Security capability that would allow the test and training communities to bring programs at classification levels above SECRET into events and significantly increase the capability to provide a realistic representation of present battlefields and possible future battlefields. In 2003, the TACCSF was officially designated the Air Force DMOC thereby establishing the facility as a foundation of the distributed mission operations (DMO) concept for supporting the joint and coalition warfighters through integration of Live, Virtual, and Constructive capabilities. DMOC supported several test community events during this period including the Joint Fires Coordination Measures JT&E, Joint Data Link Information Combat Execution JT&E, Joint Kill Chain Experiment, Missile Defense System Exerciser test events, Defense of the Homeland Against Asymmetric Missile Attacks test events, and Joint Air/Ground Unified Adaptive Replanning test events, among others. Clearly, the test community was invested in M&S.

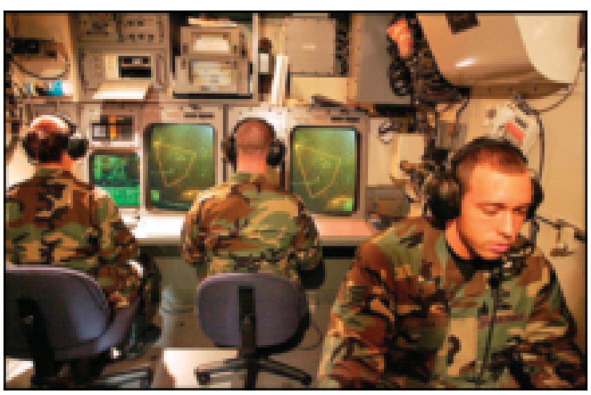


Figure 2. Real military equipment, like these Modular Control Equipment Operations Modules, was integrated into Distributed Mission Operations Center to provide the most realistic environment possible for warfighters

Figure 3. F-15E cockpit simulators (top) and an extensive data link cell (bottom) operating in a large distributed exercise allow warfighters to test and train in environments not feasible for live ranges

Today, the DMOC and the 705th Combat Training Squadron continue to support the test community with over 30 resident simulators including virtual simulators like the F-15C, F-15E, F-16, AWACS, JSTARS, MCE OMs, Cobra Ball, Predator Unmanned Aerial Vehicle (UAV), JTACS, Su-27, and MiG-29 each providing its military operator with realistic representations of the sensors, weapons, and communications (both voice and datalink) capabilities used in actual battle situations *(Figure 3)*. The environment is modeled using several simulations that can immerse the warfighter in a scenario with tens of thousands of constructive entities that represent the blue and red entities on the ground (tanks, trucks, surface-to-air missiles, radars, personnel, etc.) or in the air (jets, helicopters, UAVs, weapons, etc.) and includes the capability to represent the electronic warfare environment. The remote site connectivity spans several large classified DoD networks like the Joint Training and Experimentation Network, Defense Research and Engineering Network, Distributed Mission Operations Network, Missile Defense Agency Network, and Air Reserve Component Network with numerous additional leased lines direct to other M&S facilities.

It is not uncommon for an event to utilize more than 30 remote sites including some that are overseas in Japan or Europe. Remote sites also include various DoD ranges where live assets can now be integrated into the environment for even more realism. Multination information sharing and MLS are in place to enable coalition participation and better integration of programs above the SECRET clearance level. A dedicated staff of expert government and contractor personnel stand ready to support each customer with the planning and execution of an event, including the capability to provide a set of white force personnel to control the event and a supporting staff of engineers to ensure everything is integrated, interoperable, and functional. DMOC continues its legacy of supporting the test community for ongoing JT&Es including active support for more recent JT&Es like Joint Command and Control of Network Enabled Weapons and Joint Test and Event Methodology. Several of the new JT&Es for 2008 and beyond are also considering DMOC's capability to support their objectives. For more information on the DMOC, please visit their web site at http://www.dmoc.kirtland.af.mil/ ❑

Michael J. Mora is a senior systems engineer for Scientific Research Corporation with 17 years of experience in systems engineering and modeling and simulation. Born and raised in Albuquerque, New Mexico, he obtained his bachelor of science degree in electrical engineering from the University of New Mexico in 1991. He began working at the Theater Air Command and Control Simulation Facility in 1992. He served as the director of engineering at the DMOC for over 10 years. He currently resides in Huntsville, Alabama working at the SRC headquarters for the Simulation, Test, and Instrumentation Division where he continues his work in modeling and simulation and Live-Virtual-Constructive integration for the Department of Defense. E-mail: mmora@scires.com

ITEA Journal 2008; 29: 132–134

Testing on the Fly:

World War II Field Expedients That Kept Aircraft in Combat

Frederick A. Johnsen
Air Force Flight Test Center Museum, Edwards Air Force Base, California

Normally, solving weapon system problems requires a deliberative process of arriving at engineered solutions, and of testing and evaluating them under controlled conditions. But this process became a casualty of war when damaged aircraft needed improvised fixing in flight. During the heat of World War II air battles, with no time to ponder and test a solution, aircrews relied on their own ingenuity to bring aircraft—in this case, B-17s and B-24s—back to base on a wing and a postulation.

The B-24 Liberators of the 5th Bomb Group operated in an atmosphere rife with disaster. The runway they used at Samar in the Philippines in 1945 was packed coral, sharp enough to cut tires, yet yielding enough to cause wheels with flat tires to gouge into its surface. Deep ditches paralleled the runway, to carry off rainwater and to keep the runway functional. When the Liberators experienced asymmetrical tire failures on the coral, they often turned abruptly toward the ditches. Pilot Robert D. Houghton of the 5th Bomb Group described a crash in a letter home: "…I saw a ship swerve on the runway, burst into flames and slide a thousand feet like a Roman candle."[1]

The 5th Bomb Group, self-nicknamed the Bomber Barons, drew patrol missions in the uncertain days of August 1945, after Japan reeled under two atomic bomb attacks, but before the formal surrender. The Group's task was to reconnoiter the China coast for any evidence of recalcitrance by Japanese forces, and it took a lot of gasoline to fly from Samar to China and back. On one such sortie, a loaded Liberator had sufficient speed to give good controllability on take-off, until the left main tire blew out. Just clearing the coral at about 6 AM, the bomber was lucky to be airborne, but its crew still faced the predicament of an inevitable return to the same treacherous airstrip that had claimed lives and airframes before. A different Liberator was sent on this patrol mission in its place.[2]

Figure 1. Lt Col Albert W. James helped devise ways to land B-24s with flat tires on a crushed coral runway in the Philippines (Albert W. James collection)

But what to do about the crippled B-24 in the air? Its fate was in the hands of Lt Col Albert W. James, who brought some prior Wright Field test and evaluation savvy with him when he joined up with the 5th Bomb Group in the Pacific. The group's last wartime commander, James worked diligently with his men and their B-24 Liberators. He kept a command

Figure 2. This 5th Bomb Group B-24L Liberator shows the single-tire main gear at the end of a long strut beneath the wing. With one main wheel flat, the Liberator naturally tended to slew in the direction of the damaged tire. (Albert W. James collection)

jeep with aircraft radios so he could be in communication with his aircrews, and on this day talked to the men inside the bomber stricken with the flat tire. He knew the peril: "When you had a blowout, you usually lost everybody in the plane…" he recalled in a 1979 interview. The crew had about 10 hours to ponder their predicament as they burned off the gasoline intended for the China patrol. By that time, their bomber would be several tons lighter. If nominally less prone to catching fire in that configuration, this was insufficient comfort to the men on board or Colonel James. The crew—even those members not required for landing the Liberator—shunned the option of bailing out, a measure considered last-chance by many aviators.

In this instance, James questioned the crew at length to ascertain that the left main tire was in fact deflated. Only then did he advise the men to use physics to improve their chance of survival, by deflating the other main tire to bring the bomber's landing gear friction back into symmetry. "I didn't want to be in the position of shooting out the good tire, and then discover the other was good too," he explained. The B-24's right waist gunner, in an act of faith, removed one round of .50-caliber ammunition from its linked belt and inserted it into the breech of his machine gun. Taking careful aim at the extended good mainwheel, he punctured it with irrevocable finality. "It was the only thing I figured they could do and land and stay on the runway," Colonel James said.[3]

Around 4 PM on that humid August day, the crippled Liberator was sufficiently lightened of gasoline to permit a landing on two flat main tires. The pilot approached at about 110 miles an hour, then slowed as the B-24's Fowler flaps were lowered and extended to increase the wing's effective area. Colonel James coached the pilot to use the outboard engines to steer the lumbering aircraft if it veered one way or another as it gouged into the coral, since the impulse to tap the brakes would have an unpredictable and dubious effect on two flat tires. Moreover, as speed bled off, the Liberator's large twin rudders would lose aerodynamic authority to steer the airplane. None of this was easy; Colonel James knew there was "a terrible lag which must be anticipated by the pilot before an engine revs up enough to give the necessary thrust to steer with."

The theory worked—the B-24 with two flat main tires stayed true as it mushed down the runway. It ground to a halt quickly, using less than half the runway, causing only minor damage to its landing gear.[4]

The Bomber Barons were challenged another time by the blowout of a nose tire on takeoff. The Liberator climbed into the muggy sky while its fate was pondered. With two good mainwheels, the decision was made to land it nose high, keeping the flat tire off the coral. Crewmembers not needed for landing chores congregated near the aft bomb bay, close enough to the center of gravity to keep the B-24 manageable on landing, and poised to scramble to the tail once the mainwheels touched ground, adding human ballast to keep the nosewheel aloft. The low-slung B-24s were built with a tailskid in case they rotated too far aft, and the skid's presence gave the crew some assurance their scheme would work without causing mayhem. "The plan was that as soon as he touched down, the crewmembers would get as far back into the fuselage as they could," Colonel James explained. When the moment came, the pilot kept his control wheel hauled back in an effort to keep the nose high as long as

Figure 3. Once its structural integrity was lost as it left the coral airstrip, this 5th Bomb Group B-24 displayed the catastrophe that attended runway mishaps (A.B. Goldberg collection)

effective elevator forces remained when the bomber rolled down the runway on its mainwheels. With brakes out of the question—they would cause the B-24 to pitch down onto its flat nosewheel—the Liberator relied on the weight of the crewmen in the aft fuselage to exert force on the tailskid, plowing a furrow in the coral runway surface until it stopped, its nosewheel never touching ground.

The fliers crowded into the tail of the bomber were liberated only when a maintenance trailer arrived and was positioned under the aircraft's nose, with a tire placed as cushioning. One by one, the men eased their way forward like mass weights on a scale, until the B-24 rotated gently on the fulcrum of the main gear and came to rest on the trailer, doing no harm to itself or its crew.[5]

Other theaters of war witnessed similar, in-the-moment test and evaluation trials. Photographically documented in England and Italy, the use of crew parachutes as braking devices kept more than one Fortress and Liberator from careening off the runway when hydraulic failures rendered wheel brakes useless. In the days before drogue and braking parachutes were built into high-performance warplanes, bomber crews learned how to take advantage of the symmetry of waist windows on the sides of the fuselages of B-17s and B-24s by tying crew parachutes to the waist gun mounts and unfurling them once the free-wheeling bomber touched down.

In the end, what these extemporaneous efforts lacked in test-and-evaluation discipline, they made up in ad hoc ingenuity. They also appealed to the airman's old friend, luck. ❑

Frederick A. Johnsen is the new director/curator of the Air Force Flight Test Center Museum at Edwards Air Force Base, California. His career includes 18 years as an Air Force historian. E-mail frederick.johnsen@edwards.af.mil

Endnotes

[1]From correspondence sent by Robert D. Houghton to his father during World War II. Copies obtained by the author from Robert Houghton.

[2]Information about the 5th Bomb Group landing improvisations is included in interviews conducted by the author with Lt Col Albert W. James, USAF (Ret) between 1970 and 1979. Portions of this material were published in Echelon magazine, Vol. 1 No. 2, March–April 1979.

[3]See note 2.

[4]The Bomber Barons requisitioned B-29 Superfortress main tires when possible, since they were the same size as B-24 main tires, and were reputed to be stronger to support the heavier B-29. Nonetheless, tire problems plagued the loaded B-24s at Samar.

[5]See note 2.

ITEA Journal 2008; 29: 135–139

AFOTEC:
Creating Active Involvement and Institutionalizing Early Influence

MG Stephen T. Sargeant
Commander
U.S. Air Force Operational Test and Evaluation Center (AFOTEC)
Kirtland Air Force Base, New Mexico

The Air Force Operational Test and Evaluation Center (AFOTEC) faces many challenges in an ever evolving acquisition environment. However, these challenges present fresh opportunities for the Air Force's Operational Test Agency to refine its operations and responsiveness in order to enhance our ability to ensure warfighters are delivered the capabilities they need, when they need them, to allow our Airmen, as well as our joint and coalition partners to fight more effectively and with less risk.

As a nation that has been in continuous combat for more than 16 years, longer than World War II, Korea, and Vietnam combined, our Airmen and joint and coalition partners are engaged in the global war on terror and our Air Force is actively searching for ways to rapidly enhance our effectiveness at all levels. The warfighter is demanding the entire acquisition community rapidly develop, test, and field increasingly complex and urgently needed weapon systems despite a reduced force.

Therefore, AFOTEC is focusing our efforts on institutionalizing Early Influence in the air, space, and cyberspace domains; establishing credible AFOTEC liaison officers at the Air Force Materiel Command and Air Force Space Command Product Centers; and aggressively creating the conditions for combined development and operational testing. These initiatives are all aimed at improving our ability to ensure that required warfighting capabilities are delivered within cost and schedule constraints whenever possible. The acquisition community has talked about the *Early Influence* concept for over two decades using many definitions. AFOTEC's goal is to clearly define Early Influence, establish more robust Early Influence activities as soon as possible, and institutionalize Early Influence across all appropriate instructions and regulations.

Early Influence is not a new concept. However, the practical application has proved to be challenging and inconsistent due to the lack of definition. AFOTEC has engaged in programs across the air, space, and cyberspace domains in an inconsistent manner throughout its history. Even the operational testing community does not share a definition of Early Influence. Therefore, we begin this article by defining what we mean by Early Influence.

Defining early influence

Early Influence is AFOTEC's formalized approach to refine capability requirements and acquisition strategies, and then develop early integrated test and evaluation (T&E) strategies and plans. We don't define requirements, but we can help refine them. If we get involved early, even before Milestone A, we can ensure requirements are testable, measurable, and operationally relevant.

Early influence provides AFOTEC the greatest opportunity to affect emerging capabilities and is based upon the premise that issues discovered early, before we have a formal program, are more easily resolved and often less costly. It costs far less to identify and fix problems while acquisition strategies are still in the planning stage and designs are still in development. The warfighting, acquisition, and T&E communities working together early and throughout a program can enable this Early Influence approach.

We begin applying Early Influence standardized methodologies prior to Milestone A by engaging in the capabilities based assessment process. The best opportunity to influence warfighting capabilities is when solutions are still being analyzed. Through formal reviews of the early Joint Capabilities Integration and Development System (JCIDS), documents such as the Joint Capabilities Document (JCD) and Initial Capabilities Document (ICD) we have opportunity to influence capabilities before a material solution, or mix of solutions, is selected. By joining the operational

T&E (OT&E) professionals with the other players early and often in the acquisition program, we increase communication and coordination enabling increased teamwork, leading to fewer surprises in the later part of the acquisition process.

We recognize there are three areas that can interject change after the inception of a program and before the fielding of the program. These inevitable changes are

(1) Technology continues to advance;

(2) User growth through real world experience;

(3) The nature of combat due to changing threats.

These three factors initiating change in the course of a program must be accounted for in a transparent way if we are to avoid surprises at the end game. They can have an effect on the capabilities of systems and their testing — early and constant communications will ensure that documents and plans can be modified as necessary to keep pace with the changing world. The world does not stop evolving once we publish our requirements or a test plan, and we must remain flexible enough to react.

By institutionalizing the Early Influence approach, we are helping to stress and refine requirements from a testability and measurability standpoint, as well as from an effectiveness and suitability standpoint. I want to emphasize that although AFOTEC does not define requirements for emerging capabilities we can assist in refining requirements. We will work with all stakeholders to ensure requirements are operationally relevant and realistic, and can be tested. Program success is reached when the required capability is delivered into the warfighters' hands, as close to on schedule and budget as possible. Executing Early Influence may seem simple on the surface, however, in order to execute effectively we are revamping our organization and processes to move beyond an era of much discussion and little action.

Managing early influence

AFOTEC has a mature policy that defines Early Influence as a "major operational test and evaluation phase." Early Influence is the first phase of a multi-phased approach of our involvement that then leads to planning, execution, and finally reporting. However, much of the early interaction with the community was being conducted only by our headquarters personnel and was not very robust.

Additionally, we have developed end-to-end Early Influence processes to ensure consistency and repeatability and we will continue to evolve these processes based on lessons learned. AFOTEC's Early Test Operations Division has traditionally led our Early Influence activities and their primary responsibility included starting the initial test planning processes. We are now refining our initial test planning efforts to get as much operational testing data as possible during the developmental testing phase wherever possible. And, we are shifting much of the Early Influence responsibility from the headquarters to our detachments where our hands-on testers live, to involve current and future test directors. The headquarters will then support our detachments in this role.

We've also tailored our training to address Early Influence activities, and we now provide formal training for Early Influence with an emphasis on operational suitability. While considering the entire system lifecycle, our primary targets include operational capability requirements, early integrated test planning, and acquisition strategies to support delivery of the capabilities required by the warfighter.

We are also developing metrics to track our early influence efforts. Again, we are moving away from a primarily bureaucratic process to effect early and active involvement. We have in place the essential elements to influence programs. Now we'll take a look at one of the first steps, initiating involvement in an acquisition program.

Involvement determination

A number of considerations go into making an acquisition involvement determination ranging from statutory mandates to multi-service participation. Operational risk is always a consideration. New capabilities that result in "game changing" operations such as new or significantly enhanced mission areas warrant the fidelity and rigor of the testing AFOTEC brings to the program. However, enhancements to existing capabilities to be used under current concepts of employment and support may be more suited for Major Command testing.

AFOTEC has a robust involvement determination process and in the past we've normally identified programs in one of two ways: we would either be asked to become involved or we would identify programs ourselves for involvement. Our involvement threshold has changed and starting now we are not waiting to be asked. If we're being asked, we're late. Self-identification is where AFOTEC is placing additional emphasis and we have several initiatives underway to strategically position AFOTEC liaisons as active conduits between AFOTEC, the Major Commands, and the Product Centers. We want to discover emerging programs as early as possible, make rapid involvement determinations, and then get involved early and often to assist in promoting program success.

Capability requirements

A major element of effective Early Influence focuses on capability requirements development. The external expectation is for testers to focus on the testability of requirements supporting early milestones or other key decision points. Although they are based on valid warfighter needs, the parameters are often not defined very well and this is where AFOTEC can add value to the process. Testability is not our only focus, as we also look for completeness of requirements such as how well the key system attributes and associated thresholds address the capability gap. The refinement, rigor and fidelity we add to requirements leads to a better product delivered to the warfighter on time and on cost. In addition, while refining the requirements, we are simultaneously looking for ways to save time and money by creating opportunities to gather operational test data during the developmental test phase of testing.

AFOTEC is a core member of the Air Staff's High Performance Team and takes part in the development of initial capabilities documents and capability development documents. The document review process ensures focus on lifecycle management issues such as reliability and maintainability, logistics supportability, and training. While AFOTEC is not a signatory on these documents, we provide face-to-face and written comments and suggestions. These early document reviews help the users refine their capability requirements and lay the foundation for initial test design. AFOTEC is a voting member of the Air Force Requirements Oversight Council where final approval of these documents is achieved.

We are aggressive in our Early Influence role reviewing all of the JCIDS documents. We recently made critical comments on the Joint Heavy Lift ICD. We addressed the lack of logistic focus, and specifically recommended mission and sortie generation, material reliability, training, and other related integrated logistics elements be addressed in the ICD. These areas are significant enabling attributes and capabilities for the eventual Joint Heavy Lift solution, and should be taken into consideration early.

Initial test design

Initial test design is another focus of Early Influence. It is a systematic approach to take the test teams from capability requirements to credible OT&E constructs which, when executed, will yield the final data required by decision-makers to make program decisions.

There is no panacea to how and when testing is done, but there are opportunities where more test data can be pulled from training and actual combat sorties. The CV-22 is going through current testing with operators and the test community working together. When we have reached the level of T&E that the warfighter needs, we can issue reports that are relevant to pressing needs such as looming deployments. We are flexible enough to schedule the rest of the required testing when the test assets are once again available. So in this case, we can complete the required testing and also support the warfighting customer to fulfill his mission requirements at the same time.

We strive to use Integrated Test Teams (ITTs) to develop test designs, and we execute rigorous design efforts for Test and Evaluation Strategies for Milestone A, and Test and Evaluation Master Plans for Milestone B. In the past, inputs were often based on the experience of a few subject matter experts without using standard processes and that led to OT&E inputs with a significant number of unknowns in early documents.

Initial AFOTEC test designs are based on envisioned concepts of operations and support, designed around the operation the user intends to employ the system. We've created early test designs based on a wide range of operations from combat operations to noncombat information technology systems used for finance, and personnel management. AFOTEC designs test around the operation, but scopes the operational testing to the system.

As we design the test around the operations, high-fidelity system characteristics are not critical at this point. Designing around the operation enables very early test designs and material solutions can evolve from initial expectations. The result of early planning rigor is fewer unknowns, higher fidelity test resource projections, and early identification of test capability shortfalls that will have to be overcome. For example, initial test design was completed on the KC-X and CSAR-X programs before either program had completed the down-select process to a specific platform. We have also proven that the process can be executed on very short timelines. We recently tested the Laser Joint Direct Attack Munition with Air Combat Command on a timeline based on months, not years, as a good example of our rapid test capability and flexibility.

AFOTEC does not accomplish these initial test designs in a vacuum. Much like the Air Staff's High Performance Teams, AFOTEC uses a core team composed of both internal and external participants.

Core team approach

The core team approach invites those with a vested interest to take part in the initial test design process and is consistent with the integrated test team approach. The core team might include lead major commands, other service operational test agencies, as well as other members from the operations, acquisition, and test and evaluation communities. In some cases it will be the first time all of the integrated test team members come together. Because AFOTEC executes initial designs before material solutions are selected, responsible test organizations or developmental testers may not have been selected yet. In these cases, AFOTEC will now invite the center test authorities to these meetings. Even though operational testing is the final acquisition phase prior to the system's fielding and deployment, involving operational testers early in the entire process can lower risk and increase the chances of successfully conducting the operational test phase of a program.

It is essential for the success of any program that all participants, that include the major command sponsor, product center, program manager, contractor, developmental testers, and operational testers, collaborate early in the program, well before test articles are produced, to ensure the success of the program. Each participant brings relevant information to the table and takes away a better understanding of the projected operational test and evaluation phase. The user communities bring current operational expertise and requirements clarity, while the program offices bring specific acquisition information and clarify questions about acquisition decisions, schedules, and actual capability increments. When developmental testers participate, there is a better chance of seeing operational testing data points in developmental test plans. Developers can provide details on the system under development as well as provide their interpretation of operational requirements.

Initial test design

Early involvement in developing OT&E designs inherently results in timely planning. AFOTEC initial test designs are driven by the need for confidence in our results. By applying rigorous, repeatable processes and gaining community buy-in during initial test designs, decision risk is addressed and credibility is achieved. Instead of looking at the lack of requirements and employment or support concepts as rationale to not become engaged in programs, I've charged my staff to leverage our early influence mindset and use the information gained during our utility assessments to refine the capability requirements, evolve the employment concept of operations, and help develop the support and training concepts. Essentially, our testers will be part of the solution when it comes to supporting rapid transition from technology demonstrations to programs.

Defining the operation and selecting operational test events based on operational factors allows us to design operationally relevant test scenarios. Initial AFOTEC test designs are based on envisioned concepts of operations and support and designed around the operation the user intends to employ the system. These scenarios are initially developed as end-to-end operational activities and the first choice is always live field testing. When field testing is not practical, we consider other methods like modeling and simulation. The result of early planning is fewer unknowns, higher fidelity test resource projections, and early identification of test capability shortfalls. Operational sufficiency and technical adequacy are achieved by addressing these areas early.

If there will be a multiservice OT&E effort, representatives from our sister-service operational test agencies attend to ensure service-specific interests are addressed during early test designs. We currently have about 45 multiservice OT&E efforts in progress.

Bringing the right people together at the right time is essential toward meeting initial test design expectations. This approach also works in the space program acquisition process. We have several initiatives underway to make OT&E more relevant in the unique acquisition process for space programs. Specifically, we are looking to identify acceptable methods to execute more OT&E in less than completely operational environments — that is, prior to launch. This requires getting involved early, increasing influence on the development testing design to increase the likelihood of acquiring useful OT&E data and a reassessment of risk tolerance. AFOTEC will become much more involved in the development work and testing that goes on in government and corporate labs today to accomplish this revised testing strategy. We are looking at where and how we can have a greater influence on developing operationally sufficient test capabilities in the Space realm. These capabilities should be useful for both development and operational testing efforts. Typically, we have only become involved once systems were on orbit, so it was more or less a stan-eval effort as our reports were not informing acquisition decisions.

Finally, we are establishing a liaison officer position in place at the Space and Missile Systems Center (SMC) at Los Angeles AFB, California, to work the full range of programs as they emerge. Both AFOTEC and the SMC will gain from this, but ultimately, our warfighters will get the greatest benefit. AFOTEC is changing the way we do operational test in space and is working toward partnership and teamwork.

Strategic initiatives

While AFOTEC is actively engaged in Early Influence initiatives at the tactical level, we are also working these initiatives at a strategic level. AFOTEC is heavily involved in the Air Force Smart Operations 21st Century initiative, commonly referred to as AFSO21. As an organization, we are looking at our processes from beginning to end. This includes our involvement with the Developing and Sustaining Warfighting Systems process. The author is a cochair with the Air Force Materiel Command on the Test and Evaluation subprocess team. They are collaborating on initiatives approved by the AFMC Commander and the Chief of Staff of the Air Force, to increase confidence in early acquisition efforts by institutionalizing Early Influence across the developmental and operational test communities. AFOTEC already had rudimentary Early Influence policy and processes in place and was organized to specifically address early acquisition efforts. The early preparation allowed them to play an integral role in the Milestone B initiative designed to foster more combined developmental and operational test plans. Importantly, this work has given the AFOTEC the opportunity to hone its processes.

While the Air Force instructions did a reasonable job of addressing most issues associated with acquiring and testing new systems, we worked to strengthen them across the entire test and acquisition community to institutionalize early influence activities supporting Milestone B. AFOTEC recently participated in the annual Test and Evaluation Policy Conference in Washington, D.C., and as a result Early Influence is now more clearly codified and institutionalized in Air Force and AFMC instructions (AFI 99–103, Capabilities Based Test and Evaluation; AFI 63–101, Operations of Capabilities Based Acquisition System; AFMC 99–103, Test Management).

AFOTEC is now an advocate of institutionalizing actionable Early Influence, not just continuing to talk about it. AFOTEC can influence all areas early and consistently throughout the life of a program, to include addressing known life cycle management costs and accounting for changes as a program matures. Through early and continuous communication and coordination, AFOTEC will benefit from high confidence planning and potential schedule and cost savings. AFOTEC's goal is increased teamwork leading to fewer surprises at end of the process.

We are working closely with my detachment commanders to increase their role in developing early partnerships with the warfighting and acquisition communities. We will capitalize on their expertise to further enhance our ability to positively affect programs and early. The need for Early Influence is even greater now because of the long war we are engaged in. Increased communication and coordination leading to greater team work and fewer surprises is what AFOTEC is striving to achieve, and we are increasing our efforts to work more closely with the acquisition and warfighting communities. AFOTEC's vector is clear. We intend to have a positive influence through early activities, and to that end we will get involved early with clear priorities. Our personnel will make timely involvement determinations and apply appropriate rigor to requirements development and test design. Our test designs will seek opportunities for combined DT and OT testing whenever possible.

Creating active involvement and institutionalizing Early Influence provides better and more capable systems to the acquisition decision makers and the warfighters sooner. The need for Early Influence is even greater now because of the long war we are engaged in. The bottom line is — these efforts will help the acquisition community to provide better, more capable systems to the warfighter … sooner … to accomplish their mission more effectively, with less risk to our Airmen and joint and coalition partners! ❑

MG STEPHEN T. SARGEANT, USAF is the commander of the Air Force Operational Test and Evaluation Center (AFOTEC) at Kirtland Air Force Base, New Mexico. Gen Sargeant reports to the Air Force chief of staff regarding the operational test and evaluation (OT&E) of more than 200 major programs being assessed at 22 different locations. He directs the activities of more than 950 civilian and military personnel. As a member of the T&E community, he works directly with the Office of the Secretary of Defense and Headquarters U.S. Air Force, Washington, D.C., to ensure realistic, objective and impartial operational testing is conducted on Air Force and Joint-use systems. MG Sargeant has served as the commandant of the U.S. Air Force Weapons School at Nellis AFB, Nevada, commanded the 8th Fighter Wing at Kunsan Air Base, South Korea, and the 56th Fighter Wing at Luke AFB, Arizona. He has also served in numerous Air Force, Joint, and Coalition staff assignment, including 18 months in Baghdad, Iraq as the C-5 for CJTF-7 and MNF-I. He is a command pilot with more than 3,000 flying hours in the A-10/A and F-16A/B/C/D. E-mail: steve.sargeant@afotec.af.mil

ITEA Journal 2008; 29: 140–141

Test and Evaluation:
Transforming to Enable Successful Systems Development and Fielding

MG Roger A. Nadeau
Commanding General, U.S. Army Test and Evaluation Command, Alexandria, Virginia

The key to making transformational changes in the test and evaluation process is one program at a time. Army Test and Evaluation Command (ATEC) is looking at itself to further applications of the principles of testing and evaluation discussed in this article. Personnel of the Command are being further trained in these principles to facilitate a closer working relationship with industry, government agencies, and Project Program Managers.

Army Transformation is not just two words on a briefing chart — it is a reality. The U.S. Army has been in the act of a major transformation for several years. Every organization in the Army is changing its processes and procedures to facilitate the transformation process. The Army Test and Evaluation Command (ATEC) is no exception.

Although ATEC has implemented several transformational changes in support of an army at war, there is much more that can be done. It is not change for the sake of change. Every change should, to some degree, save time and/or money, without sacrificing the quality of our mission.

ATEC's mission is simple — assess the performance capabilities of every piece of equipment used in any way by Soldiers. The two basic questions we ask are: Does it work? How do I know?

Our end-state assessments address effectiveness, suitability, and survivability. Based on that assessment, Army leadership decides if Soldiers will ultimately use the item.

Throughout the process of doing our job, we work closely with industry, requirements developers, Program Executive Officers, rapid equipping organizations, and others who are involved in determining the need for or developing materiel solutions to provide capabilities to the Army and other Services. Although usually productive, the relationship between ATEC and the aforementioned groups can improve.

Changing the perception

The major challenge right now in improving our relationship with others is one of perception. We believe most agencies see ATEC as a hindrance, rather than an enabler, in providing equipment to satisfy Soldier needs. Some of the perception challenges are rooted in current policies and regulations. As a general rule, most mandated test events, developmental and operational, are scheduled just before a major milestone decision; and no independent agency test events are required before Milestone A at all. Industry and government materiel developers must work through technology and design growth challenges early in the program, often challenged by schedule compression and funding constraints, leaving no schedule or funding flexibility by the time a mandatory test event occurs. The time and cost due to the test and evaluation (T&E), and performance shortfalls revealed in test, invariably creates the perception that the test agency is an impediment to progress. The challenge — change the perception by involving the T&E throughout the acquisition lifecycle as a proactive part of the process.

ATEC is aggressively reaching out to industry and government agencies, where appropriate, to prove the value of bringing the test community in early in the requirements development, equipment design, and development processes. If brought into the process early and used throughout the acquisition lifecycle as a partner, we are convinced we can aid in building quality into the product from the beginning, saving time and money and minimizing costly redesign when major performance problems go undetected until late in the developmental cycle.

Early and constant involvement in the acquisition process

As we work to earn (and we mean earn) the right to be brought into the acquisition process earlier, ATEC is also revisiting how we test. The current process is generally linear. Developmental Testing (DT) is separate from, and always prior to, Operational Testing (OT). The evaluation process in many cases begins after testing is complete. Industry or other government agency test data are generally not considered. Early coordination between industry, government materiel developers, and ATEC concerning T&E methods and expectations often fails to occur. These traditional processes are inefficient, archaic, not transformational, and are no longer acceptable. They waste time and money and certainly do not foster team spirit.

Change needs to start from the beginning. ATEC's involvement should start in developmental activities prior to Milestone A. How much involvement will vary according to the nature of the end item. From that point, ATEC's involvement should be constant through every milestone and all aspects of the phases between the milestones. Generally speaking, we believe Project/Product Managers should always have to deliberately decide not to have an ATEC representative in meetings vice the alternative of always having to remember to invite an ATEC representative to a meeting.

Success for ATEC, and the Army, is being recognized as a trusted T&E advisor to the acquisition/developmental team to assist in program cost, schedule, and performance risk management.

The same approach applies to ATEC involvement with the defense industry. The government can include lessons learned from T&E of similar equipment in the advertisement of Requests For Proposals to aid industry proposal responses. Once under contract, dialogue with ATEC can help guide the thought process in contractor testing which, in turn, can influence progressive design changes to enhance system performance. Early, coordinated government/industry test planning can ensure common test procedures, which will allow for potential use of contractor data in government test performance reporting, minimize repetitive testing, and better indicate system readiness for government testing. ATEC involvement in contractor testing also allows for early development of data collection and modeling and simulation tools to support informed development and acquisition decisions throughout the life of the program. Once an item enters government DT/OT, appropriate timely feedback to the prime contractor can be helpful in identifying corrective actions to fix emerging problems, saving everyone time and money.

Internal transformation

Processes internal to ATEC need to change as well. Consistent with the December 22, 2007 Test and Evaluation Policy Revisions memo jointly signed by the USD (AT&L) and the DOT&E, how we do what we do must be less linear and less parochial than in times past. Core T&E principles spelled out in the OSD policy memo include:

1. DT/OT integration throughout program lifecycle
2. Compare to current capabilities
3. Focus – Measure improvements to capability
4. Evaluate in mission context at time of fielding
5. Use all available information
6. Exploit the benefits of M&S

Transformational change internal to ATEC is mandatory to keep pace with the transforming Army. Anything less is a disservice to the Soldiers we ultimately serve. In the end, we are convinced that the quality of our services will increase, saving time and money too.

Is this easy? No.

Is it the right thing to do? Yes.

Will it take time? Yes.

The Army Test and Evaluation Command is committed to this transformational challenge. The process has begun and is picking up steam every day! ❑

Acknowledgment

James Wells, Assistant Technical Director, US Army Test and Evaluation Command, contributed to this article.

MG Roger A. Nadeau took command of the Army Test and Evaluation Command (ATEC) on June 28, 2007. Prior to accepting command of ATEC, he served as the commander of the U.S. Army Research, Development and Engineering Command from October 2004 to June 2007. His other significant assignments include program executive officer for Ground Combat Systems; program executive officer for Combat Support and Combat Service Support; deputy for Systems Acquisition (DSA), Aviation and Missile Command (AMCOM); assistant deputy for Systems Management and Horizontal Technology Integration in the Office of the Assistant Secretary of the Army for Acquisition, Logistics and Technology; and chief of staff to the Assistant Secretary of the Army (Acquisition, Logistics and Technology)/Army Acquisition executive. For more on ATEC — www.atec.army.mil.

ITEA Journal 2008; 29: 142–146

OPTEVFOR:
U.S. Navy Operational Test and Evaluation Perspective: Collaboration in an Uncertain Environment

RDML Stephen Voetsch
Commander Operational Test and Evaluation Force (COMOPTEVFOR)
Steven Whitehead
Executive Director,
COMOPTEVFOR, Norfolk, Virginia

It is the mission of the Operational Test and Evaluation Force (OPTEVFOR) to ensure that all systems are as effective, suitable, and survivable as possible and convey their capabilities and limitations to the warfighter. In this regard, the test and evaluation (T&E) community has the unique opportunity to provide advice, guidance, and the wisdom gained from our exposure to the T&E of hundreds of systems providing warfighting capability over the past 60 years and the lessons learned from their successes and failures.

About OPTEVFOR

The U.S. Navy's OPTEVFOR, headquartered in Norfolk, Virginia, is the Navy's sole independent Operational T&E command. It is our mission to report to the Chief of Naval Operations (CNO) on the operational effectiveness and operational suitability of new and improved warfighting capability.

In effect, we're information providers. However, we do not limit our service to the final examination of a warfighting capability. We are active participants in all phases of product development providing operational insight and identifying both real and potential operational shortfalls to a system's end-state capability. As Dr. McQueary, (Director, Operational Test and Evaluation, Office of the Secretary of Defense) has said, "OT&E should be a period of confirmation, not discovery" and we can only achieve that by being there throughout the process.

OPTEVFOR
Who we are:
- Navy's Independent OT&E Agent
- Reporting Directly to CNO
- Information Providers to Decision Makers and Warfighters
- Supporting both Acquisition and Experimentation

How do we do that and with what resources? *Figure 1* is a high level view of OPTEVFOR's portfolio. This figure provides a view of how we are organizationally aligned, both internally and externally, and our resources. Bottom line is, fewer than 300 people, an operating budget of just over $11 million annually, and an average per year of $40 million of reimbursable funding from program managers.

OPTEVFOR is currently assigned OT&E responsibility for over 300 programs. These programs span the Navy's warfighting enterprises of aviation, surface, subsurface, command and control, and expeditionary warfare. The wide range of product types that support new and improved warfighting capability requires a diverse level of operational and technical experience within the staff at OPTEVFOR in order to adequately support our mission. This diversity is obtained by the continuing rotation, average 3 year tour, of fleet operational personnel that serve as the operational test coordinators and operational test directors.

OPTEVFOR initiatives

The Navy T&E community in general and OPTEVFOR specifically are actively engaged in developing methods, procedures, and policies to help ensure that the financial investments of the U.S. Department of Defense (DoD) and the Department of the Navy are provided the opportunity to generate the highest rate

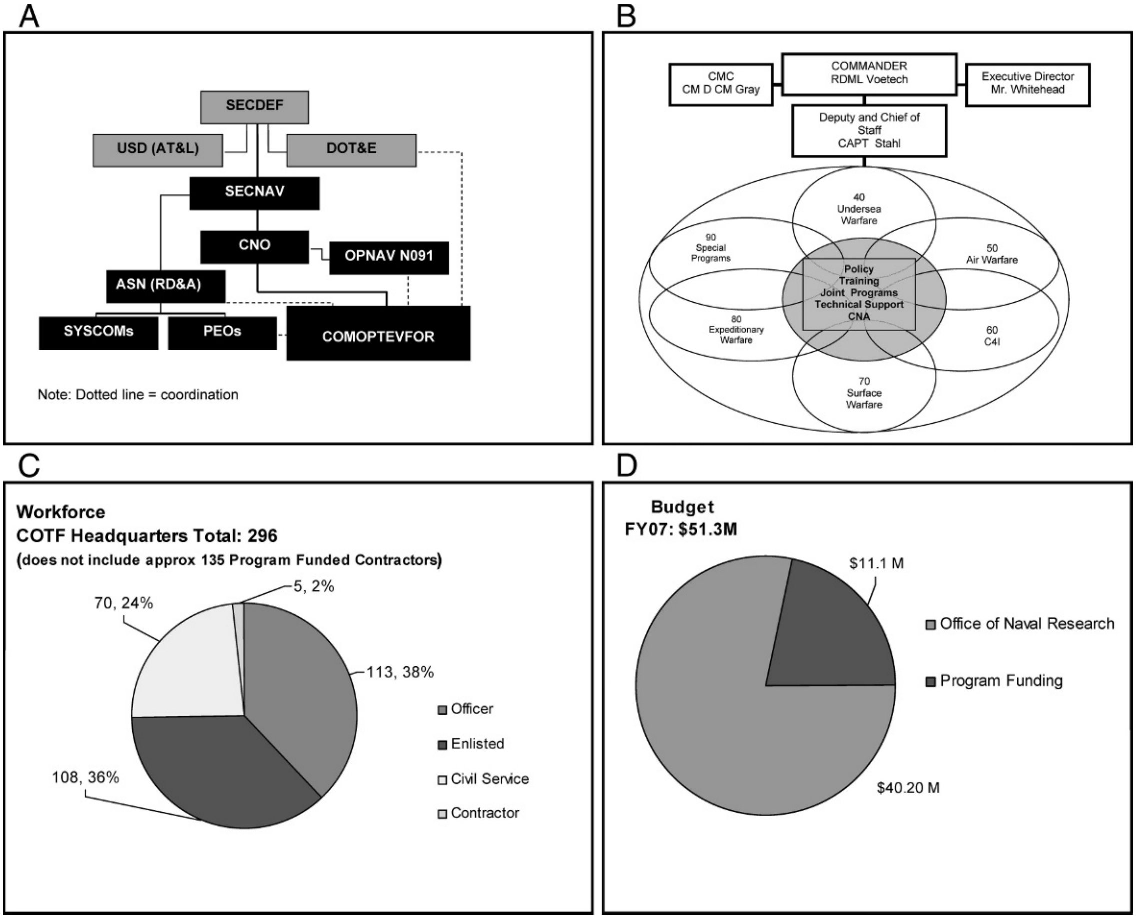


Figure 1. Operational Test and Evaluation Force portfolio

of return possible in warfighting capability and preservation of human life. Four specific initiatives currently ongoing are:

1. The Navy's enterprise approach to T&E (T&E Board of Directors)
2. T&E cycle time reduction
3. Mission-based test design
4. Integrated testing

Navy's enterprise approach to T&E

As the result of a recommendation to the chief of Naval Operations (CNO) in July 2005 that stated:

> "The need for a T&E capability that is synchronized with product procurement across the spectrum is critical to support the Navy's future war fighting needs. As part of the enterprise integration of T&E there is a need for a single Navy T&E process owner. An additional dimension is the need for a more effective and efficient business model in the relationship between government and industry, the desire to achieve synergism, and produce a "win-win" situation for both entities."

As a result of this recommendation, the Navy's Test and Evaluation Board of Directors was formally established by the Assistant Secretary of the Navy for Research, Development, and Acquisition in April 2007. This board of directors is composed of Flag and Senior Executive stakeholders from the CNO staff, Navy System Commands, Program Executive Offices, and the Marine Corps, and is co-chaired by the Department of the Navy T&E Executive and COMOPTEVFOR. It is the function of the Board of Directors to apprise department leadership of T&E enablers to ensure that the needs of acquisition programs are met and balanced with overall Navy warfare enterprise goals. The Board resolves issues among T&E enablers, recommends priorities to be executed by the responsible organizations, supports the Navy enterprise and integrated T&E process, and champions improvement initiatives to meet program requirements while continually improving T&E cost efficiency.

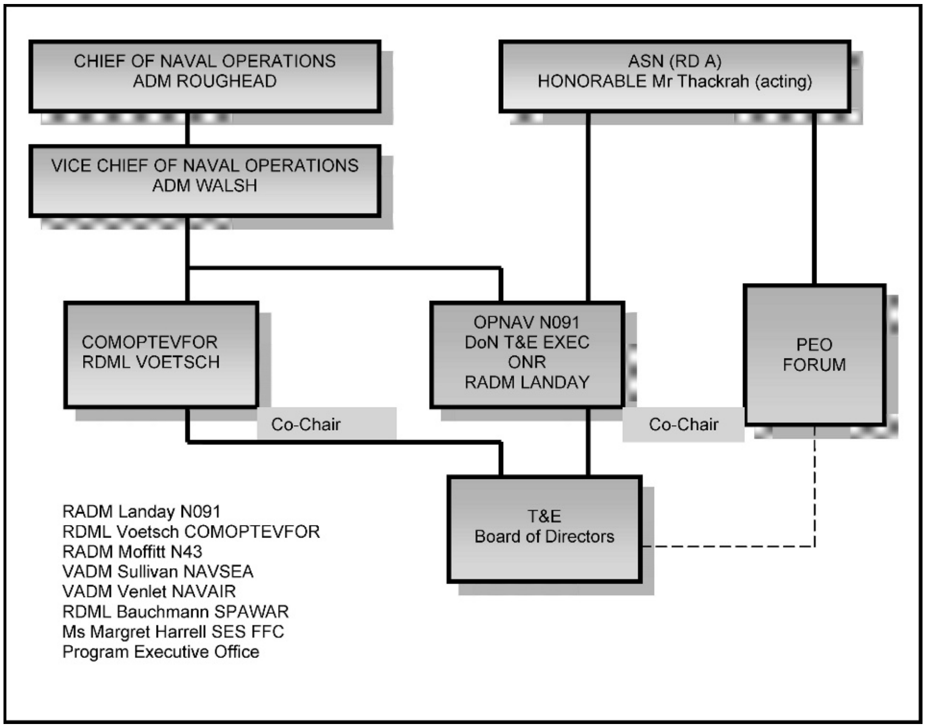


Figure 2. Navy test and evaluation board of directors

To this end, the vice chief of Naval Operations (VCNO), the assistant commandant of the Marine Corps, and the assistant secretary of the Navy for Research Development and Acquisition realigned the reporting requirements of the Navy T&E executive to not only the VCNO but also to the assistant secretary for Research Development and Acquisition. What this means is now the Department of the Navy, Navy, and Marine Corps, have a single process owner for T&E. And, that process owner reports to and represents not just the CNO but also the Navy's service acquisition executive on all T&E policy. This is a very powerful move because now there is a chain of command between the CNO, the person ultimately responsible by law for training and equipping the fleet, and the business decision maker on the acquisition side that provides the CNO with warfighting products.

T&E cycle time reduction

Over the past several years the decision cycle by the development and acquisition community to provide products to the fleet has, in many instances, been reduced beyond the ability of the operational tester to provide, in a timely manner, value-added information to the decision maker. Reduced cycle time of product development and fielding for highly software intensive systems now ranges from several months to less than a year. In addition, the final capability that is delivered to the fleet is more often than not, not determined until the platform is ready for deployment. *And let's never forget it is all about the warfighters needs!* The ability of the operational tester to plan, based on proposed yet uncertain final system capability and configuration, schedule warfighter resources, execute the test, conduct the analysis and write an evaluation report is severely hampered by the time it takes to conduct these activities in the long established methods and processes currently utilized (*Figure 2*).

Operational testing has, over the past 10 years taken on a more technical analysis of the capabilities provided by developers. The ability to collect system performance data at levels significantly below the operational level (that is the operator and supervisor level) has resulted in a belief that it is the responsibility of the operational tester to conduct extensive "failure analysis" on system performance when a shortcoming or failing is identified during test. This belief is counter to evaluating a system's performance capability at the operational mission level. By elevating the level of evaluation back to the operational mission level, it is possible to reduce the evaluation and report time on a warfighting capability. This elevating of the evaluation level, combined with leveraging all available data and information, wherever it is created, contractor, program office, or operational test will allow for significantly greater insight in a systems performance capability as well as allow for greater confidence in that system after it is deployed.

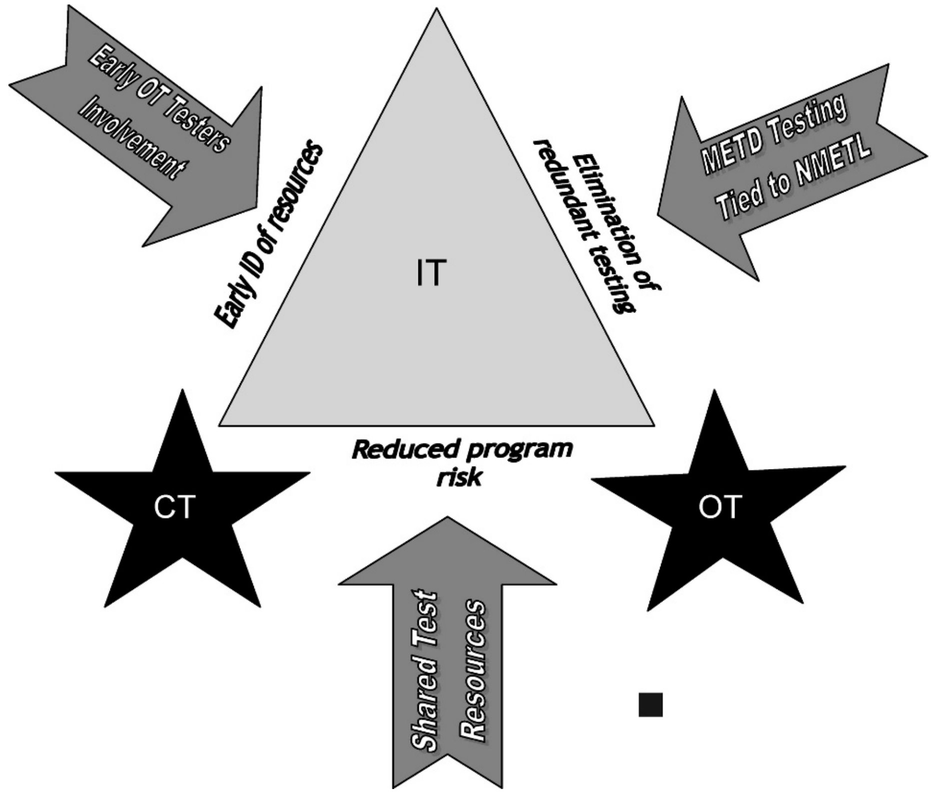


Figure 3. Integrated test

Mission-based test design (MBTD)

MBTD is a detailed and disciplined systems engineering approach to test design. Test design is done early in the acquisition life cycle (prior to MS B if possible). Simply put, the methods and processes we have developed utilize existing measures, the Universal Joint Tasks Lists (UJTS) and the Navy Mission Essential Task Lists (NMETLS) to collect data on the system and the operators' capability to execute the mission. The foundation document produced from this process, known as the OT framework, provides the groundwork for all subsequent detailed test plans. In fact, the OT framework contains enough detail itself to serve as a test plan. Test size is based on the conditions which impact the system and warfighter in their environment. Statistical methods known as *Factorial Designs* and *Design of Experiment* produce permutations of the tasks based upon the number of conditions.

Tasks are broken into logical executable segments called vignettes. Vignettes can be as small as a maintenance demonstration or as large as an end-to-end exercise. The end product out of the MBTD, the OT framework, contains a detailed description of test vignettes, conditions in which the test is to be conducted, the measures of success and a robust, detailed list of resources needed to execute. The resource list can then be fed into the T&E Master Plan.

Integrated testing

The May 2003 version of the DoD Directive on Defense Systems Acquisition specifically states "Test and evaluation shall be integrated throughout the defense acquisition process." What has been lacking until recently is the answer to the question "What does that mean?" For the Navy, that means, and I quote from the Navy's T&E Board of Directors, "Integrated testing is the collaborative planning and collaborative execution of test phases and events to provide data in support of independent analysis, evaluation, and reporting by all stakeholders particularly the developmental (both contractor when appropriate and government) and operational test communities." (See *Figure 3*) A policy memo is expected out soon on this subject.

We are also working with DOT&E and USD (AT&L) in revising the format for T&E Master Plans to reflect a more integrated approach.

How do we proceed?

All of this raises some very challenging questions. What is the impact of these initiatives on the industry? How does it affect the acquisition community and the total test community? How will it affect the government/industry relationship? Are there statutory or regulatory barriers to succeeding in these initiatives?

All good and valid questions, do we have the answers? No. Can we work together to come up with the answers? That's the whole point, collaboration. Meaningful and open dialogue and action about what we all know needs to be done but have avoided because it's believed too difficult. It is difficult, but not impossible. Government and industry have different agendas. Admit it, understand it, and move on. We once again are in a time when we are faced with significant challenges. This time let's look at these challenges as opportunities to move beyond our parochial stove piped interests and really do what we say: Support the warfighter. ❑

Rear Admiral (RDML) Steve Voetsch was born in Fort Lewis, Washington. He received his commission in 1979 from the United States Naval Academy and was designated a naval flight officer in 1980.

After initial F-4 Phantom training in VF-171 he was assigned to VF-74 where he completed his first deployment to the Indian Ocean on board USS Forrestal (CV 59), in the last East Coast F-4 Phantom squadron. In 1983 he transitioned to the F-14 Tomcat and completed a follow-on deployment on board USS Saratoga (CV 60) with VF-74, deploying to the Mediterranean Sea.

After a brief shore tour, he was assigned to VF-143 embarked in USS Dwight D. Eisenhower (CVN 69). From 1988–1991, RDML Voetsch served as the operations and maintenance officer in VF-41, operating from USS Theodore Roosevelt (CVN 71), flying numerous combat missions over Iraq and Kuwait during Operations Desert Shield/Storm and Provide Comfort. Assuming command of the "Fighting Diamondbacks" of VF-102 in September 1995, RDML Voetsch was directly responsible for leading the squadron toward earning the Battle "E", Safety "S", the coveted Clifton Award, Grand Slam Award, Tactical Air Reconnaissance POD System (TARPS) Award, and the NAS Oceana Athletic Award. RDML Voetsch commanded the "Grim Reapers" of VF-101 from July 1998 to December 1999. In July 2000 he reported as deputy commander, Carrier Air Wing One and assumed duties as commander, Carrier Air Wing One in July 2001.

Serving ashore, RDML Voetsch was assigned as a fighter instructor in VF-101. Other shore assignments include the Armed Forces Staff College and assistant Washington placement officer in the Bureau of Naval Personnel; aide/flag lieutenant to the chief of naval personnel, Admiral Ronald J. Zlatoper; two tours as executive assistant to the commander, North American Aerospace Defense Command and United States Northern Command. RDML Voetsch served as the deputy chief of staff for operations, training and readiness (N3/N7) on the staff of commander, U.S. Pacific Fleet from July 2005 to May 2007. On May 24, 2007, RDML Voetsch assumed command of Operational Test and Evaluation Force in Norfolk, Virginia.

Steven K. Whitehead currently serves as the executive director to the Commander, Operational Test and Evaluation Force (COMOPTEVFOR), U.S. Department of the Navy. He was selected to a senior level (SL) position as a senior executive on June 17 2001. He enlisted in the U.S. Navy out of high school in 1975 as a surface electronic warfare technician and continued to serve the Navy as a civilian while attending the University of Rhode Island, earning an undergraduate degree in electrical engineering. He transferred to the Naval Warfare Assessment Center (NWAC), Corona, California where he served in a variety of engineering and weapon systems analysis positions from 1985 to 1991. His experience includes weapon system and missile flight analysis in the Surface Weapons Department, Flight Analysis Division, Point Defense Systems Branch, performing engineering evaluations of Point Defense (RIM-7E, &H, &M and RAM) surface missile firings and weapon system exercises; project leader in the Systems Analysis Division, Special Projects Branch for the Vertical Launch ASROC weapon System, the Mk 116 Mod 0 Vertical Launching System, and the TRIDENT II (D5) Strategic Weapon System (SWS). The latter two projects directly supported DEPCOMOPTEVFOR, Pacific and COMOPTEVFOR, respectively. For the TOMAHAWK SEARA Branch, he continued his duties as NWAC project lead in support of OPTEVFOR for the TRIDENT II (D5) SWS, and was assigned the duties of TOMAHAWK quality assurance service test representative to the Cruise Missile Program Office, San Diego, California. He also served on the department staff, Surface Weapons Department, coordinating the establishment of an antisubmarine warfare division.

In May 1991, he transferred to OPTEVFOR Headquarters, Norfolk, Virginia, as an operations research analyst. He was selected as the deputy assistant chief of staff for C4I Systems in October of the same year and held that position until June 1996 at which time he was selected to assume the duties as the technical director. In November 2007, his duties were realigned to better support the commander and his position was designated as the executive director.

Steven Whitehead received his bachelor's degree in electrical engineering from the University of Rhode Island. He also holds a master's degree in management from Troy University and is a 1994 graduate of the Naval War College. His awards include the Navy Civilian Superior Service Medal, Navy Civilian Meritorious Service Medal, two Sustained Superior Performance awards, Special Mission Support award, and three Letters of Commendation.

ITEA Journal 2008; 29: 147–155

Operational Realism Via Net-Centric Test & Evaluation:
From Concept Development to Full-Rate Production and Sustainment

COL Ronald C. Stephens, Randon R. Herrin, Danielle Mackenzie, and Daniel W. Knodle
Joint Interoperability Test Command, Fort Huachuca, Arizona

As the Defense Information Systems Agency's Operational Test Agency since 1987, the Joint Interoperability Test Command (JITC) has integrated operational realism into major test programs and efforts through a variety of methods and strategies. Developing and using net-centric test approaches is a fundamental concept that JITC uses to enhance the operational testing environments for Information Technology and National Security Systems. Complementing JITC's net-centric testing strategies are three other key mission activities that enable JITC to assess and test Information Technology and National Security Systems in operationally realistic environments: JITC's roles and responsibilities with numerous military exercises around the globe, JITC's core mission of interoperability certification through testing, and expanding and executing the Department of Defense Interoperability Communications Exercise.

Based upon lessons learned since the middle 1990s (beginning with the Global Command and Control System), and lessons that continue to be learned with programs like Net-Enabled Command Capability (NECC) and Net-Centric Enterprise Services (NCES), the Joint Interoperability Test Command (JITC) has concluded that successful testing of Information Technology and National Security Systems (IT and NSS) in operational environments must incorporate a net-centric testing approach that accurately represents the Net-Centric Information Environment (NCIE) as defined in the Net-Centric Operations and Warfare Reference Model (NCOW-RM). This article addresses the JITC's current efforts of integrating operational realism into net-centric testing strategies and highlights this work through a discussion of JITC's substantial involvement with interoperability certification testing and military exercises around the globe. These testing activities involve systems and programs in various system acquisition lifecycle stages from concept development to full-rate deployment and even legacy systems in sustainment. Finally, as an illustration of assessing and testing IT and NSS in an operationally realistic environment, this article addresses the Department of Defense (DoD) Interoperability Communications Exercise (DICE) that JITC develops and executes with the Joint Staff (JS).

Background

JITC is a field element of, and directly reports to, the Defense Information Systems Agency (DISA). DISA is the DoD Agency responsible for planning, developing, and supporting the IT and NSS that serve the National Command Authority under all conditions of peace and war. JITC is considered by the Office of the Secretary of Defense's (OSD's) Director, Operational Test and Evaluation (DOT&E) to be the DoD's only joint Operational Test Agency (OTA). Counting the four Service OTA's, this gives the DoD a total of five OTAs.

JITC directly contributes to DoD combat operations success by providing interoperability test and evaluation (T&E) services for emerging and legacy IT and NSS, while also providing on-call support to the deployed warfighters. As designated by OSD and the JS, JITC is the only DoD organization with the mandate and authority to certify DoD IT and NSS to ensure they meet interoperability and net-readiness requirements for joint military operations. JITC's mission is to provide a full range of agile and cost-effective T&E and certification services to support rapid acquisition and fielding of global net-centric warfighting capabilities. Our vision is to be a world-class T&E organization that advances global net-centric testing in support of warfighting capabilities.

As DoD's sole joint interoperability certifier of IT and NSS, in addition to directly reporting to DISA, JITC has direct reporting responsibility to the JS. Also, since JITC is designated as DISA's OTA, we have a direct reporting responsibility to the OSD DOT&E.

JITC works closely with the DISA T&E executive, Dr. Steven Hutchison, and DISA's Test and Evaluation Management Center (TEMC) director, Luanne Overstreet. Among other key roles and missions, Dr. Hutchison, Ms. Overstreet, and the TEMC provide long-term, strategic management of DISA's T&E resources and investments and work T&E policies with JS and OSD. JITC reports directly to the DISA chief-of-staff, COL (USA) Alan R. Lynn.

Operational realism via net-centric T&E

Over the past decade, the DoD acquisition and testing communities have slowly but surely realized that acquisition and testing concepts and approaches for standard weapon systems (e.g., ships, tanks, and aircraft) cannot be used effectively for IT and NSS systems. This is particularly true with respect to testing net-centric IT and NSS in operationally realistic environments. As we transition to a fully net-centric enterprise, the operational environment for IT and NSS will be the global information grid (GIG). As a result, it is critical that a persistent, distributed test environment be stood up on the GIG to allow for the testing and certification of net-centric capabilities. Likewise, it is imperative that the operational community embrace net-centricity as the way ahead. JITC has recommended to JS that every program developing net-centric capabilities be required to address net-centricity as a critical operational issue. A program's ability to support net-centric operations by operating in the NCIE and, as appropriate, provide for the ongoing evolution and maintenance of the NCIE, is critical to realizing the DoD vision of transformation to a net-centric enterprise.

As a major step towards supporting our mission and our vision statements and for integrating operational realism into JITC's net-centric testing, JITC realigned in 2007 so that test portfolios would be better aligned with the DoD IT portfolio management mission areas *(Figure 1)*. This alignment enables JITC to use T&E as a means to measure the actual contributions of the portfolio against established outcome-based performance measures. In doing this, T&E becomes a means by which recommendations are made to continue, modify, or terminate the individual investments within the portfolio.

In partnership with DISA's TEMC, JITC is searching for more rapid and agile ways to test and certify net-centric capabilities. TEMC representatives on JITC capability test teams (CTTs) provide JITC CTT leads with recommended solutions for conducting agile T&E, improving collaboration, reducing redundancy, and speeding the delivery of capabilities to the warfighter. A key enabler to realizing these concepts will be the Federated Development and Certification Environment (FDCE). During FY07 and continuing into FY08, JITC has partnered with DISA to execute a series of T&E pilots using the NECC program as a test case. This has yielded a collection of test processes, tools, and lessons learned that will greatly benefit the JITC, and DISA's NECC program. Using the pilots as a learning process, JITC has also come to realize the necessity of the FDCE as a means to realize the capability T&E mantra of "one team, one time, one set of conditions." The FDCE is conceived as a virtual environment that is intended to address the challenges associated with concurrent and distributed Service management. Its purpose is to provide the policies, processes, and infrastructure that allow capabilities to be progressively refined, tested, and certified in increasingly rigorous situations leading to an operational environment.

The NECC Joint System Team composed of representatives from JITC, Service OTAs, DISA, and Joint Forces Command has begun using the FDCE as a means of streamlining the T&E lifecycle for NECC. The FDCE provides a collaborative workspace on the GIG for net-centric capabilities to register, define requirements, plan and track test status, and share information. The use of this tool allows for testers across the systems acquisition lifecycle to leverage each others work, as well as influence and define test events so that they meet the objectives of multiple stakeholders. In doing this, operational testers are able to provide guidance and direction on incorporating operational realism into developmental testing activities. This approach can be seen in *Figure 2*, JITC's Lifecycle T&E Approach. By having early JITC involvement in the acquisition lifecycle, joint requirements are better defined to support the

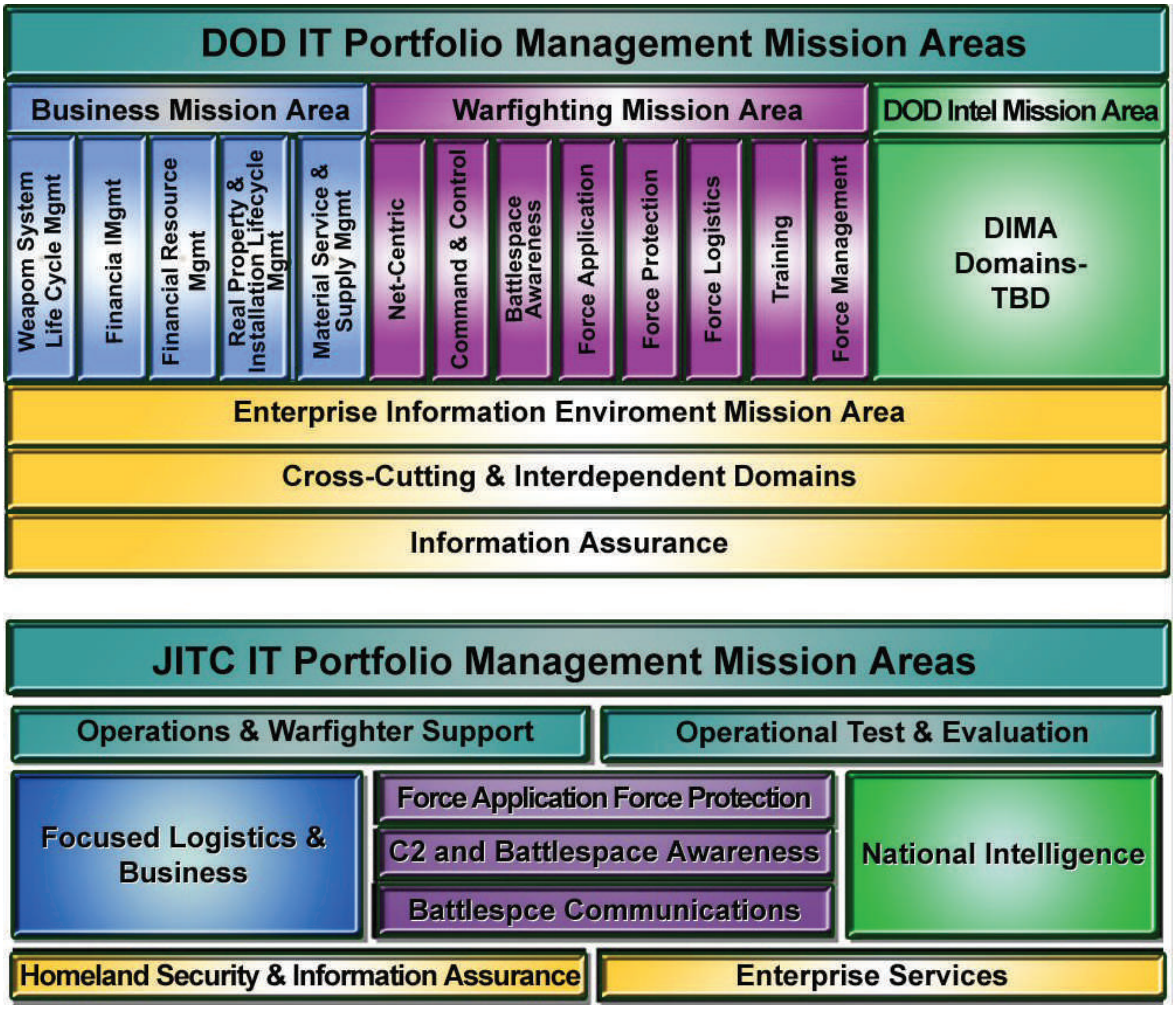


Figure 1. Joint Interoperability Test Command's portfolio relationship to Department of Defense

rigorous T&E required for fielding suitable and effective net-centric warfighting capabilities. JITC's "lifecycle" T&E approach is to leverage the entire requirements and acquisition lifecycle depicted at the top of *Figure 2*.

While the Service OTAs are heavily involved in testing major weapon systems as well as IT and NSS, JITC *specializes* in IT and NSS testing. A significant component of JITC's IT and NSS charter is interoperability testing through assessment of compliance with the elements of Net-Ready Key Performance Parameter (NR-KPP). Building upon experience with the elements of NR-KPP, JITC is uniquely poised to conduct operational T&E of net-centric capabilities. JITC has full-spectrum knowledge and understanding of all doctrine, organization, training, materiel, leadership and education, personnel, and facilities considerations for the net-centric enterprise to include policy and governance, community of interest requirements, and enterprise level architectures focusing on joint capability areas. Net-centric testing methodologies involving service oriented architectures and the elements of NR-KPP help develop test documents that better support the integration of operational realism into testing IS and NSS.

Rigorous development of operationally relevant net-centric test methodologies requires an understanding of what it means to be truly "net-centric." Net-centricity is a concept, not a particular technical implementation. As defined by John G. Grimes, DoD Chief Information Officer, "net-centric means people, processes, and technology working together to enable timely and trusted *access to* information, *sharing of* information, and *collaboration among* those who need it most." (Grimes 2006) Supporting technologies include net-centric standards, Service Oriented Archi-

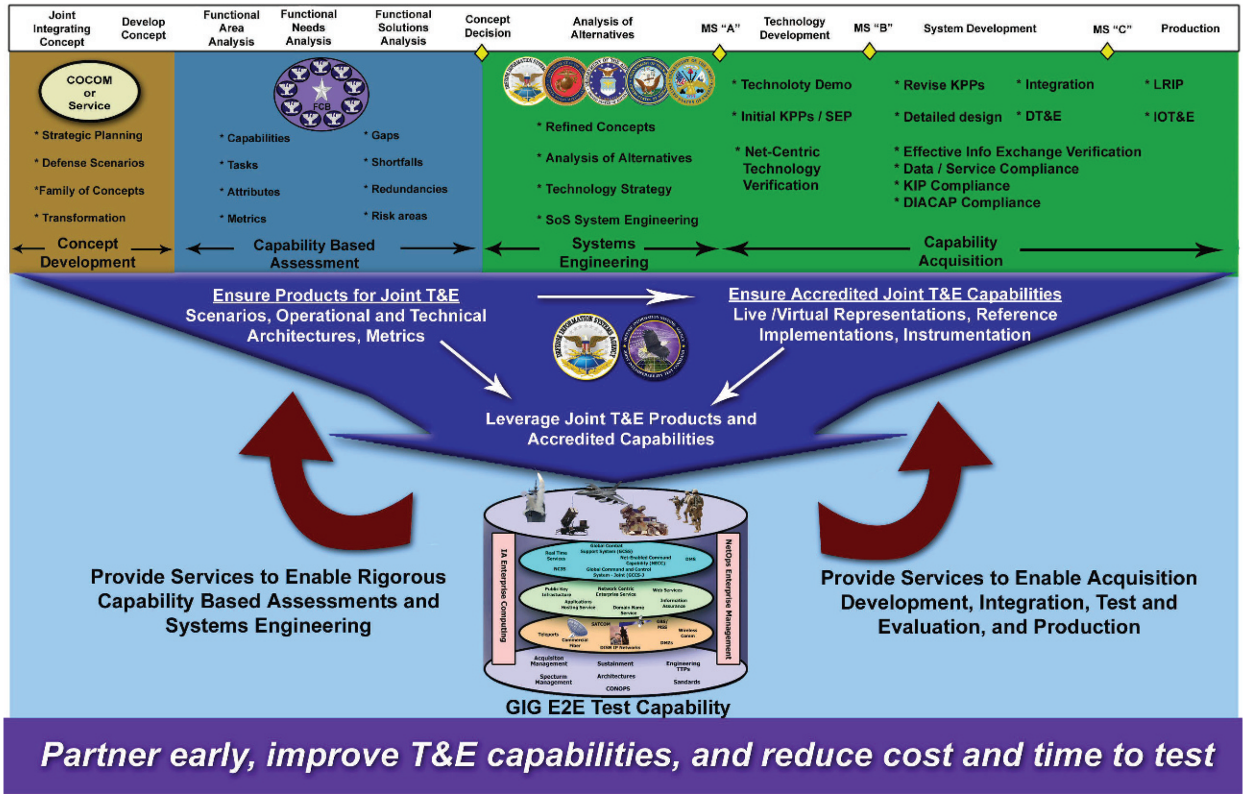


Figure 2. Joint Interoperability Test Command's lifecycle test and evaluation approach

tectures, and collaboration tools, but it is important to make the distinction between the concept of net-centricity and the various architectural implementations that support that concept. From an operational perspective, the implementation behind the concepts is only as important as its overall effect on the ability to complete the operational mission.

Figure 3 outlines a notional NR-KPP testing timeline. While not every IT and NSS program follows the same development path, *Figure 3* provides a general idea of the NR-KPP certification process. It's important to note that the NR-KPP evaluation spans the entire life-cycle, from development of the NR-KPP assessment approach prior to Milestone B to periodic reevaluation based on significant system changes. This is not intended to be viewed as a serial process, but rather a cohesive integrated test approach resulting in net-ready warfighting capabilities that have been evaluated under operationally realistic conditions.

Integrating operational realism into net-centric testing strategies is paramount. As stated by Dr. Hutchison, DISA's T&E Executive, in the Fall 2007 Military Information Technology Journal, "The ultimate objective of an agile T&E strategy is to involve the tester early, to focus on what's important to the Warfighter, and to test as one team, one time and under one set of operationally realistic conditions (Capability Test & Evaluation mantra)." (Hutchison 2007) As stated in the December 22, 2007 memorandum signed by Dr. Charles E. McQueary, DOT&E, and John T. Young, Jr., Under Secretary of Defense for Acquisition, Technology, and Logistics: "To realize the benefits of modeling and simulation, T&E will be conducted in a continuum of live, virtual, and constructive system and operational environments."[1] Dr. McQueary's and. Young's memorandum emphasizes various aspects of T&E in operational environments, the combining of developmental testing with operational testing, and a variety of other T&E policy changes that will be incorporated in the next revision of DoD Instruction 5000.2 and will impact all DoD T&E activities, to include those of the five OTAs.

Operational realism in interoperability certification testing

IT and NSS interoperability is an important component of operational realism. Among many other testing activities, JITC's core mission involves interoperability certification testing of DoD IT and NSS. Paralleling the Service OTAs that provide independent testing for Service acquisition executives and milestone decision authorities to make procurement and fielding decisions

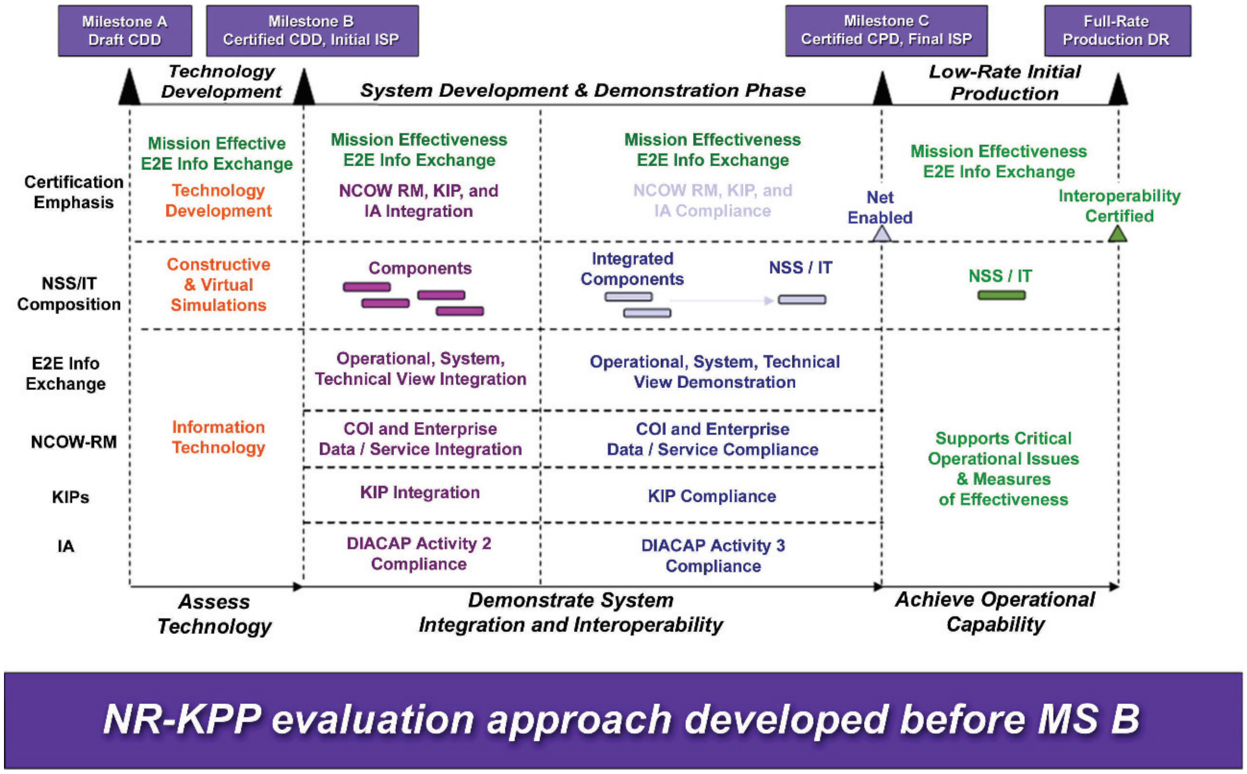


Figure 3. Net-Ready Key Performance Parameter testing timeline

with DOT&E as oversight for major programs, JITC provides independent testing for the DISA component acquisition executive, federal agencies, combatant commands, and the Assistant Secretary of Defense for Networks and Information Integration (ASD/NII) to provide procurement and fielding decisions with DOT&E as oversight for major IT and NSS.

A significant aspect of JITC's IT and NSS testing is interoperability and net-readiness assessments in operationally realistic environments. As DoD migrates to a net-centric enterprise, the operationally realistic environment will rely heavily on incorporation of net-centric concepts such as service oriented architectures, net-centric standards, and collaboration capabilities.

Figure 4 depicts the overall JS J6 interoperability and supportability certification process. Detailed information on the entire process is in the Chairman, Joint Chiefs of Staff Instruction (CJCSI) 6212.01D. This process consists of three primary phases. Program sponsors first submit the requirements/capabilities documents for consideration to the JS requirements/capabilities for an interoperability and supportability certification. Upon receiving an interoperability and supportability certification, JITC tests and evaluates the system.

If JITC issues a Joint Interoperability Test Certification, Joint Staff then considers the system(s) for Joint Staff J6 system validation. The amount of testing required to make an interoperability certification decision is based on several factors, including: the number and complexity of the interfaces; the interoperability requirements; the criticality of the information exchanged; the risks involved with the technology being used; and, perhaps most significantly, the need for an operationally realistic test environment. The PM/sponsor and JITC work closely to establish a strategy for evaluating interoperability requirements in the most efficient and effective manner, and in an operationally realistic environment.

Operational realism in military exercises

Over its 20-year history, JITC has supported a variety of military exercises and coalitions involving various roles and responsibilities. This is a more practical application of supporting the warfighters in their own backyards. Currently, JITC is supporting approximately 15 different exercises and coalitions. The following is a partial list of Fiscal Year 2007 exercises that JITC was involved with:

- **COMBINED ENDEAVOR** — European Command (EUCOM) sponsored event that involves the conduct of interoperability assessments between various North Atlantic Treaty Organization and Euro-Atlantic Partnership Council countries.

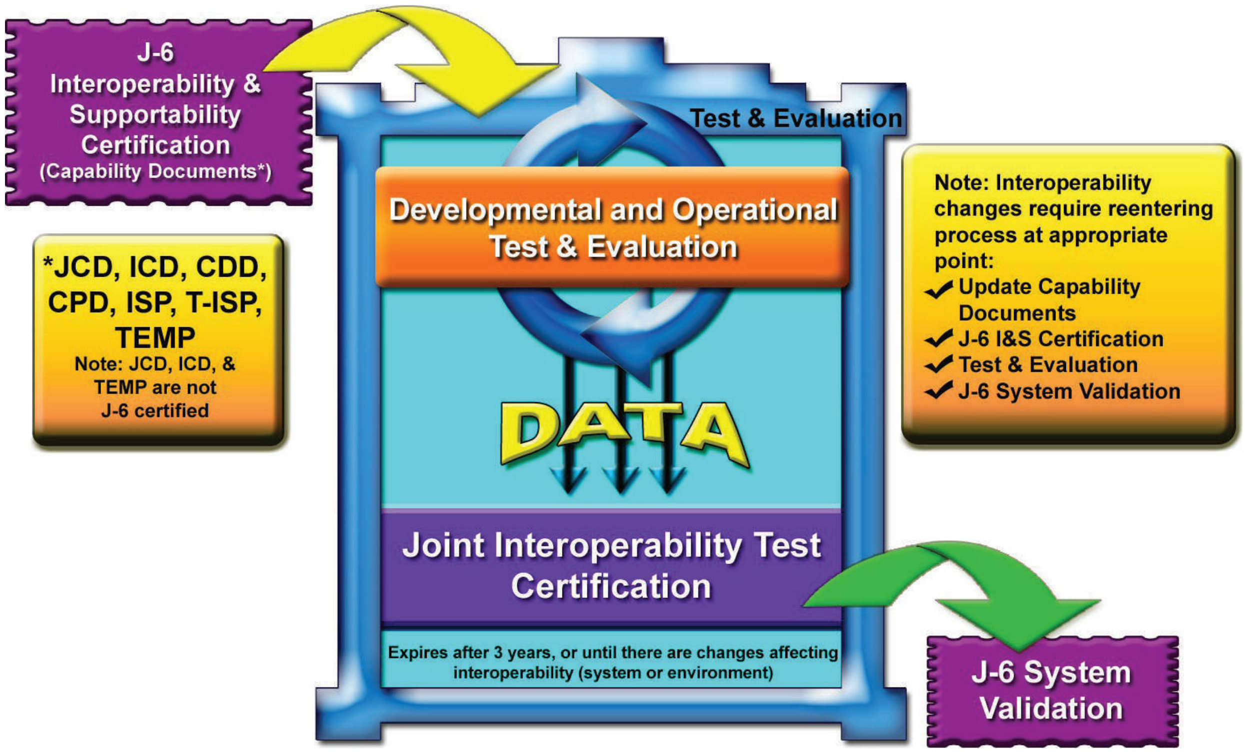


Figure 4. Joint staff Net-Ready Key Performance Parameter test and evaluation process

- **AFRICA ENDEAVOR** — EUCOM sponsored event that involves the conduct of interoperability assessments between the African nations.
- **PACIFIC ENDEAVOR** — Pacific Command sponsored event that involves the conduct of interoperability assessments between various Pacific Rim countries.
- **BALIKATAN** — Annual bilateral U.S./Republic of the Philippines exercise that consists of a staff exercise, humanitarian and civic assistance/civil military operations, and field training exercises.
- **TALISMAN SABER** — Australia/U.S. bilateral exercise merging exercises TANDEM THRUST, KINGFISHER, and CROCODILE.
- **COBRA GOLD** — Thai/U.S. field exercise. This year's exercise is the 26th anniversary of Cobra Gold, and training will focus on a field training exercise, a command post exercise, and civic action projects.
- **AIR FORCE-INTEGRATED COLLABORATIVE ENVIRONMENT** — Support both technical and operational assessments to determine the ability of participating systems to digitally support Joint Close Air Support.

We have a long standing history of working closely with combatant commander sponsored exercises and various contingencies. JITC supports various exercises by providing technical consultation at planning conferences and through review of documents for interoperability issues. We're able to take advantage of exercise involvement and assess and/or evaluate programs and systems in final fielding environments. JITC testers and evaluators assess operational environments and integrate information from the warfighters to our testing efforts to add credibility to our test planning and test executions. JITC provides "Warfighter Hotline Support" by answering operational and technical questions 24 hours a day, 7 days a week. If needed, we deploy task organized teams that are custom tailored for combatant commands and joint operational requirements. Subject matter experts provide interoperability and information assurance throughout the planning and execution phases of each exercise.

Our involvement with military exercises and coalitions along with testing experiences in various operational environments for a variety of IT and NSS has led to JITC becoming a leader in building and developing joint mission environments. Using our expertise, we've been a pathfinder and this has improved our capabilities of testing and certifying IT and NSS in joint distributed mission environments. Our federation expertise has included evolving operationally realistic joint T&E mission environments for a number of IT and NSS systems. As part of a continuing effort to integrate operational realism into our testing, we intend to leverage our unique experiences

Figure 5. Operational realism Q.E.D.: The 9,500-foot-high Huachuca Mountains serve as a backdrop for a variety of Joint Interoperability Test Command test shelters and antennas, including: high-gain spiral satellite communication ultra-high frequency (SATCOM UHF) antennas; line-of-sight Army-Navy transportable radio communications-170 antennas; and, a 20-foot parabolic Army-Navy transportable SATCOM-85B antenna

and expertise in establishing joint test environments for support of key transformational DoD programs. In the past four years, JITC has supported approximately 35 joint test environments with more than 50 CONUS/OCONUS sites involving several functional areas/programs. Some of the major functional areas supported include: Joint Tactical Data Link, Integrated Air and Missile Defense, Joint Battle Management C2, and the Distributed Common Ground Station.

As a continuation of adding operational realism to testing of IT and NSS, we continue to support the development and execution of net-centric testing using service oriented architectures and NR-KPP concepts. Additionally, our core interoperability certification mission and involvement with various military exercises coalitions and contingencies present opportunities for assessing and testing in operationally realistic environments.

JITC owns and operates critical test assets that contribute to integrating operational realism into testing programs for IT and NSS at Indian Head, Maryland, the National Capital Region, and Fort Huachuca, Arizona *(Figure 5)*.

Operational realism in the DoD interoperability communications exercise (DICE)

JITC is committed to being a leader in the DoD for federation expertise including evolving operationally realistic joint mission environments for T&E. As such, JITC, with Joint Staff support, develops and executes the largest annual DoD communications exercise, known as the DoD Interoperability Communications Exercise (DICE).

DICE has been conducted annually since 1989 and has certified over five-hundred (500) systems. In 2007, DICE became a triannual event. By moving to a triannual event, DICE better supports the JS Command, Control, Communications, and Computer

Figure 6. During a recent Department of Defense Interoperability Communications Exercise event, participation included communications equipment such as Northern Command's (NORTHCOM's) Mobile Command Platform (MCP). The left photo is an external view of the MCP and the right is a view inside the MCP.

Systems Directorate (J6) Interoperability and Supportability Certification process as well as facilitating acquisition to the field for systems needed to support the warfighter.

DICE is a major JITC initiative supported by the J6 JS, and the Joint Forces Command. It allows the warfighter to acquire, certify, and deploy systems in less than one year which supports DISA's "Surety, Reach, Speed" strategic vision. DICE also reduces the cost of interoperability testing by maximizing network efficiencies and providing a venue where the Services, combatant commanders, and agencies can interact at no additional cost. DICE is growing and evolving in stride with the warfighter and the warfighter's communication needs. JITC conducts DICE in support of joint interoperability experimentation, testing and evaluation of communication systems as well as execution and transformation initiatives.

DICE replicates a geographically dispersed Joint Task Force environment and is the only DoD exercise dedicated solely to interoperability testing and certification. It also provides an excellent forum for testing and experimentation with emerging technologies, allied communications initiatives, regressive testing with legacy equipment, and realistic joint communications training. DICE provides an outstanding venue for integrating operational realism for planners, operators, developers, and testers from the DoD and its partners to collaborate on communications and interoperability solutions—whether the solutions are equipment or procedural in nature.

While the main focus of DICE is the warfighter, this exercise further mitigates risk by creating a dynamic training environment. This environment provides participating warfighters and civil responders opportunities to develop and improve their proficiencies in information systems-related mission essential task lists and tactics, techniques, and procedures, and helps ensure that systems can communicate prior to actual operational fielding need *(Figure 6)*.

Ultimately, DICE reduces the Warfighters' risk of operational failure by aggressively testing new versions of software, equipment, and employment techniques in an operationally representative Joint Task Force communications network while allowing operators and developers to train and collaborate in various operationally representative environments and architectures.

Summary

The future test environment will involve network-centric warfare and an architectural foundation for the GIG to integrate DoD IT and NSS test programs. JITC's vision for the future of net-centric T&E in an FDCE ties the Defense Information Systems Network to the operational community, the capability area proponents, the Service ranges and facilities, the DISA acquisition and development communities, the Service acquisition and development communities, and our industry partners. The integration of these communities, proponents, ranges, and facilities will help integrate warfighters, IT and NSS architectures, scenarios, metrics, live and simulated capabilities, Enterprise IT services, Component IT services, and capability developers into one master test environment that can be used by all DoD elements. A careful integration of this future master test environment into the existing DoD major range and test facility base (MRTFB) infrastructure is vital to the future success of DoD operational testing and the existing MRTFB.

The integration of operational realism into testing IT and NSS has proven to be a formidable challenge for JITC and the DISA TEMC. IT and NSS cannot be tested like standard weapon systems during developmental, operational or combined testing. As a result, DISA's IT and NSS test efforts will continue to expand in military exercises, interoperability certification testing, and DICE. Most significantly, JITC and the DISA TEMC will continue to develop net-centric test capabilities and establish policies and procedures to field net-centric capabilities faster as integrated mission capabilities in operational environments. ❑

COL Ronald C. Stephens is the Joint Interoperability Test Command (JITC) Commander. In addition to commanding a company during Desert Storm, COL Stephens was the signal officer for D7 (Enterprise Integration) at the Defense Information Systems Agency, Washington D.C.; a battalion executive officer for the 54th Signal Battalion, Army Signal Command, Dhahran, Saudi Arabia; a G3 plans officer for 3rd Army, Fort McPherson, Georgia; deputy commander, 2nd Signal Brigade, 5th Signal Command, Mannheim, Germany; a commander, 29th Signal Battalion, Fort Lewis, Washington; an information assurance branch chief, J6, Joint Staff, Pentagon, Washington D.C., and, a combat support arms division chief, Human Resources Command, Alexandria, Virginia. He holds an associate of science degree in computer electronics and a bachelor of science degree in industrial electronics from Eastern Kentucky University; a masters degree in telecommunications from the University of Colorado; and, a masters degree in national resource strategy from the National Defense University. COL Stephens is also a graduate of the Command and General Staff College and the Industrial College of the Armed Forces.

Randon R. Herrin works in JITC's Operational Test and Evaluation Division as the branch chief for Sustaining Base Systems. A former Air Force officer, he has more that 28 years of information technology and national security systems (IT and NSS) research, development, test and evaluation, and acquisition experience. He's a distinguished graduate of the "Joint C4I Systems Technology" master of science (MS) curriculum at the Naval Postgraduate School and a graduate of two other MS curriculums. Previous IT and NSS T&E assignments include the Tactical Air Command Warfare Center, the Air Force Operational Test and Evaluation Center, and the Global Positioning System Joint Program Office. He is a graduate of the USAF Squadron Officer School, the USMC Command and Staff College, and the USAF Command and Staff College, and has published in a wide variety of professional journals.

Danielle Mackenzie works in JITC's Operational Test and Evaluation Division as the Net-Enabled Command Capability (NECC) Capability Test Team lead. A graduate of the Army intern program, she has nine years of experience in both government and industry focusing on the research, development, engineering, test and evaluation, and acquisition of command and control systems. She holds a bachelor of arts degree in mathematics and a master of engineering degree in systems engineering from Stevens Institute of Technology, Hoboken, New Jersey. Previous assignments at Fort Monmouth's Communications Electronics Research, Development, and Engineering Center include project lead for the Network Enabled Battle Command Science and Technology Objective, and information management integrated product team lead for the Objective Force Warrior program. A former member of the Army Acquisition Corps, Ms. Mackenzie received Level III DAWIA certification in systems engineering in 2005, and is currently qualified for Level III DAWIA certification in test & evaluation.

Daniel (Dan) W. Knodle works in JITC's Test Engineering Branch as a computer scientist. A former Air Force officer, DoD contractor and commercial telecommunications manager, he has more than 29 years of information technology and command, control, communications computers and intelligence test and evaluation, engineering and operational experience. He's a graduate of the Teleprocessing Science Master of Science program from the University of Southern Mississippi. Previous assignments include the 485th Engineering Group (communication system design), the Air Force Tactical Communications Test Team, the Air Force Operational Test and Evaluation Center, DISA Europe and, as a senior manager at Sprint Internet Operations. During his tenure at JITC, he has assisted with the developments of the Joint Distributed Evaluation Plant and the Net Ready Key Performance Parameter test concepts.

Endnotes

[1]McQueary, C. E., Young J. T. Jr. 2007. "T&E Policy Revisions." Memo signed December 22, 2007 by Dr. Charles E. McQueary, Director, Operational T&E, and John T. Young, Jr., Under Secretary of Defense for Acquisition, Technology, and Logistics.

References

Grimes, J. D. 2006. "From the DoD CIO: The Net-Centric Information Enterprise." *CrossTalk: The Journal of Defense Software Engineering*, Volume 8, No. 8, July 2006, p. 9.

Hutchison, S. 2007. "Test and Evaluation Transformation—DISA Organization Was Rebuilt to be the Premier Tester and Evaluator of Net-Centric Capabilities in the DoD." *Military Information Technology Journal*, Volume 19, No. 7, Fall 2007, p. 4.

ITEA Journal 2008; 29: 157–159

Test and Evaluation Policies and Practices: A New Emphasis

Darlene Mosser-Kerner
Office of the Deputy Director,
Developmental Test and Evaluation,
Office of the Secretary of Defense, Washington, D.C.

The Department of Defense (DoD) recently issued new and revised test and evaluation (T&E) policies that represent a shift in emphasis toward the evaluation side of T&E and promote a continued emphasis on integrated testing. The revised policies focus on using T&E throughout the system life cycle in a seamless continuum. This revision of T&E policies represents one of many actions the Department is taking to revitalize T&E and to ensure that the T&E is timely, effective, and efficient.

Key words: 231 report; capabilities and limitations; Department of Defense; evaluation and reporting; system life cycle; test and evaluation.

In December 2007, the Under Secretary of Defense for Acquisition, Technology, and Logistics, and the Director of Operational Test and Evaluation jointly issued a memo to introduce new and revised policies for test and evaluation (T&E) of Department of Defense (DoD) programs. The memo affirms, "The fundamental purpose of test and evaluation is to provide knowledge to assist in managing the risks involved in developing, producing, operating, and sustaining systems and capabilities" (OSD 2007).

The revised policy responds to a 2007 review of DoD T&E and its applicability to emerging acquisition approaches. The Director of Operational Test and Evaluation and the Office of the Deputy Director, Developmental Test and Evaluation conducted the review and delivered the resulting report to Congress in July 2007 in compliance with Section 231 of the John Warner National Defense Authorization Act for Fiscal Year 2007, Public Law 109-364. The report, known as the "231 report," is the latest in a series of reviews and studies of DoD T&E that signaled the shift in DoD T&E policy.

The December 2007 policy and the findings from the 231 report can be grouped into four broad themes:

1. Emphasis on evaluation
2. Focus on capabilities and limitations
3. Integrated and seamless T&E
4. Developmental T&E reporting.

Emphasis on evaluation

In recent years the Department has focused on the testing side of T&E, creating an imbalance toward measuring technical parameters, but the new policy assumes the "knowledge to assist in managing risk" (OSD 2007) comes mainly from the evaluation step of the T&E process. Testing is perhaps the most visible part of T&E and consumes most of the resources; however, people conduct testing because someone in a decision-making role needs credible knowledge of how a system works or does not work to make an informed decision.

The effectiveness of the evaluation depends on decisions about what to test and the applicability of the data from testing. If program managers assume that they cannot test all aspects of a system or capability, then the questions become twofold: What do they test, and how much testing is enough? The answer at a strategic level is to test enough, and in specific areas, to mitigate the key risks for the system or capability being developed.

Who defines the key risks? The program manager for one, and all the decision makers in the program management chain, which includes the milestone decision authorities, and even Congress, which authorizes and appropriates funding for the program. Other "decision makers" who need T&E-generated knowledge to manage risk include systems engineers who need knowledge of system and subsystem performance to assist in maturing the technologies and design. The manufacturing decision makers need knowledge of system performance to mature and control the manufacturing processes. The operator uses the knowledge of system capabilities and limitations to mitigate the inherent risks in operating and employing the equipment. The maintainers need knowledge to inspect, service, and repair systems.

How much testing is enough? The obvious answer is "it depends." It depends upon how much risk the decision maker is willing to accept. If the decision maker is not willing to accept much risk, then the amount of required testing will increase. If the decision maker is willing to accept more risk, then the amount of required testing will decrease. In general, the expectation is that you will never have enough time or money to test to achieve absolute certainty; there will always be an element of uncertainty or residual risk.

By shifting the emphasis to evaluation and the knowledge generated through T&E, the customers are empowered, the decision makers are empowered to help testers determine what to test and how much testing is necessary. In some respects, this shift in emphasis will increase the importance of communication between the T&E community and the various decision makers.

Focus on capabilities and limitations

The second theme of the new policy on T&E is the focus on determining or assessing capabilities and limitations of the system(s). One of the purposes of the Defense acquisition system is to "acquire quality products that satisfy user needs with measurable improvements to mission capability" (DoDI 5000.1). One of the new policies is that "Evaluations shall include a comparison with current mission capabilities…, so that measurable improvements can be determined" (OSD 2007). This policy statement was driven by the use of relative performance in system requirements and during milestone reviews. For example, "System X shall be twice as good as Legacy Y," or "I know it doesn't meet the users' requirements, but it's better than what they currently have." The new policy recognizes the use and utility of comparative assessments and provides some appropriate guidance for the acquisition community.

In addition, the policy revision states that these improvements to mission capability "should be reported in terms of operational significance to the user" (OSD 2007). The focus on determining capabilities and limitations is not a mandate or a blank check to test everything in a search for capability or potential limitations. The amount of testing is still bounded by the risk tolerance of the various decision makers, especially the ones paying for the program. On the other hand, the focus on capabilities and limitations also means that T&E is more than just specification compliance. T&E does measure progress in system and capability development, and one of the ways to do that is by measuring progress against the specification; however, T&E should also develop an understanding of basic capabilities and limitations, so the systems engineer and the program manager can both assess the relative technical maturity of the system. The understanding of capabilities and limitations informs discussions of current mission performance and potential issuance of new capability requirements. The results of T&E need to be linked in some mission context and stated in terms of relevance to the user. Our purpose in defense acquisition is to provide capability to the user, so it makes sense that evaluators should be able to tie the results of T&E to capability for the user.

The focus on capabilities and limitations generated considerable discussion during the drafting and coordination of both the 231 report and the policy memorandum. The concern was specifically about the requirement to compare the new system capabilities with current mission capabilities and whether that requirement became an "unfunded mandate" to retest legacy systems. Such a mandate was not the intent, and the policy memo specifically included a provision that if the "evaluation is considered cost prohibitive the Service Component shall propose an alternative evaluation strategy" (OSD 2007). The new policy let the program managers know that if they wanted to use the rationale that the new system was better than the old system, they would need to provide a basis for that evaluation.

Integrated and seamless T&E

The third theme of the new policy is integrated and seamless T&E, meaning T&E conducted in a continuum throughout the system life cycle. The traditional focus of T&E has been during the system development phase and early production. One focus of the new policy is getting the T&E community involved earlier in the system life cycle, when requirements and concepts are first developed. The goals of this early involvement are to establish better requirements that are more fully understood, and the "early identification of technical, operational, and system deficiencies, so that appropriate and timely corrective actions can be developed prior to fielding the system" (OSD 2007).

In addition, "Developmental and operational test activities shall be integrated and seamless throughout the system life cycle" (OSD 2007). The focus on integrated developmental and operational testing is consistent with prior policy; however, now the role of T&E in the system life cycle is being expanded, so all testing should be as seamless as possible, with minimal or no stops and starts for different types of testing. This seamless T&E will require continued emphasis on the use of live, virtual, and constructive modeling and simulation (M&S), or as the policy memo puts it, "T&E will be conducted in a continuum of live, virtual, and constructive system and operational environments" (OSD 2007). Another focus in making T&E integrated and more efficient is the policy that "evaluations shall take into account all available and relevant data from contractor and government sources" (OSD 2007). This may not be as easy as it sounds, given the typical issues with data authentication, archival, and retrieval, in addition to potential proprietary issues; however, it is essential if programs are to realize the promise of integrated testing in increasing the efficiency of the test programs and effectively shortening the time required to acquire new or improved capabilities for the warfighter.

T&E also should consider the deployment and sustainment period in the system life cycle. The new policy states in part, "As technology, software, and threats change, follow-on T&E should be used to assess current mission performance and inform operational users' during the development of new capability requirements" (OSD 2007). Since the majority of the life of a system is spent in operations and sustainment, T&E will have a role to play in providing system modifications, and assessments for end-of-life and disposal decisions. Some of the testing in this phase of the system life cycle is already being performed by operational units, so the new policy should not change that testing; however, it should cause a reassessment of all T&E throughout the system life cycle to ensure the full benefits of T&E are being realized in an efficient and effective manner.

Developmental T&E reporting

The fourth theme of the T&E policy memorandum is the renewed emphasis on evaluation and reporting by the developmental evaluators. This is one of the key aspects in revitalizing T&E, especially the government's Developmental Test & Evaluation role and mission. The operational evaluators already fulfill their statutory roles in providing assessments of operational effectiveness and suitability. In a similar manner, the developmental evaluators formerly provided assessments of system maturity and technical progress at each milestone decision review, but over the years that assessment has been lost. The new policy provides for a developmental evaluation of system "strengths and weaknesses in meeting the warfighters' documented needs" (OSD 2007). The program manager is tasked with providing the results of this evaluation at the Milestone B and C reviews, so the new policy just adds a new element to the program manager's presentation. It does not create any additional independent reporting requirement.

Summary

The 231 report and associated policy memorandum are not the last word in revitalizing T&E in DoD. The Department is taking ongoing actions, in areas such as system of systems T&E for example, to revitalize the role T&E plays in the acquisition of new and modified systems and capabilities. The revised policy does provide a shift in emphasis on the role of T&E, and especially evaluations. The 231 report and policy memo also make adjustments in T&E policy to accommodate both existing and emerging acquisition approaches. The revised policy is another step toward achieving the end goal of efficient and effective testing to deliver timely knowledge to all stakeholders to help manage the risks in developing, producing, operating, and sustaining systems and capabilities for the Department of Defense. ❑

Darlene Mosser-Kerner is the senior policy lead within the Office of the Secretary of Defense for Developmental Test & Evaluation matters. She has more than 20 years of test and evaluation experience at NASA and the Department of Defense, including chief engineer on the NASA Systems Research Aircraft, assistant senior technical advisor at Edwards Air Force Base, and assistant program manager for T&E with the Naval Air Systems Command. She is the recipient of an Aviation Week & Technology Laurels Award and was the 2007 Department of Defense Civilian Tester of the Year. She has a bachelor's degree in electrical engineering and a master's degree in technology management. E-mail: Darlene.Mosser-Kerner@osd.mil.

References

Department of Defense. Department of Defense Directive 5000.1: The Defense Acquisition System. U. S. Department of Defense, 12 May 2003.

Department of Defense. Department of Defense Report to Congress on Policies and Practices for Test and Evaluation, July 2007.

Office of the Secretary of Defense Memorandum. Test and Evaluation Policy Revisions, 22 December 2007.

ITEA Journal 2008; 29: 160–166

Joint Mission Environment Test Capability (JMETC)

Richard Lockhart
Test Resource Management Center, Arlington, Virginia
Chip Ferguson
JMETC, Arlington, Virginia

During the past six years, utilizing the concept of Joint Operations has become standard for successful U.S. military combat. The need for Joint interdependency and Joint systems interoperability during combat operations in Afghanistan and Iraq is irrefutable. Consequently, Joint lessons learned are rapidly being incorporated in combat operations and Service/Joint training programs.

However, Joint lessons have proven to be elusive and slow to take hold in the Department of Defense (DoD) acquisition and testing communities. Despite the popularity of the common catch phrase "Born Joint," many systems are fielded without a rigorous and credible test of their capabilities in a Joint operating environment. Performance of Joint systems must be evaluated in a Joint environment, and the only way to do that effectively, efficiently, and early in acquisition is to link distributed high-fidelity test facilities creating a realistic Joint environment. However, the proliferation of unique, noninteroperable, and expensive testing infrastructures remains widespread within the DoD.

To effectively and efficiently test systems in a Joint environment, the DoD needed an enterprise-level LVC distributed test capability using a common infrastructure. The Joint Mission Environment Test Capability (JMETC) program was initiated in FY 07 to meet that need. This article provides a brief description of the JMETC program and the capabilities it is providing, as well as its recent accomplishments.

Key words: Integral Fire 07; InterTEC Spiral 2, Build 1; JMETC program; JMETC Users Group; JMETC Virtual Private Network (VPN); LVC; Object Models; reuse repository; warfighter.

The proliferation of unique, noninteroperable, and expensive testing infrastructures remains widespread within the DoD. The current DoD test community contains a number of stand-alone test resources that:

- Lack a standard capability to collaborate and exchange data between facilities, which can result in duplicate efforts among similar programs
- Contain unique software that must be integrated for each test activity, adding to test preparation time and expense
- Use data definitions that are often unique and noninteroperable, complicating integration into Joint systems and system capabilities
- Require long lead times to establish security agreements needed to link them together on a network for each test (most such agreements are generally in effect for one test event).

The lack of universal tools and a common infrastructure also impacts test planning, coordination, and execution. Each program spends time and money establishing or re-establishing a live, virtual, and constructive (LVC) test environment and a network configuration for each test.

The JMETC program is a new program, formally established in October 2006. In actuality, the need for JMETC was established years earlier as reviewed in this section.

In March 2004, the Strategic Planning Guidance (SPG) on Joint Testing in Force Transformation stated that developing and fielding Joint force capabilities requires adequate, realistic test and evaluation (T&E) in a Joint operational context. The SPG directed development of a Roadmap to define changes to ensure that T&E is conducted in a Joint environment and facilitates the fielding of Joint capabilities.

The DoD *Testing in a Joint Environment Roadmap*, approved by the Deputy Secretary of Defense, November 2004, identified actions to implement testing in a Joint environment. One action was to establish a corporate capability for Joint distributed testing.

The JMETC Program originated with the 2005 Department direction for the stand-up of the JMETC Program in FY 07 under the Under Secretary of Defense for Acquisition, Technology, and Logistics (AT&L). AT&L assigned responsibility for execution to the Director, Test Resource Management Center (TRMC). The JMETC Program Office was stood up in October 2006 and in the past year has made great strides in establishing and executing the program.

Program description

The JMETC mission is to provide a DoD corporate approach for linking distributed facilities on a persistent network, thus enabling customers to develop and test warfighting capabilities in a realistic Joint context. Customers include *Program Managers* (Acquisition Program Managers, Portfolio Managers, Advanced Concept Technology Demonstration Managers, etc.), *Test Agents* (Organizations designated by Program Managers to lead their event test planning and execution, e.g., White Sands Missile Range and Edwards AFB) and *Resource Owners* (capabilities owned across the department and in industry used to test warfighting capabilities).

In addition to DoD customers, industry is a key participant in a successful DoD "corporate approach" to linking distributed facilities. Industry owns many of the resources that must be linked together to create a Joint environment. There are two ways in which industry can participate in the JMETC infrastructure. One is to perform tests in support of a government contract. The second is for industry to fund the addition of their facilities onto the JMETC Virtual Private Network (VPN). In addition, industry is welcome to participate in the JMETC users group or the Testing and Training Enabling Architecture (TENA) Architecture Management Team (AMT).

The JMETC program relies heavily on the collaboration of the Services, Joint Forces Command (JFCOM) and agencies to build an infrastructure relevant to current and future requirements. In order to facilitate and formalize this exchange, the program office instituted the JMETC Users Group. The group is composed of representatives from acquisition program offices, technical experts, and representatives from ranges that are potential users of the JMETC infrastructure and products. Its focus is on technical requirements and solutions. It makes recommendations to improve JMETC processes and procedures and determines the TENA priorities for the Test Community prior to TENA AMT meetings.

Just as the JMETC Users Group provides a collaborative environment for JMETC, the TENA AMT meetings provide a technical forum for open dialogue between users and TENA developers. The group identifies issues, vets concerns, debates solutions, and agrees on a way forward. Currently, over 27 companies are members of the TENA AMT. TENA middleware and object models are freely available at www.tena-sda.org.

JMETC is used whenever you need to link resources together to conduct a distributed test event and supports events, such as Developmental Testing, Operational Testing, Interoperability Certification, Net-Ready Key Performance Parameter (KPP) compliance testing, and Joint Mission Capability Portfolio testing. JMETC's ability to support the full spectrum of testing makes JMETC a true DoD corporate solution for distributed testing.

The JMETC program consists of both products and services. Products include a core reusable and easily reconfigurable infrastructure that also provides compatibility between test and training. JMETC services include a customer support team to assist in the use of JMETC products and to assist in the planning and execution of distributed testing.

Specifically, the JMETC infrastructure consists of the following six products that form the bedrock of JMETC capability. Each is being developed and matured through active coordination with the Services, JFCOM, test programs, and other T&E agencies. Each of these products will be refined as the capability and user requirements within the Joint infrastructure mature.

1. Persistent Connectivity. The JMETC program has established and is maintaining a dedicated virtual private network. The current sites will be expanded based on customer requirements and potential for reuse. This is a readily available, corporate integrated network centrally managed and configured to provide long haul information transfer services. In technical terms, it is not a true VPN, yet it is "VPN like," operating over a broad mesh network of active sites on the Secret Defense Research and Engineering Network (SDREN). After an analysis of alternatives, JMETC selected SDREN due to existing sites at test facilities, existing security agreements and procedures, available encryption devices needed for secure connectivity, and cost considerations. Key to this capability are the persistent security agreements that will reduce event preparation time and effort by virtue of the fact that they are persistent.

Once an initial entry on to the net is achieved and the location has the authority to connect, that location will then have a persistent network connection on the JMETC VPN. There will be no need for multiple security agreements with every other location on the network or the need to redo the agreements for each subsequent event, since the network connections will not be event or time specific. This persistency lowers test preparation time and cost of future test events.

2. Middleware. The JMETC program provides a data exchange software used by range systems, laboratories, and simulations to send and receive data. It allows for common functionality (data distribution, filtering, etc.) to exchange data or information between systems on the JMETC VPN. Many JMETC locations already use the Test and Training Enabling Architecture (TENA). In an analysis of alternatives study of LVC integration capabilities, JMETC selected TENA for the infrastructure's data exchange software since it was already being used, could satisfy demanding real-time performance requirements, was easy to integrate, and had a well-established improvement process to incorporate new technologies (TENA 6.0 will be available in FY 08). Furthermore, gateway devices have been developed that enable TENA to connect to other data exchange protocols such as distributed interactive simulation (DIS) and High-Level Architecture (HLA). TENA is also used by the Joint National Training Capability (JNTC) thus providing a solution common to both the test and training communities.

3. Standard Interface Definitions and Software Algorithms. This is a collection of Object Models that provide a common language used in data exchanges between the systems integrated together in a distributed event. Most significantly, Object Models provide the standard data definitions and interfaces of the numerous configured systems (such as radars, tracking systems, GPS instrumentation, hardware-in-the-loop laboratories, display systems, analysis terminals, etc.), specifying what data can be generated by and collected from each system. Object Models not only define the data elements being exchanged, but also provide common algorithms (such as, coordinate conversions, unit conversions, dead-reckoning, etc.), aiding interoperability, and data analysis. In other words, they serve as the object-oriented interface to a service available over the network or system being integrated. Such an interface is said to be the *Object Model* of the represented service or system. JMETC uses the TENA Object Models to provide the standard data definitions set.

4. Distributed Test Support Tools. The JMETC program will use a collection of common software applications that help test managers plan, prepare, set up, check out, monitor, and analyze the distributed LVC integration. The JMETC Users Group will use a "best-of-breed" process to identify the existing tools across the Services and Agencies. This suite of universal test planning, set-up, monitoring, and control tools will provide easy, reliable, and uniform test tools for integration of test assets. The standard set of tools enables reuse from event to event, diminishing a program's need to rely on a unique toolset. This improves coordination and reduces event planning and preparation time—not only saving test programs time and money, but also allowing them to capitalize on similar efforts and programs already completed.

5. Data Management Solutions. In the future, the JMETC program will have a suite of data archiving solutions designed to store and transport collected test data from multiple locations. It will provide a rapid and efficient method of retrieving data from distributed test locations for analysis and enhanced evaluation of test results. The Central Test and Evaluation Investment Program (CTEIP), under the TRMC, is conducting an assessment of T&E data management requirements. Leveraging the efforts of the CTEIP activity, JMETC will adopt and advocate the proposed data management solutions. JMETC will also participate in and represent LVC integration interests in the CTEIP data management requirements assessment effort, and will keep JMETC customers informed of the latest status through the JMETC Users Group.

6. Reuse Repository. The repository will provide a Web portal with a variety of services valuable for all JMETC users. Available via the world-wide web, the Reuse Repository will provide customers with access to JMETC information, including the latest middleware, object models, available software tools, documentation, how-to guides, and Web-enabled collaboration services. In addition, the Reuse Repository will also contain information and references to past events support by JMETC, capturing general configuration information, lessons learned, recommended practices, and other relevant LVC integration information. Allowing test programs to leverage the success of similar efforts and events, the Reuse Repository will reduce event planning time by providing customers a "one-stop-shop" for existing and available LVC capabilities. The planners will not have to start from scratch for each event, and operators, testers, and the acquisition community will have access to a tremendous readily accessible library of lessons learned.

Figure 1 is an operational view of JMETC's infrastructure. The systems under test, depicted in

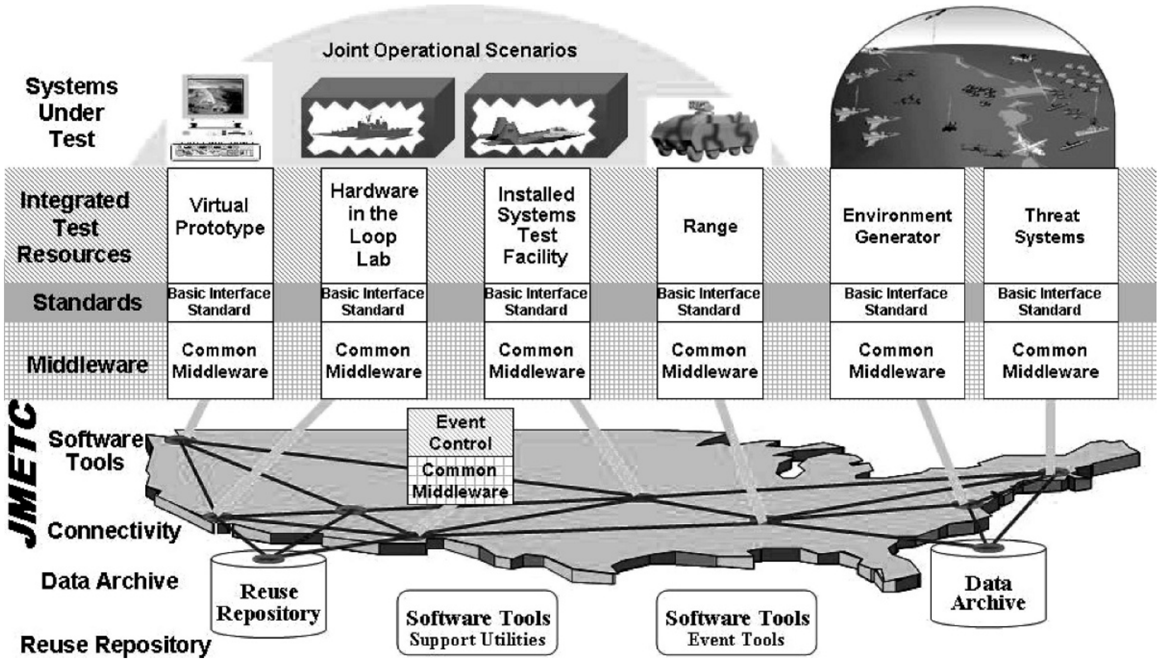


Figure 1. JMETC infrastructure capability

the top row along with the test environment and threat systems, are represented by virtual simulations, hardware-in-the-loop facilities, constructive simulations, and live systems. The solid colored row represents the standard TENA interfaces needed for those LVC representations to communicate. The grid patterned row represents the middleware used to control the flow of data across the network. The lines on the map represent the JMETC VPN which uses the SDREN. The rounded rectangles represent the software tools needed to support and control the distribute test. The cylinders represent the JMETC reuse repository which will contain items relevant to distributed testing, such as information for test planning, available integration software and tools, test facilities descriptions, and past event lessons learned, among others.

JMETC is not developing a new or unique DoD network. It is leveraging existing connectivity and proven network tools to provide a common infrastructure linking test facilities and laboratories to provide a Joint distributed environment for the system under test. Reuse is an important tenet for establishing the JMETC infrastructure. Rather than having each acquisition program establish special network agreements with each site, develop specific data converters, and create unique integration tools, JMETC provides a foundation of networking and supporting infrastructure that future customers can reuse for their distributed test events. For example, the security agreement established with the F/A-18 laboratory at China Lake for the JMETC VPN can be (and has been) reused with other customers, enabling future customers to connect to the F/A-18 laboratory without several months of effort.

In addition to the six products, JMETC also provides a customer support team to assist in using JMETC products as well as expertise in distributed LVC testing. JMETC can provide a dedicated technical representative for each customer to assist with infrastructure requirements definition in addition to planning, preparing, and executing the distributed test event.

Establishing the JMETC VPN

In May 2007, the dedicated JMETC VPN was established on the SDREN. The JMETC program worked closely with the High Performance Computing Modernization Program Office (HPCMPO) on connectivity and security issues to create the JMETC VPN in time to support its first distributed test event, Integral Fire 07, in July 2007. Establishing and verifying the JMETC VPN is a significant step in providing the persistent LVC environment needed to conduct distributed Joint testing. The JMETC VPN provides readily available, persistent connectivity using standing security agreements and common tools to integrate system representations for distributed testing.

In support of the Integral Fire 07 test event, five sites (White Sands Missile Range, Eglin, Redstone Technical Test Center, China Lake, and Patuxent River) were established on the JMETC VPN with standing security agreements. In addition, JMETC used the Network Aggregator Router at Patuxent River, originally spon-

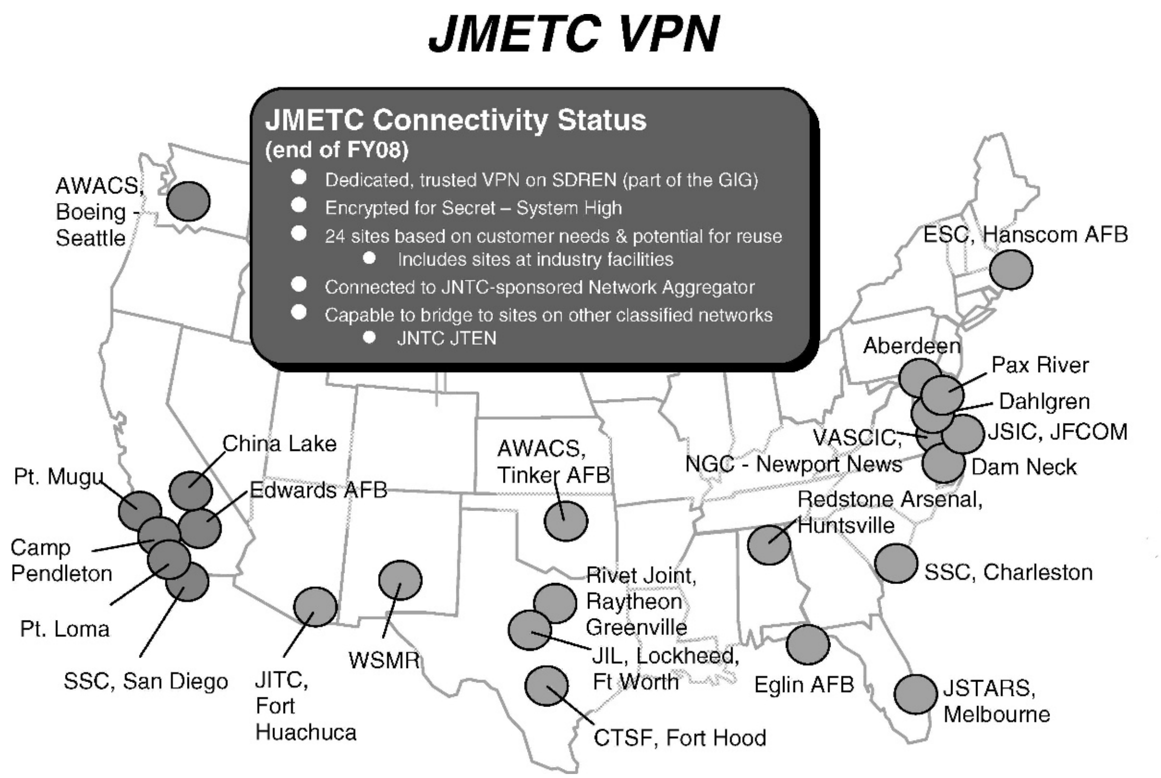


Figure 2. JMETC VPN sites FY07-08

sored by JNTC, to link an additional ten sites, leveraging existing connection agreements to satisfy Integral Fire 07 requirements. Most significantly, this Network Aggregator Router enables JMETC to connect to many other DoD networks, including the Joint Training and Experimentation Network (JTEN) used by JNTC, providing connectivity to all the joint training locations. An additional two sites (Pt. Mugu and Ft. Huachuca) were added to the JMETC VPN to support the second test event in September 2007 called Interoperability Test and Evaluation Capability (InterTEC) Spiral 2 Build 2. (A more detailed description of these tests is provided later in this article.) By the end of FY 08, JMETC expects to create a total of 24 sites in response to customer requirements on the JMETC VPN (see *Figure 2*). This infrastructure will remain available for reuse by future programs for their distributed test events.

Integral Fire 07 - First distributed test event supported by JMETC

Integral Fire 07, completed in August 2007, represented the inaugural use of the JMETC VPN to formally support a distributed test. Integral Fire 07 was a distributed test event sponsored by the Air Force Integrated Collaborative Environment (AF-ICE) and involved all military Services and JFCOM. The customers were JFCOM's Joint Systems Integration Command (JSIC), the Joint Test and Evaluation Methodology (JTEM) Joint Test and Evaluation (JT&E) project, and the Warplan-Warfighter Forwarder (WWF) initiative, sponsored by the U.S. Air Force Command and Control Intelligence, Surveillance, and Reconnaissance Battlelab.

For Integral Fire 07, the JMETC program created a single infrastructure that served all three distinct customers with different requirements, as described below, who were able to test independently in the same time frame, thereby making multiple use of the same infrastructure.

- JFCOM's JSIC conducted an assessment of Joint Close Air Support (CAS) to evaluate the capability of Joint Terminal Air Controller equipment to support digital operations in a distributed Joint environment. Specific emphasis was on the immediate request process for CAS, with the goal to identify gaps, shortfalls, and overlaps with current systems.
- The JTEM JT&E project led a test activity exercising their methods and processes while also providing insight to the Army's Non-Line of Sight/Precision Attack Missile (NLOS/PAM) emerging weapons concepts.
- The USAF WWF initiative tested Non-Line of Sight/Network Enabled Weapon (NLOS/

NEW) Command and Control concepts leveraging the Joint Air-to-Surface Standoff Missile-Extended Range weapons system. Specifically, WWF assessed machine-to-machine data transferred from the Air Operations Center to an airborne platform and then directed to NEW.

The JMETC program was the infrastructure lead for Integral Fire 07. As such, JMETC was responsible for connecting not only the five JMETC sites, but also connecting two other separate enclaves with the JMETC enclave using the JFCOM sponsored JNTC aggregation router. TENA was successfully used to exchange all simulation and instrumentation data between sites. Specifically within their laboratories, nine sites used DIS Protocol Data Units. At each of these local DIS sites, the data was converted to TENA using a DIS-to-TENA Gateway device prior to being sent to another site, mitigating the configuration challenges of using DIS over wide-area networks.

JMETC conducted systems integration, site surveys, and multiple dry runs. All site preparations were completed in time for the three test customers to achieve all test objectives during execution of the two-week event. JMETC also achieved its own four main objectives:

1. Stand up the JMETC VPN
2. Successfully use the Aggregation Router to link three enclaves
3. Support three customers conducting tests using the same network in the same time frame
4. Record lessons learned to improve support in future events.

JMETC's success in providing the infrastructure for Integral Fire 07 is a significant stepping stone for other Service and Joint programs to leverage a core reusable and easily reconfigurable infrastructure.

InterTEC Spiral 2, Build 1 – Second distributed test event supported by JMETC

InterTEC Spiral 2, Build 1, completed in October 2007, represented the second use of the JMETC VPN to formally support a distributed test.

Interoperability T&E Capability (InterTEC) is an OSD-sponsored, U.S. Navy-led project under the Central T&E Investment Program (CTEIP). The purpose of the InterTEC project is to develop an accredited test capability to conduct joint interoperability certification and joint mission thread testing. Spiral 2, Build 1 objectives were developing and assessing tools to test joint threads and assessing the C2 messages sent from sensors to shooters through command and control systems (GCCS-J, GCCS-M, GCCS-A, and TBMCS).

JMETC program responsibilities included taking the overall lead for creating the infrastructure; integrating six sites (five sites on the JMETC VPN and one through the Aggregation Router); conducting systems integration, site surveys, and dry runs in preparation for the event; and overseeing operation of the network and dataflow among all sites during the event. In addition, just as in Integral Fire, TENA was used as the integrating solution to link each site.

The JMETC program successfully supported execution of InterTEC Spiral 2 Build 1 along with completing the following significant accomplishments:

- Established three new sites on the JMETC VPN within 90 days
- Demonstrated reuse (three sites from Integral Fire 07 test)
- Successfully use the Aggregation Router for the second time

JMETC's successful support for this second distributed test event validated the reusability of the JMETC infrastructure, emphasizing the efficiencies of cost and schedule provided by using the JMETC program.

Benefits

The warfighter is the ultimate beneficiary of the JMETC capability. Using the JMETC infrastructure, Program Managers can more effectively and efficiently link distributed test facilities to create a realistic joint environment for testing weapon systems in the environment they are intended to operate. This means they can conduct a more rigorous test of weapons systems and the capabilities of those systems in the environment they are intended to operate, thus improving system interoperability during joint combat operations. Also, since JMETC will be interoperable with the JNTC, program managers will be able to easily link to JNTC sites in order to conduct joint test and training events—a significant benefit to both the test and training communities.

Program managers will save both cost and time when conducting distributed test events using the JMETC VPN sites, the TENA common middleware, standard test planning tools and reuse repository. Instead of creating their own infrastructure and security agreements for each test event, program managers can use JMETC's readily available, corporate integrated network with persistent security agreements. The standing security agreements for the JMETC VPN sites alone will save significant event preparation time. Using the TENA common middleware will make it easier to integrate distributed systems and avoid compatibility issues. TENA will also negate the need to develop unique software solutions so systems can exchange data. Also, TENA will cost the program manager nothing; it is free to download by any user, government or industry.

Another benefit that will save time and cost is the standard test planning tools. Using common tools enables reuse from event to event; this diminishes a program's need to rely on a unique toolset and will improve coordination. In addition, this reduces event planning and set up time. The reuse repository will also save cost and time by allowing programs to leverage the success of similar efforts and events as well serve as a one-stop-shop for LVC capabilities. A significant benefit to the test community is the gateways that JMETC is developing that allow users to connect TENA sites to other legacy data exchange software solutions such as DIS and HLA. This will allow users to connect to these sites through the JMETC VPN without having to modify them.

Users will also find the technical expertise provided by JMETC extremely beneficial. JMETC engineers will assist not only in using JMETC products but also in LVC test event planning and execution. These types of distributed events can be very challenging and JMETC engineers have significant distributed testing experience. Also, unique to JMETC is that customers can help guide the program in the direction that meets their needs and requirements through participation in the users group and the TENA AMT. Collaboration with the Services and Industry is a key element to JMETC implementation and expansion.

In the one short year that JMETC has existed, the Department is already recognizing its benefits. These benefits will significantly increase as the JMETC infrastructure expands and need for joint systems interoperability testing increases.

Conclusion

JMETC is no longer a proposed capability, it is an established capability. The JMETC corporate approach for linking distributed facilities enables T&E and acquisition community customers to evaluate new and legacy systems in a Joint context. Increasing customer requirements for the JMETC capability is driving rapid expansion and demand for the JMETC VPN resulting in increased benefits to each customer.

The JMETC Program Office is aggressively working with multiple programs to determine how requirements can be met. The collaborative support from the Services and JFCOM, as well as their contribution to the accomplishments achieved in the first year, is appreciated. It is encouraged that potential customers contact the organization so they can begin early coordination for their program test planning and event support.

JMETC is the corporate solution for joint distributed testing and is here now! ❑

RICHARD LOCKHART is a member of the Senior Executive Service and has over 30 years of service in the Department of Defense. He currently serves as the principal deputy director, Test Resource Management Center (TRMC), under the Office of the Under Secretary of Defense for Acquisition, Technology and Logistics (OUSD[AT&L]). He is the principal staff assistant and advisor to the TRMC director for strategic planning and assessment of the Major Range and Test Facility Base. In addition to his principal deputy responsibilities, he also serves as the director for the Investment Programs and Policy Division within TRMC. E-mail: Richard.Lockhart@osd.mil.

BERNARD "CHIP" FERGUSON is the program manager for Test Resource Management Center's Joint Mission Environment Test Capability (JMETC) Program. Since joining the Army in 1965, he has held leadership positions in combat units, varied level staffs, the Army's Operational Test and Evaluation Command, and Office of the Director, Test and Evaluation, Office of the Secretary of Defense. Upon retirement of active duty, he became a division manager and operations manager with SAIC supporting test and evaluation in DoD. With his vast experience in distributed testing and evaluation, he was selected for his current position in 2006. E-mail: Chip.Ferguson@osd.mil.

GEORGE RUMFORD, systems engineer, chairs the JMETC Users group, an entity of customers and potential customers that meets quarterly to review JMETC features, sites, and future plans. The test community is encouraged to participate in these meetings. E-mail: George.Rumford@osd.mil.

ITEA Journal 2008; 29: 167–174

A Framework for Understanding Experiments

Richard A. Kass, Ph.D.
GaN Corporation,
U.S. Army Operational Test Command, Fort Hood, Texas

While experimentation is an integral aspect of the capability development and acquisition process, its methods may be less familiar to testers. This article provides a framework for understanding the essence of an experiment, its central components, requirements for validity, and programmatic ways to increase experiment validity thru experiment campaigns. A follow-up article will compare experiments to tests.

Key words: Causality, experiment design, experiment campaigns, capability development, model-exercise-model, hypothesis, validity requirements.

An article on experiment techniques[1] should be an interesting read for this audience of testers. When asking test engineers and analysts whether testing and experimenting are similar activities, about half might agree they are similar. A similar question to experimenters located in Service battle labs would find far fewer considering test and experiment similar. The U.S. Department of Defense (DoD) has differentiated between testing and experimenting; tying tests to the acquisition process and experiments to the concept and capability exploration process. So is there a difference in test and experiment techniques?

The answer to this question is in two parts. Readers of this journal are familiar with the nature of testing and test design. This initial article will therefore characterize warfighting experiments and their design requirements. A follow-up article in the next issue will then compare experiments with tests.

Experiments and the capability development process

Tests are conducted on early capability modules, subsystems, prototypes, and production items to quantify the degree of design success. Experiments are also employed throughout this process. Experiments provide a scientific empirical method to identify capability gaps, explore alternative solutions, and develop and continuously update implementation techniques.

Prior to initialization of a capability development process, early experiments identify future warfighting gaps and assess relative merits of proposed doctrine, organization, training, materiel, leadership, personnel, and facilities (DOTLMPF). Analyses of alternatives (AOA) include experiments conducted with combat simulations.

Early in the acquisition process, experiments compare alternative designs and alternative competing solutions. Later, prior to testing of early prototypes, experiments assist combat developers in assessing new tactics, techniques, and procedures (TTP) required for optimizing employment of the new capability. After capability fielding, warfighting experiments can continuously examine opportunities to further enhance capability employment as environments and threats evolve.

Definition of a warfighting experiment

In its simplest formulation, to experiment is to try. In this sense, experimentation is a characteristic of human nature and has existed from earliest times. When early humans attempted different ways to chip stone into cutting edges or selected seeds to grow sturdier crops, they were experimenting.

More formally, "…to experiment is to explore the effects of manipulating a variable." (Shadish, Cook, and Campbell 2002).

This definition captures the basic themes of gaining new knowledge (explore), doing something (manipulating a variable), and causality (the effects). Based on their general definition, the author offers the following derivatives for warfighting experimentation:

Warfighting Experimentation—to explore the effects of manipulating proposed warfighting capabilities or conditions.

Experiment cause and effect and hypotheses

Identifying experiments with the investigation of causality is the key to understanding experiments and

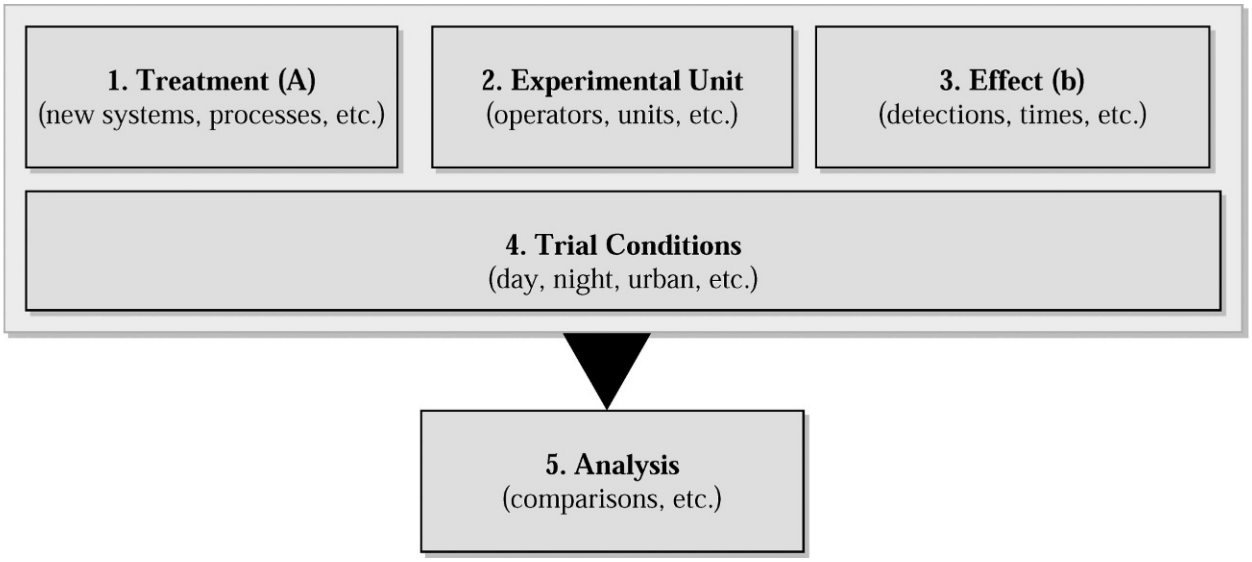


Figure 1. Five elements of an experiment

linking experiments to the transformation process. Causality is central to the transformation process. Military decision-makers need to know what to change in order to improve military effectiveness. The antecedent causes of effectiveness must be understood in order to change effectiveness. Effectiveness is improved by altering its antecedents, its causes. "Today, the key feature common to all experiments is still to deliberately vary something so as to discover what happens to something later—to discover the effects of presumed causes." (Shadish, Cook, and Campbell 2002). The notion of cause and effect is inherent in the language of experimentation and in its basic paradigm "let's do this and see what happens." All warfighting innovation questions can be translated into cause-and-effect questions expressed as: "does A cause B?" Does the proposed military capability (A) produce (cause) an increase in warfighting effectiveness (B)? This theme is fundamental to constructing the experiment hypothesis:

If a unit uses the new capability (A),
then it will increase in effectiveness (B).

Hypotheses are expectations about A causing B. The nature of experiment hypotheses prepares us to understand the five key components common to all experiments.

Five elements of an experiment

In large experiments with many moving parts it is sometimes difficult to see the forest for the trees. All experiments—large or small, field or laboratory, military or academic, applied or pure—can be described by five basic components (Cook and Campbell 1979) as depicted in *Figure 1*; and all five are related to causality.

1. The treatment, the possible cause (A), is the proposed capability, the proposed solution that is expected to influence warfighting effectiveness;
2. The experimental unit executes the possible cause and produces an effect;
3. The possible effect (B) of the treatment is the result of the trial, an increase or decrease in some aspect of warfighting effectiveness;
4. The trial is one observation of the experimental unit employing the treatment (A) or its variation (-A) to see whether effect (B) occurs and includes all of the contextual conditions under which the experiment is executed; and
5. The analysis phase compares the results from one trial to a different trial to quantify the impact of A on B.

Four requirements for a valid experiment

While defense experiment agencies have developed lists of lessons learned and best practices[2] to increase experiment rigor (validity); experiment validity is rarely formally defined. The adjective valid is defined as follows:

"Valid: well-grounded or justifiable, being at once relevant and meaningful, logically correct. [Synonyms: sound, cogent, convincing, and telling.]"—Merriam-Webster Dictionary online, 2006

When this definition is combined with the notion of cause-and-effect, a definition of a valid experiment is apparent: A valid experiment provides sufficient evidence to make a conclusion about the truth or

Hypothesis: If A, then B

Requirement		Evidence of Validity	Threat to Validity
1	Ability to use the new capability.	A occurred.	The asset did not work or was not used.
2	Ability to detect change.	B changed as A changed.	Too much noise. Cannot detect any change.
3	Ability to isolate the reason for the change.	A alone caused B.	Alternate explanations for the change are available.
4	Ability to relate results to actual operations.	Change in B due to A is expected in	The observed change may not be applicable.

Figure 2. Four requirements for a good (valid) experiment

falsity of the causal relationship between the manipulated variable and its effect.

How does one design an experiment to ensure sufficient validity? All of the good practices for designing warfighting experiments can be organized under four logically sequenced requirements[3] that must be met to achieve a valid experiment *(Figure 2)*. A simple example will illustrate these four requirements. Suppose a capability-gap analysis postulates that new sensors are required to detect time-critical targets. An experiment to examine this proposition might be a 2-day military exercise in which the current array of sensors is employed on the first day and a new sensor suite is used on day two. The primary measure of effectiveness is the percent of targets detected. The hypothesis is: "If new sensors are employed, then time-critical target detections will increase." This experiment is designed to determine whether the new sensors (A) will cause an increase in detections (B).

Ability to use the new capability

In most warfighting experiments, the majority of resources and effort are expended to bring the new experimental capability to the experiment. In the ideal experiment, the experimental capability (the new sensor) is employed by experiment players to its optimal potential and allowed to succeed or not succeed on its own merits. Unfortunately, this ideal is rarely achieved in experiments. It is almost a truism that the principal lesson learned from a majority of experiments is that the new capability, not withstanding all effort expended, was not ready for the experiment.

The experimental capability may not be ready for a number of reasons. The hardware or software does not perform as advertised. The experiment players are undertrained and not fully familiar with its functionality. Because it is new, techniques for optimum employment are not mature and by default, will be developed by the experimental unit during the initial experiment trials. If the experimental sensors (A) cannot be functionally employed during the experiment, there is no reason to expect they will detect targets (B) more often than the current array of sensors.

Ability to detect change

If the first experiment requirement is met, then transition from current to new sensors should be accompanied by a change in detections observed. If change in detections does not occur, the primary concern now is too much experimental noise. Ability to detect change is a signal-to-noise problem. Too much experimental error produces too much variability, making it difficult to detect change. Many experiment techniques are designed to reduce experiment variation: calibrating instrumentation to reduce data collection variation, limiting stimuli (targets) presentation to only one or two variations to reduce response (detections) variation, and controlling external environment variations (time of day, visibility, etc.). Sample size also affects the signal-to-noise ratio. Computation of statistical error variability decreases as the number of observations increases.

Ability to isolate the reason for change

Let us suppose the experimenter meets the first two requirements: the new sensors are effectively employed and the experiment design reduces variability and produces an observable change (increase) in detections. The question now, is the detected change due to the intended cause (changing from old to new sensors) or due to something else. The scientific term for alternate explanations of experimental data is confounded results. In this example, an alternate explanation for any increased detections on day two is that it was due to a learning effect. The sensor operators may have been more adept at finding targets on day two because of their experience with target presentations on day one, and consequently, would have increased target detections on day two whether the sensors were changed or not. This potential learning effect dramatically changes the conclusion of the detected change.

Scientists have developed experimental techniques to eliminate alternate explanations for observed change. These include counter-balancing the presentation of stimuli to the experimental unit, use of placebos in drug research, inclusion of a control groups, and randomizing participants between treatment groups.

Ability to relate the results to actual operations

Again, let us suppose that the experiment is successful in employing the new capability, detecting change, and isolating the cause. The final question is whether experimental results are applicable to operational forces in actual military operations. Experiment design issues supporting generalization include operational realism, representativeness of surrogate systems, use of operational forces as the experimental unit, and use of operational scenarios with a realistic reactive threat.

Tradeoffs in designing experiments

A fundamental implication from these four experiment requirements is that a 100 percent valid experiment is not achievable. The four experiment requirements cannot be fully satisfied in one experiment. Satisfying one works against satisfying the other three. Thus, decisions need to be made as to which validity requirements are to be emphasized in any given experiment.

All experiments are a balance between the four validity requirements. Precision and control increase the ability to detect and isolate change but often lead to decreases in ability to relate results to actual operations. Experiments that emphasize free play and uncertainty in scenarios reflect conditions found in existent operations and satisfy external validity Requirement 4, the ability to relate results. Conversely, experiments emphasizing similar conditions with diminished free play across multiple trials serve to reduce experiment noise and confounding, thus satisfying internal validity Requirements 2 and 3, the ability to detect and isolate change.

Validity priorities differ for any given experiment. Experimenters need to minimize the loss of one validity requirement because of the priority of another. However, tradeoff is inevitable. In settings where one expects a small effect and it is important to determine the precise relationship between the experiment treatment and its effect, the priority should be internal validity. On the other hand, if one expects a large effect and it is important to determine if the effect will occur in the operational environment with typical units, then external validity is the priority.

Different warfighting experiment methods provide different strengths

Warfighting experiments can be grouped into one of four general methods: Analytic war-game, constructive, human-in-the-loop, and field experiments. The experiment requirements just discussed provide a structure for recognizing the strengths and weaknesses of these four experiment methods. Relative strengths in meeting a requirement when employing a particular method is depicted by the number of plus signs in *Figure 3*.

Analytic war-game experiments typically employ command and staff officers to plan and execute a military operation. At certain decision points, the Blue players give their course of action to a neutral, White Cell, which then allows the Red players to plan a counter move, and so on. The White Cell adjudicates each move using simulations to help determine the outcome. Typical war-game experiments might involve fighting the same campaign twice, using different capabilities each time. The strength of war-game experiments resides in the ability to detect any change in the outcome, given major differences in the strategies used. Additionally, to the extent that operational scenarios are used and actual military units are players, war-game experiments may reflect real-world possibilities. A major limitation is the inability to isolate the true cause of change because of the myriad differences found in attempting to play two different campaigns against a similar reactive threat.

Constructive simulation experiments reflect the closed-loop, force-on-force simulation employed by the modeling and simulation community. In a closed-loop simulation, no human intervention occurs in the play after designers choose the initial parameters and then start and finish the simulation. Constructive simulations allow repeated replay of the same battle under identical conditions while systematically varying parameters: insertion of a new weapon or sensor characteristic, employment of a different resource or

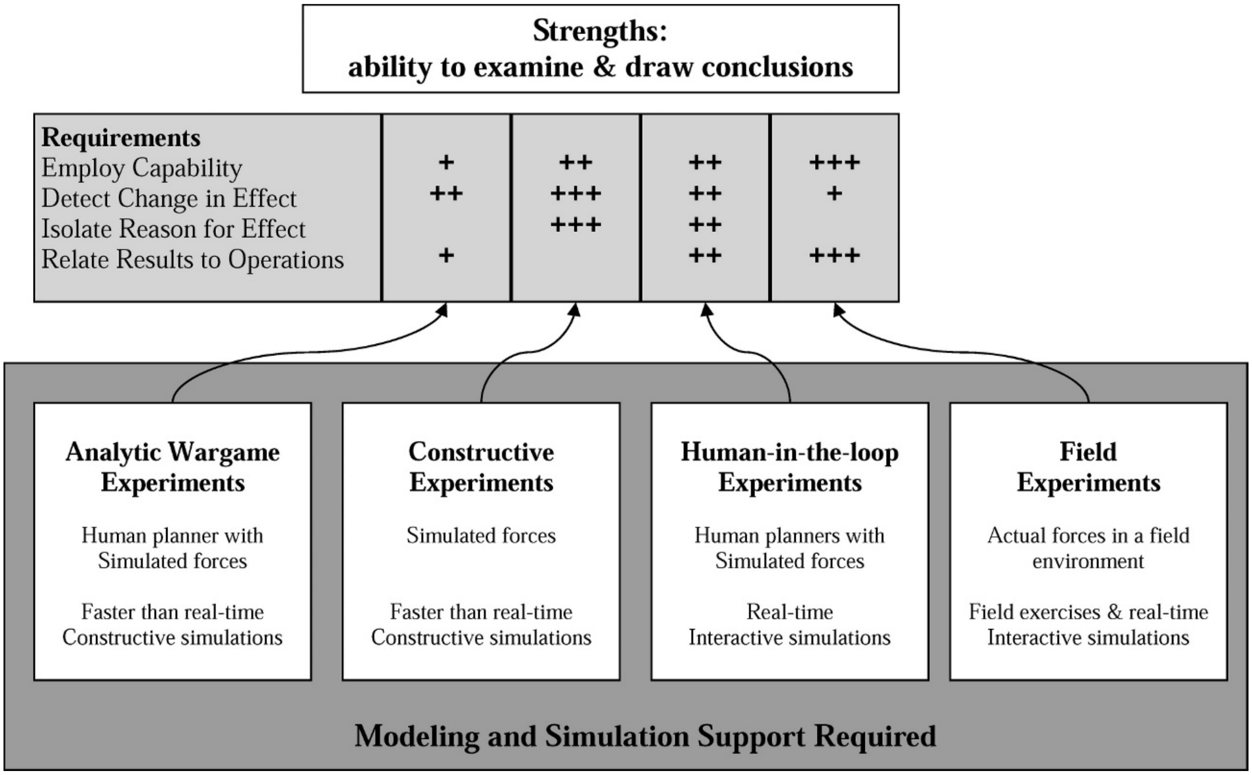


Figure 3. Different experiment venues have different strengths

tactic, or encounter of a different threat. Constructive simulation experiments with multiple runs are ideal to detect change and to isolate its cause. Because modeling complex events requires many assumptions, critics often question the applicability of constructive simulation results to operational situations.

Human-in-the-loop virtual experiments are a blend of constructive experiments and field experiments. In a command and control human-in-the-loop warfighting experiment, a military staff receives real-time, simulated sensor inputs, makes real-time decisions to manage the battlespace, and directs simulated forces against simulated threat forces. The use of actual military operators and staffs allows this type of experiment to reflect warfighting decision-making better than purely closed-loop constructive experiments. However, humans often play differently against computer opponents than against real opponents. Additionally, when humans make decisions, variability increases, and changes are more difficult to detect.

Field experiments are war-games conducted in the actual environment, with actual military units and equipment and operational prototypes. As such, the results of these experiments are highly applicable to real situations. Good field experiments, like good military exercises, are the closest thing to real military operations. A major advantage of the previous three experiment venues is their ability to examine capabilities that do not yet exist by simulating those capabilities. Field experiments, on the other hand, require working prototypes of new capabilities. Interestingly, while field experiments provide the best opportunity to examine practical representations of these new capabilities, field experiments are the most difficult environment to employ a new capability—the new capability has to function and the operators need to know how to employ it. Difficulties also reside in detecting change and isolating the true cause of any detected change because multiple trials are seldom conducted in field experiments and the trial conditions include much of the uncertainty, variability, and challenges of actual operations.

Employing a campaign of experiments to increase validity

Since a single experiment method cannot satisfy all four requirements, a comprehensive experiment campaign is required. A campaign of experiments[4] can consist of a number of successive, individual experiments to fully examine proposed solutions to complex military problems. It can also consist of a set of experiments conducted in parallel with information and findings passed back and forth. A campaign of experiments can accumulate validity across the four requirements.

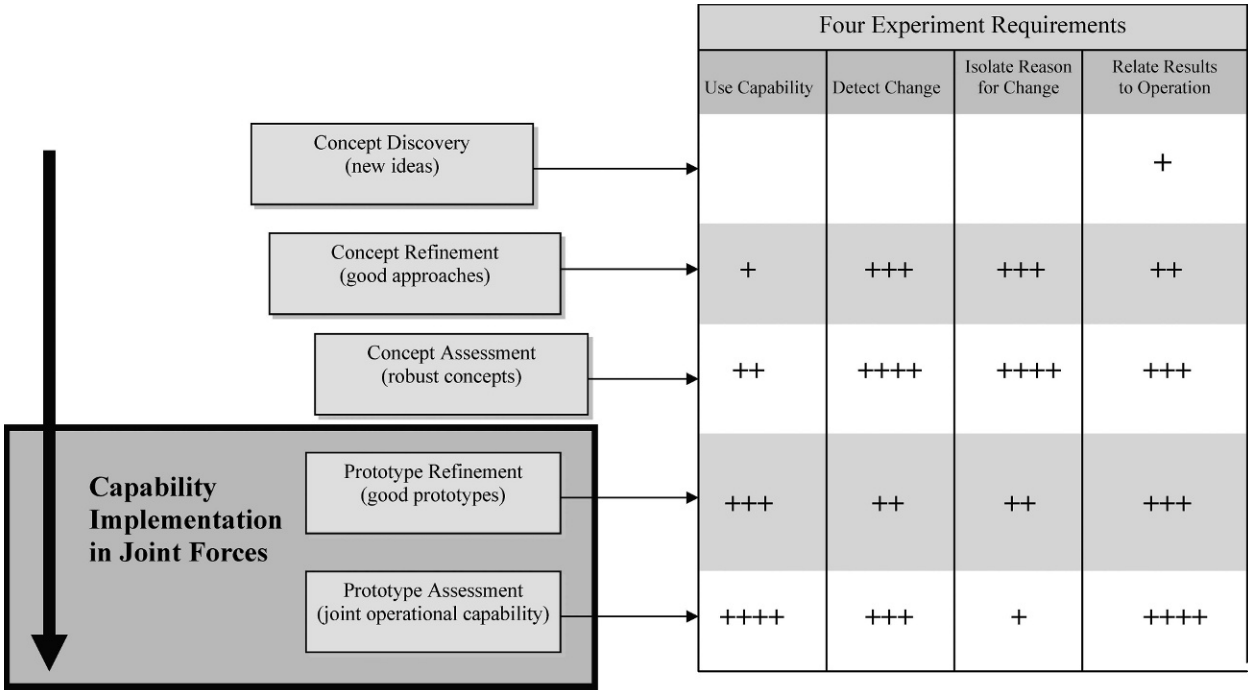


Figure 4. Experiment campaign requirements during the capability development process

Emphasizing different experiment requirements throughout the capability development process

A comprehensive capability-development program should include a campaign of individual experiments that emphasize different experiment requirements. *Figure 4* illustrates one example. The campaign starts at the top with discovery activities and proceeds to the bottom with capability implementation into the joint force. Each step in the campaign identifies possible experimentation goals. On the right of the experiment goals, the "pluses" portray the relative importance of the four validity requirements for that experimentation step. The following discussion identifies possible experiment venues that can be employed at each capability-development step to address the goals and validity requirements.

The primary consideration during concept discovery is relevance and comprehensiveness. To what extent do initial articulations of future operational environments include a comprehensive description of expected problems along with a full set of relevant proposed solutions? Relevancy, however, should not be overstressed. It is important to avoid eliminating unanticipated or unexpected proposals that subsequent experiments could investigate further.

Finding an initial set of potential capabilities that empirically show promise is most important in concept refinement. Early experiments here examine idealized capabilities (future capabilities with projected characteristics) to determine whether they lead to increased effectiveness. Initial concept refinement experiments are dependent on simulations to represent simulated capabilities in simulated environments. Accurately isolating the reason for change is less critical at this stage in order to permit "false positives." Allowing some false solutions to progress and be examined in later experiments under more realistic environments is more important than eliminating potential solutions too quickly. Concept refinement is dependent on the simulation-supported experiment such as constructive, analytic war-game, and human-in-the-loop experiments.

Quantifying operational improvements and correctly identifying the causative capabilities are paramount in providing evidence for concept assessment. Concept justification is dependent on experiments with better-defined capabilities across multiple environments. Constructive experiments can provide statistically defensible evidence of improvements across a wide range of conditions. Human-in-the-loop and field experiments with realistic surrogates can provide early evidence for capability usability and relevance. Incorporating human decision-makers into human-in-the-loop and field experiments is also essential early in the capability-development process. Human operators tend to find new ways to solve problems.

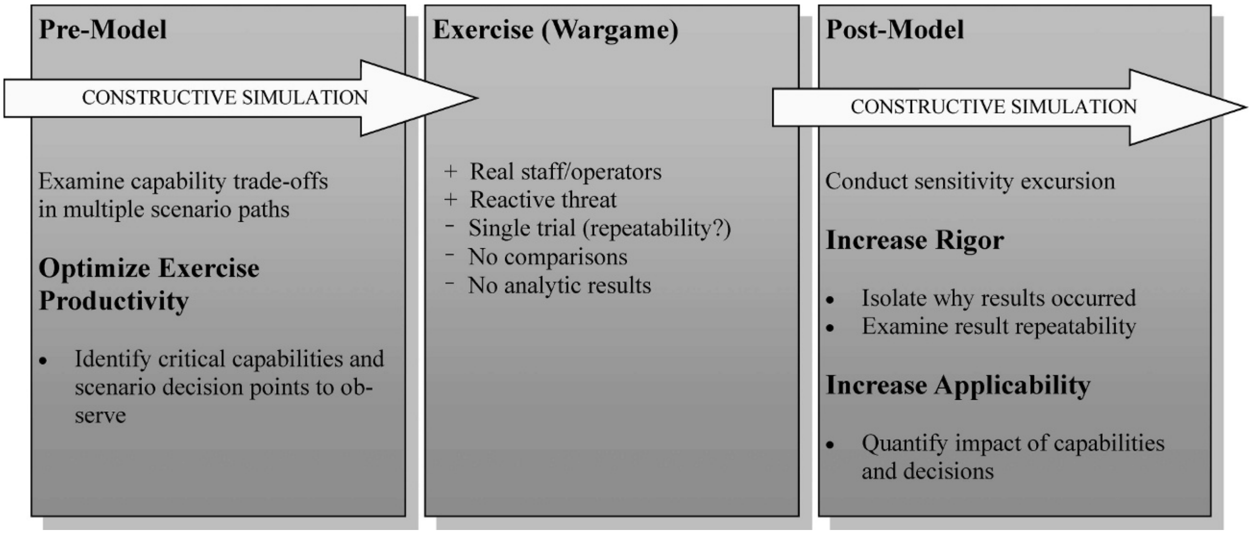


Figure 5. Model-exercise-model process

In prototype refinement, one should anticipate large effects or the implementation might not be cost effective. Accordingly, the experiment can focus on the usability of working prototypes in a realistic experiment environment. To do this, the experiment must be able to isolate the contributions of training, user characteristics, scenario, software, and operational procedures to prototype improvements in order to refine the right component. Human-in-the-loop and field experiments with realistic surrogates in realistic operational environments provide the experimental context for assessing gains in effectiveness. Human operators find unexpected ways to employ new technologies effectively.

Applicability to the warfighting operational environment is paramount in prototype assessment. If the capability is difficult to use or the desired gains are not readily apparent in the operational environment, it will be difficult to convince combatant commanders to employ it. Uncovering exact causal chains is less important while human operators are essential to ensuring that the new technology can be employed effectively. Prototype assessment experiments are often embedded within joint exercises and operations.

Emphasizing different experiment requirements via a model-exercise-model process

Another type of experiment campaign can be organized around the requirement to conduct large war-games or large field exercises to investigate the effectiveness of new capabilities. Because these large events are player resource intensive and often include multiple experimental capabilities, few opportunities exist to examine disentangled alternative capabilities or alternative situations that would allow meaningful comparisons. The model-exercise-model paradigm depicted in *Figure 5* can enhance the usefulness of war-games and exercises. This paradigm consists of conducting early constructive simulation experiments prior to the war-game or exercise and then following these events with a second set of postexercise constructive experiments.

Early constructive simulation experiments use the same Blue and Red forces anticipated to be played in the exercise. This pre-event simulation examines multiple alternative Blue force capability configurations against different Red force situations. This allows experimenters to determine the most robust Blue force configuration across the different Red force scenarios. It also helps to focus the exercise by pinpointing potential critical junctures to be observed during the follow-up exercise.

The war-game or exercise executes the best Blue force configuration identified during the pre-event simulation. The "best configuration" is the one indicated by pre-exercise simulation that the new capability dramatically improved Blue's outcome. The exercise reexamines this optimal configuration and scenario with independent and reactive Blue and Red forces. Choosing the scenario that provides the best opportunity for the new capabilities to succeed is best because large exercises include the "fog of war"—and experimental capabilities rarely perform as well in the real environment as in simulation. Therefore, it makes sense to give the new capability its best chance to succeed. If it does not succeed in a scenario designed to allow it to succeed, it most likely would not succeed in other scenarios.

Experimenters use the exercise results to calibrate the original constructive simulation for further postevent simulation analysis. Calibration involves adjusting simulation inputs and parameters to better match the play of the simulation to the play of the exercise. This adds credibility to the simulation. Rerunning the pre-event alternatives in the calibrated model provides a more credible interpretation of differences now observed in the simulation. Additionally, the postevent calibrated simulation can substantiate (or not) the implications of the exercise recommendations by conducting causal analysis. Causal analysis is a series of "what if" sensitivity runs in the simulation to determine whether the exercise recommendations make a difference in the calibrated simulation outcome. Postexercise simulation runs can also examine what might have occurred if the Red or Blue forces had made different decisions during the exercise.

Summary

Can experiments fail? Yes, they can fail to provide sufficient evidence to determine whether the manipulated variable does (or does not) cause an effect. If the experimenter is unable to answer each of the four requirements in a positive manner, a meaningful conclusion is not possible concerning the impact of a proposed capability.

Designing individual warfighting experiments is an art because every experiment is a compromise. The logical approach in this article provides an understanding of the choices available to meet the four experiment validity requirements and the strengths and weaknesses inherent in typical experiment venues. Designing an individual experiment involves making cognizant tradeoffs among the four requirements to provide sufficient credible evidence bounded by explicated limitations to resolve the hypothesis.

While a single experiment will not satisfy all four requirements, a campaign of experiments can accumulate validity and overall confidence in experiment results. A comprehensive experiment program includes a series of individual experiments, each emphasizing different experiment requirements. In this campaign, no single experiment is expected to carry the entire weight of the decision. Each experiment contributes and the final results are based on accumulated confidence with each individual experiment contributing its strength to the final conclusions. The whole is greater than any part.

So, how much of this is applicable to acquisition testing? The follow-up article in the next issue will discuss the similarities and difference between tests and experiments in several areas: The planning process—especially designing valid tests and experiments—along with the execution and reporting process. The next article will focus on clearing away misperceptions of where efficiencies could be gained by sharing resources and expertise. ❑

RICK KASS has 25 years in designing, analyzing, and reporting on operational field tests and military experiments. He held multiple positions as test officer, analyst, and test director for 18 years with the U.S. Army Test and Evaluation Command (USATEC) and was chief of analysis for 7 years with the U.S. Joint Forces Command (USJFCOM) joint experimentation program. Currently, Rick works for GaN Corporation supporting the Army's Operational Test Command at Fort Hood, Texas. He has authored over 25 journal articles on methods for research, experimentation, and testing and was the primary architect establishing the permanent Warfighting Experimentation Working Group in the Military Operations Research Society (MORS). Rick is a graduate of the National War College and holds a Ph.D. in psychology from Southern Illinois University. E-mail: rick.kass@us.army.mil

Endnotes

[1]This article draws heavily from portions previously printed in my book Kass, R. A. The Logic of Warfighting Experiments published in 2006 by the Command and Control Research Program (CCRP) of the ASD/NII which has graciously granted permission to include that material in this work. *Figures 1* through *5* here are *Figures 9, 8, 20, 39*, and *40* in that work. Readers can download or order the larger document from the CCRP website at http://www.dodccrp.org.

[2]A good discussion of many best-practices is found in Alberts, D. S. and Hayes, R. E. 2002 Code of Best Practices for Experimentation. DoD CCRP publication series, D.C.: U.S. Government Printing Office.

[3]*The Logic of Warfighting Experiments* devotes a separate chapter to each of the four validity requirements.

[4]For a comprehensive examination of the value of experiment campaigns to address warfighting problems see Alberts, D. S. and Hayes, R. E. 2005 Campaigns of Experimentation. DoD CCRP publication series, D.C.: U.S. Government Printing Office.

References

Cook, T. D. and Campbell, D. T. 1979. Quasi-Experimentation: Design and Analysis Issues for Field Settings. Houghton Mifflin: Boston.

Merriam-Webster Dictionary online. http://www.m-w.com

Shadish, W. R., Cook, T. D. and Campbell, D. T. 2002. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Houghton Mifflin: Boston. Page 507.

Shadish, W. R., Cook, T. D. and Campbell, D. T. 2002. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Houghton Mifflin: Boston. Page 3.

ITEA Journal 2008; 29: 175–178

Achieving Information Assurance Through Operational Test and Evaluation

Anil N. Joglekar, Ph.D.
Institute for Defense Analyses, Alexandria, Virginia

Over the last decade the U.S. Department of Defense (DoD) has moved to a network-centric environment in which the sensors, weapons, and command and control systems are networked for executing military missions at tactical and operational levels. Although not all systems are Internet Protocol based, they are all networked with different means, including data links. This reality makes the operations of combatant commands and the Services vulnerable to network attacks by casual hackers as well as nation states. Information assurance, which consists of protection, detection, and reaction, has become a critical test issue for net-centric systems, whether it is a sensor, weapon, or command and control system. Information assurance must be assessed systematically during operational test and evaluation. Technical aspects of information assurance (availability, confidentiality, integrity, authentication, and nonrepudiation) are generally assessed during developmental test and evaluation or during the certification and accreditation process, which supports interim/final authority to operate certification. The emphasis of this article is on subsequent information assurance assessments during operational test and evaluation, focusing on system operation after receiving the authority to operate. Information assurance posture is not just dependent on system designs, but also involves how systems are networked and operated; hence, information assurance is dependent on the abilities of system operators, systems administrators, and network designers to detect, protect from, and react to potential attacks. This article references the policies and guidelines issued by the DoD, operational metrics to be used in evaluation, and the process to conduct operational test and evaluation for most acquisition programs. The article also references a congressionally-mandated effort underway in which fielded systems are assessed for information assurance.

Key words: Computer network defense, cyber attack, information assurance, information security, net-centric operations, network vulnerability.

Over the past 10 years, U.S. military operations have become fairly network centric, even if not all systems are Internet protocol based. The Nonsecure Internet Protocol Network (NIPRNet) and Secure Internet Protocol Network (SIPRNet) have become the backbone of U.S. military operations along with the Link 16 and other data links. Most of the military missions involving sensors, command and control (C2) systems, and weapon systems are networked together because these systems are likely to be physically separated. Net-centric operations increase the vulnerabilities to cyber attacks by casual hackers, adversarial persons, and enemy nations. The United States needs to ensure that joint military campaigns can be conducted using the networks and linked systems and be defended against such adversaries.

The U.S. Department of Defense (DoD) has issued Directives 8500.1 and Instruction 8500.2 for Information Assurance (IA), and Instruction 8510.01 for the certification and accreditation process. However, the focus of these documents is on assigning and assessing IA controls to obtain authority to operate. Much of this process is based on an online checklist that does not address IA in operational configuration; hence, the process by itself does not provide the IA posture of the system for Services and Combatant Command operations.

Each Service and Combatant Command uses the NIPRNet, SIPRNet, Satcom, data-links and fiber,

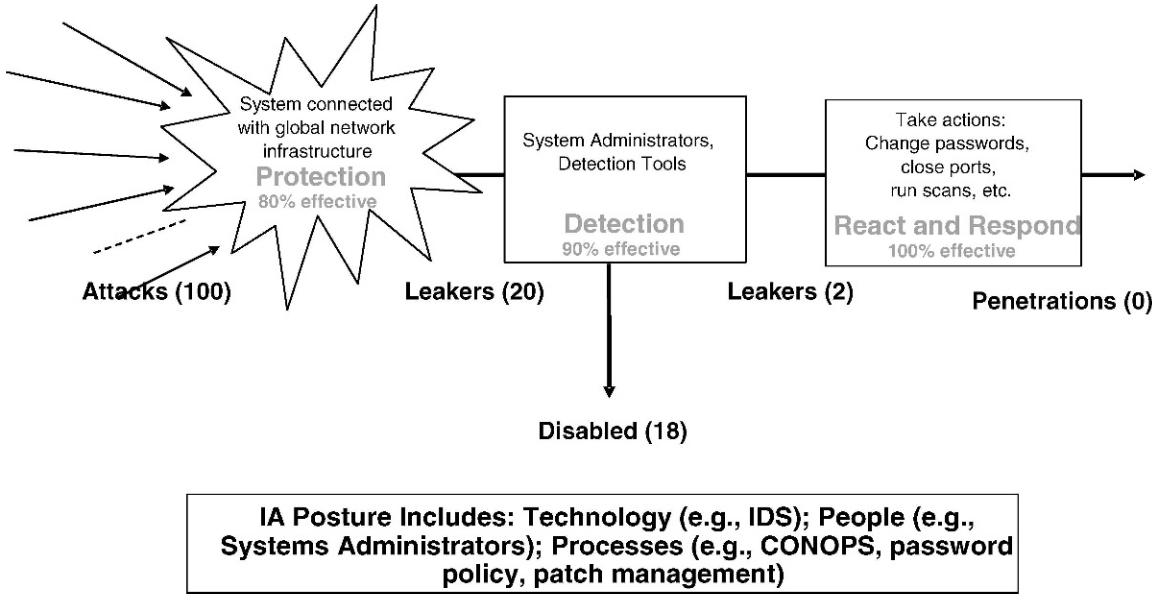


Figure 1. Protect, detect, and react/respond components of Information Assurance

microwave routers, and wireless devices to design their networks. However, the network architecture for each organization is different and needs to be assessed separately for information assurance. Although the Joint Task Force for Global Network Operations (JTFGNO) is responsible for IA, the ultimate burden is on systems administrators belonging to the Services and combatant commands who are responsible for securing and maintaining the network.

In order to ensure that systems and networks delivered to the warfighter are assured for information security, the Director, Operational Test and Evaluation (DOT&E) promulgated a policy for Operational Test and Evaluation for Information Assurance on November 16, 2006 for all acquisition programs that are on DOT&E oversight. Recognizing the need to assess IA of systems and networks already fielded, Congress has been funding DOT&E for the past four years to assess each combatant command's information assurance posture.

Definition of information assurance

The DoD definition of information warfare is broad and includes electronic warfare, cyber warfare, and psychological operations. For this article, the definition of information warfare means only cyber defense or computer network defense and will not deal with electronic warfare or psychological operations. With this definition, the technical components of information assurance are: availability, integrity, authentication, confidentiality, and nonrepudiation of information. While combatant commands are not overly concerned with the technical aspects, they want to ensure that the mission-critical information is secure and available when needed. However, they must understand that information assurance is not achieved just by protecting networks, but must operationally involve detection of penetrations and reaction and responses to those penetrations by systems administrators. The protection, detection, reaction, and response functions are known as defense in depth, as shown in *Figure 1*.

As illustrated in *Figure 1*, information assurance is provided by technology (firewalls, intrusion detection devices, network visualization tools, etc.), people (systems administrators, users, etc.), and processes (password policies, concept of operations, vulnerability patch updates, closing of unused ports, wireless encryption, etc.). Thus, achieving information assurance is a closed-loop process and requires all four components of protection, detection, reaction, and response.

How to achieve information assurance

In order to achieve high information assurance posture, leadership must be aware of potential information warfare threats and be educated about their own system vulnerabilities. Users still lack adequate appreciation concerning their system and network vulnerabilities. There are enough policies, directives, and regulations prescribed by the Department of Defense. Yet, many warfighters still perceive information assurance as related to information technology systems, "geeky" systems that do not apply to ships, aircraft, and land combat vehicles. Achieving information assurance starts with system designs.

Today's philosophy is largely to provide just the perimeter defense rather than designing each system to protect itself to some degree. Each system designer needs to ensure that firewalls are included and configured correctly, and unneeded ports and services are closed by default.

From an operational test and evaluation perspective, achieving information assurance requires the following five steps.

Step 1: Determine if the system is a candidate for assessing information assurance. Any system (sensor, weapon, or command and control system) that provides information to another system or network, or receives information from another system or network (even if the information is received by insertion of a media), needs to be assessed for information assurance. The networks can be local area networks (LAN) or wide area networks (WAN). If wireless channels are used, then information assurance testing is a must.

Step 2: The higher the mission criticality and confidentiality of the system, the more critical is the need for information assurance. Unclassified systems that maintain personnel records, medical information, and payroll information should be assessed as rigorously as secret or top secret systems.

Step 3: In order to connect to Global Network Infrastructure, Certification and Accreditation (C&A) procedures are prescribed by DoD. These procedures primarily involve assessment via checklist as prescribed in DoDI 8500.2. Based on this documentation, a system receives an interim/final Authority to Operate authorizing connection to a network. The data in this process are contained in the DoD Information Assurance Certification and Accreditation Process (DIACAP) Knowledge Service. However, for multiple reasons a Designated Accreditation Authority (DAA) may grant a waiver to the C&A process. If DIACAP data exist, then the operational testers need to review this data and identify the potential system vulnerabilities.

Step 4: The operational tester should review the DIACAP data to determine where the system is most vulnerable. However, if DIACAP data do not exist, then a Blue Team assessment should be conducted to identify system vulnerabilities. Even if DIACAP data exist, a Blue Team assessment helps identify vulnerabilities in the operational configuration prior to assessment by a Red Team.

Step 5: Vulnerabilities identified in Step 4 should be corrected prior to the Red Team penetration testing. A Red Team should be threat-representative and the system under test must be connected to all systems and networks with which it will be interfacing during peacetime and wartime operations. When the Red Team is conducting penetration testing, the system administrators should be required to detect those penetrations, and the time difference between actual penetration and detection by the system administrators should be measured as one of the metrics. Generally, Red Teams are not authorized to do any damage to the network; hence, if a penetration is detected, then system administrators should be simply asked to record what their reaction and response to that penetration would be. In addition, there should be documented procedures for restoring the system and data.

Blue and Red Team contrasts

The Blue Team generally has access to the system and network data and primarily focuses on identifying system vulnerabilities. The system under test is not necessarily connected to other systems and LANs and WANs. There is no function for the system administrators during the Blue Team assessment.

The Red Team, on the other hand, is representative of the threat. In order to minimize Red Team testing costs, they can be provided some information on the system. Red Teams can exploit all types of vulnerabilities such as social networking, physical access, phishing, trojans, and so on. When the Red Teams are attempting to scan and penetrate, the system administrators should attempt to detect the scanning/penetrations and then respond accordingly. For example, they can reboot, change passwords, close open ports, install necessary security patches, etc.

The capabilities of both Blue and Red Teams are similar. However, the scope of testing, environment for testing, and approach are generally significantly different. Blue Teams are advisors to systems administrators as to how to protect information and improve information assurance postures, whereas Red Teams want to exploit physical, social, and cyber vulnerabilities aiming to achieve root access to systems and networks.

Where to seek help

Many organizations have been created primarily to support Blue Team assessments and some can support Red Team assessments. The National Security Agency leads this effort. Other organizations are: Army First Information Operations Command, Air Force Information Operations Center, Navy Information Operations Center, Marine Corps Information Assurance Assessment Team, Defense Information System Agency Field Security Office, Joint Interoperability Test Command, and Defense Logistics Agency Computer Emergency Response Team, to list a few key organizations.

How to conduct Red Team Assessment

Red Team penetration testing must be planned early enough to ensure that testing is adequately resourced and testing would include detection, reaction, and response. Red Teams should have the capability representative of a threat for that system. Just hiring any Red Team would not be adequate. The Red Teams can leave their mark on the system under test without doing any damage. They should not tell the system administrators when they have penetrated the system and the burden should be on the system administrators to detect and react to the penetration. System administrators should be required to record the time of detection and response they would recommend to mitigate the impact of the penetration.

Improvement for IA Testing Policy and Guidance

Presently the Department of Defense policy and guidance is fragmented. DoD Directive 8500.1 and Instruction 8500.2 primarily focus on the DoD Information Assurance Certification and Accreditation component of the process; the Joint Staff guidance requires assessment of net-ready key performance parameter (NR-KPP); and the Director, Operational Test and Evaluation policy requires assessment in a truly operational environment. All these three policies and guidance documents should be consolidated into single, clearly articulated, policy guidance in order to help program managers, combatant commands, and network managers to do what is right to protect information and assure that net-centric operations will become a force multiplier for the allies and will not be exploited by the enemy.

Summary

Information assurance has become critical for military and civilian operations. Information assurance must be taken into consideration during the design phase of the system and should not be considered as an add-on after systems are designed. Further, information assurance cannot be achieved only by protecting systems and networks. Information assurance can only be achieved by systems administrators operating the systems and networks and employing processes to detect penetrations and react and respond to those penetrations from unauthorized and hostile personnel. The certification and accreditation process should be strengthened and should be made mandatory prior to granting initial authority to operate.

Repeat vulnerabilities should not be ignored, and inabilities to detect and react should not be taken lightly. Investments should be adequate to keep security patches up-to-date. ❑

DR. ANIL JOGLEKAR is a senior research staff scientist with the Institute for Defense Analyses (IDA) in Alexandria, Virginia. He oversees the operational test and evaluation of net-centric command, control, and communications systems, the Joint Test and Evaluation (JT&E) program, and was responsible for an effort to develop the national personnel recovery architecture that covers all government agencies including the Department of Defense. Dr. Joglekar received the National Defense Industrial Association's 2000 Tester of the Year Award. Dr. Joglekar has been with IDA since 1979 and has led studies related to NATO airspace management, combat identification systems, Patriot fielding in NATO, and the Joint Air Defense Operations/Joint Engagement Zone Joint Test. Dr. Joglekar led the DOT&E's team for the CINC year-2000 (Y2K) operational evaluations. He was instrumental in the development of operational test and evaluation policies for Interoperability and Information Assurance (IA). Dr. Joglekar received his Ph.D. degree from Stanford University in 1973 in aeronautics and astronautics. E-mail: ajogleka@ida.org

Acknowledgments

The author would like to acknowledge the contributions and initiatives taken by Bill McCarthy, Col Ralph Harris, Kirk Johnson, and Steve Gates, all from the Office of Director, Operational Test and Evaluation, over the past many years to place emphasis on this emerging area and supporting assessment in this area. Similarly, the author would like to acknowledge his colleague, Dr. Nat Subramonian, who has done tremendous work to advance testing for information assurance for acquisition programs and who also reviewed this article.

ITEA Journal 2008; 29: 179–189

Mission-Model Driven Process: Test and Evaluate Composition of SOA Services

Prem P. Jain, Ph.D. and Ben King
MITRE Corp., McLean, Virginia

The Service Oriented Architecture (SOA) paradigm offers an opportunity to rapidly deploy new mission capabilities by composing verified SOA services thereby shifting the critical path from development of services to integration, validation, test, and evaluation tasks. Realization of this potential requires a well defined semi-automated process to achieve the needed agility and speed to deploy validated new mission capabilities within the constraints of doctrine and net-centric infrastructure. Fortunately, fast-growing commercial-off-the-shelf (COTS) support of converging SOA and business process modeling standards provide the U.S. Department of Defense an excellent opportunity to improve this process. This article describes the semi-automated mission-model-driven MLM2SOA process, a COTS-based prototype implementing the process, and its illustrative use in a missile defense vignette composed of sensor track processing and Google Earth services. Current COTS maturity provides a potential to execute the first MLM2SOA process iteration for a simple operational mission thread in weeks and subsequent iteration in days.

Key words: Business process modeling notation (BPMN); capability modules; commercial-off-the-shelf solutions; mission threads; MLM2SOA process; service oriented architecture.

The hundreds of U.S. Department of Defense (DoD) information systems in the field today are expensive, complex, and specialized. Each was built independently by various program offices across the services and agencies. Acquisition in most cases took many years. These systems are generally not interoperable with each other. Testing is an onerous and expensive process that occurs at various discrete milestones in the development process.

In contrast, DoD's "net-centric" strategy intends that all future information systems will efficiently share information and processing capability across a common computer network. To accomplish that goal, architectural design must focus on information flow across systems and on how to modularize, rather than specialize, capabilities. The Service Oriented Architecture (SOA) paradigm offers such a potential.

To realize this net-centric vision, the current system-centric acquisition process must evolve to capability based acquisition (CBA) as defined by the DoD Instruction 5000.2 (DoD Instruction 5000.2, 2003). The long-term goal of CBA is to field modular capabilities that can be composed on-the-fly by warfighters in the field to create the particular mission capability they need. Further, CBA intends to field these mission capability modules continuously over the course of weeks and months, rather than fielding complex end-to-end systems over the course of years. Accordingly, rather than discrete and onerous, the test model must become continuous, and adaptive.

Introduction

Mission threads, widely identified as a desirable method to capture DoD acquisition requirements, modeled in a machine-executable form provide the needed mission context to verify and validate (V&V) operational performance early in the acquisition process. Available cost-effective commercial-off-the-shelf (COTS) solutions automate the integration of executable mission models in an SOA environment, providing agility and speed to the V&V process. COTS solutions are also available that can automatically generate a SOA test environment in the form of workflows expressed in the business process executable language (BPEL) standard (OASIS, 2007). Multiple COTS vendors currently support execution of BPEL (via BPEL workflow engine) in their respective environment for functional testing of the SOA services deployed at disparate nodes on a network. These services can be instrumented to collect performance parameters, which can then be fed back to the mission thread model to improve accuracy of earlier projection of mission performances.

MITRE Corporation (MITRE) has been exploiting these standard-based trends over the last two years for developing a mission level modeling (MLM) and analysis capability. MLM captures operational mission thread information flows among operational and systems nodes using the executable graphical business process modeling notation (BPMN). (OMG, 2006). MITRE is also experimenting with a mission-model-driven semi-automated process titled MLM2SOA to meet CBA emerging needs. The MLM2SOA process is based on BPMN and BPEL standards and supports a Model-Test-Model iterative development paradigm. Iteration consists of modeling a test event operational mission thread (OMT), analyzing mission performances for a specific composition of verified SOA services and specific workload using simulation, generating the SOA test ready workflow, and testing composition of services.

This article describes the MLM2SOA process, a COTS-based prototype automating some of the process steps and exercising the prototype for a missile defense vignette, processing sensor data, and displaying tracks via a composition of SOA services. Initial results offer significant encouragements.

T&E challenges for DoD net-centric capabilities

Current DoD T&E practices use constructive, virtual, and real simulation environments that inter-operate via standards such as the High Level Architecture. Realizing net-centric potential requires significant changes to material acquisition processes and an overhaul to T&E practices. Net-centric capability development provides the T&E community with new challenges:

1. Non-availability of dedicated and controlled physical environment for end-to-end T&E.
2. Mitigating risk rather than avoiding risk to field capabilities in weeks-to-months.
3. Significant degree of unknowns for the services interactions with other services and the configuration of infrastructure to execute these services.
4. Projecting confidence with limited test and incorporating the results from different programs.
5. Isolating and debugging functional and timing errors, identified in end-to-end testing in a distributed environment.

Net-centric environments have a significant degree of concurrency both at the operator and system level. Significant improvements can be made in operational effectiveness by exploiting the inherent concurrency of net-centric environments. This, however, requires the understanding of different types of dependencies such as data, control, resource sharing among concurrent activities, and the impact of larger workloads such as the number of targets and inter-arrival time of targets. These dependencies lead to significant T&E challenges for the temporal behavior such as measures of effectiveness (MOE), measures of performance (MOP), and key performance parameters (KPP) under variable conditions. In general, test methods must fundamentally change to support capability-based rather than current scenario-based acquisition. Following are some of the proposed tenets for dealing with evolving T&E challenge:

1. Separate testing of SOA services within an architecture from testing the architecture itself.
2. Derive the test confidence from testing throughout the acquisition process using models and not merely from an operational test event at 5000.2 Milestone “C”.
3. Reduce test cost by exploiting the SOA architecture’s loose coupling among services and data.
4. Reduce test time by testing at multiple levels simultaneously.
5. Test service integration and services using simulated services.
6. Use modeling & simulation (M&S) for testing scalability and variability.
7. Improve M&S accuracy by feeding measured parameter values to model.
8. Use model-test-model paradigm.
9. Rapidly re-configure mission test environment using a set of constructive and virtual controllable components.

The MLM2SOA process uses these tenets and is described in a later section in this article, “The MLM25SOA process.” This process is just a step for the major transition, T&E community must undertake supporting CBA using SOA paradigm.

Role of mission threads in DoD acquisition

Historically the DoD T&E requirements process is a years’ long, formal, serial process that delivers long expensive paper documents. Similar documents are created for similar programs, but not shared. Translating these documents into measurable and testable criteria evaluating mission (as distinct from "system") performance is difficult.

The Joint Capabilities Integration & Development Systems (JCIDS) process envisions using mission threads to support CBA. Specifically, mission threads will be cited across the JCIDS family of Capability Development Documents to define testable MOE, MOP, and KPP. Mission threads are used in multiple stages of the DoD acquisition process from 5000.2 (DoD Instruction 5000.2 2003) pre-milestone "A" to operational testing in milestone "C". Test event OMTs provide the context for DoD T&E.

A mission thread is a precise, objective, description of an important task. In other words, a mission thread is a time ordered operational event diagram that captures discrete, definable, interactions among human operators and/or technological components. Unfortunately, current forms of mission thread documentation in various non-executable static forms (Microsoft Office products, Visio files and/or DoDAF products) are not adequate to deal with the rising joint operational and systems complexities.

By reducing tasks to their basic components, mission threads provide a means to discover concurrency among operational tasks and systems across mission domains. Further, because they define discrete objective events, mission threads are inherently programmable into machine readable format. Thus the task of describing likely multi-use capability modules can be automated.

Exploiting SOA and business process management (BPM) for DoD

SOA is an approach to software design that aims to improve business process and leverage economy of scale by sharing common software capability modules. If we use machine executable mission threads to design and test multi-use capability, and SOA to deploy it, we can achieve the compose-able capability objective of DoD net-centric vision. That said, there is a great deal of devil in the details!

SOA separates services from data and exposes data for discovery and concurrent use across a myriad of DoD services and agencies and mission threads allow us to properly design and test the shareable services. The question is how mature are the commercial standards associated with capturing machine readable mission threads and creating reusable and discoverable services?

Commercial industry is making steady progress adopting and adapting standards (e.g., WSDL, SOAP, OWL, RDF, and UDDI) for data discovery, access and understanding. Commercially deployed SOA rapidly provisions capability for various applications, e.g., online auctions, banking, travel services, searching, etc. In those business applications, SOA allows quick adaptation to changes in operational needs and system implementation technologies and enable organizations to quickly deal with emergent threats and opportunities.

Fortunately, in the last few years, progress has been made in executable standard modeling languages such as BPMN [4], which are suitable to capture mission threads. BPMN standard language contrasts with historical trends of using proprietary languages in simulation tools such as OPNET and EXTEND. This trend has significant implications to DoD:

1. This now allows multiple simulation vendors to compete reducing the price to a few hundred dollars compared to tens of thousands dollars for tools based on proprietary languages.
2. This also allows modelers to use their favorite tool and enables the composition of models developed on different tools.
3. Business analysts, rather than engineers, can model using highly intuitive graphical language and can simulate business behavior using different COTS tools.

COTS support of converging SOA and BPM standards is growing at a fast rate leading to agile and speedy deployment of capabilities. These modeling technologies are converging to realize business efficiencies from the SOA paradigm.

The need for agile and speedy business-to-business (B2B) interactions is driving the commercial industry to adapt to the SOA paradigm for deploying information systems. End-to-end test of these information systems is still an open challenge. This being mission specific, ad hoc commercial solutions are used for each implementation. SOA adoption by DoD critically needs a repeatable mission-driven T&E process with predictable outcomes.

As the information interactions among DoD Joint missions mimic B2B interactions, where each DoD Service can be treated as a separate business, DoD can adopt the available standard-based COTS solutions to realize the long-term SOA benefits of agility and speed.

The MLM2SOA process takes advantage of the commercial support for both business modeling and SOA technologies, and especially, their mutual convergence. The iterative MLM2SOA process enables the DoD operational analysts to define, refine and simulate mission threads captured in BPMN executable standard graphical language, in contrast to current static documentation practice. It exploits the available COTS support for automated generation of test scenario and test configuration for live and virtual test; and feeding the test measurements into the mission thread model.

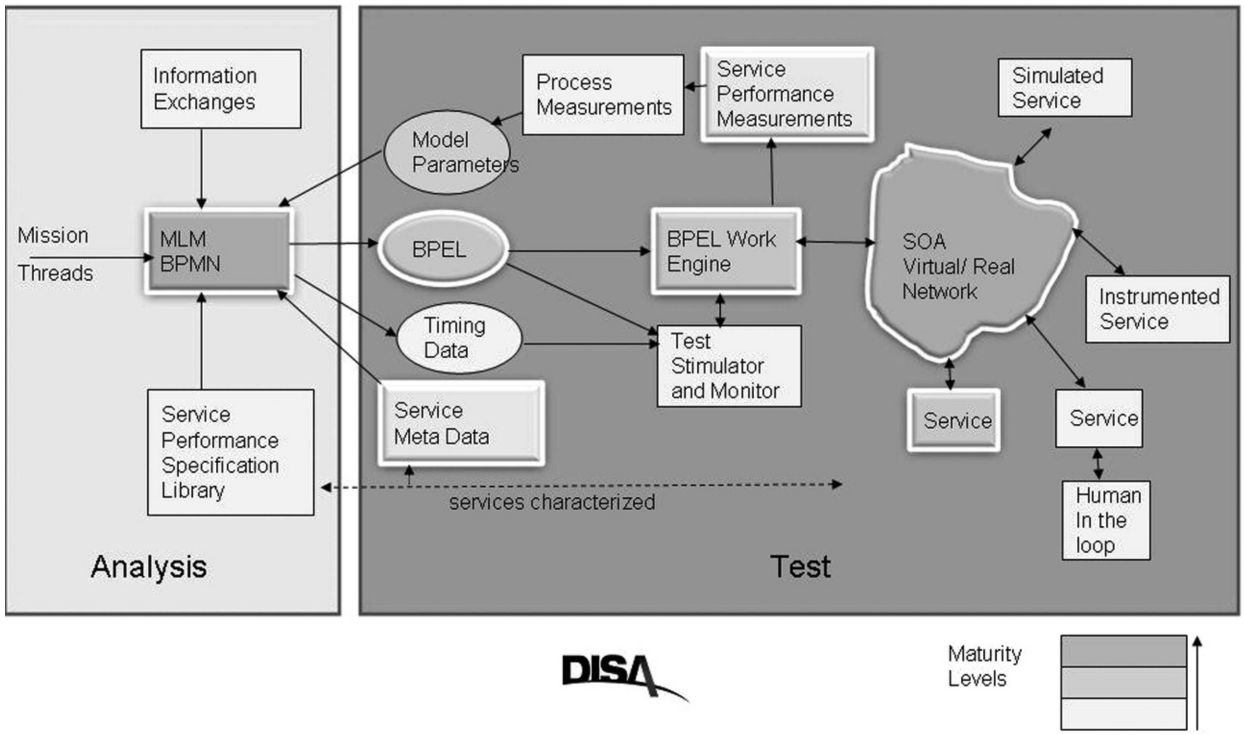


Figure 1. MLM2SOA process

The MLM2SOA process

The MLM2SOA process, as shown in *Figure 1*, semi-automates the iterative model-test-model methodology using modeling & simulation and the real test environment, thereby significantly improving T&E efficiency. It provides an early opportunity to analyze overall mission performance using M&S. The MLM2SOA environment integrates evolving COTS products supporting BPMN, BPEL, and SOA standards. Mission threads are modeled using BPMN standard and can be simulated and analyzed using COTS tools supported by over 40 vendors.

MLM2SOA process facilitates certification at dynamic interoperability level (Tolk and Maguira 2003) and complements syntax and semantics interoperability test supported with SOA COTS solutions. It facilitates T&E of concurrent and sequential dependencies among services under varying conditions.

The MLM2SOA process uses MLM to define, refine, and verify mission performance within a defined scope via simulation. MLM simulation projects scalable operational performances under variable conditions using initially estimated and later the measured parameter values. MLM captures the end-user workload and the end-user computing, communication, and storage environment. MLM uses a model library of validated and/or proposed SOA service performance specifications (SPS).

Mission activities can use alternative SOA services to achieve the desired mission performances, which can be analyzed using the COTS BPMN simulation capability. MLM parameters include human and SOA service times and can be updated from measurements from a real/virtual environment.

The process guides the scenario selection for real testing and automates test generation from MLM of system configuration, test stimulus and test monitoring. Using the COTS products, the mission thread with selected SOA services and their composition configuration can be exported in a standard BPEL format. BPEL model includes the service interface metadata in Web Service Definition Language (WSDL) format, which is initially imported in the BPMN model.

Timing data is exported in Excel format as BPEL standard does not support time representation. Timing data is used to create a stimulator controlling external events and to validate the monitored test results. The process automates the bi-directional integration of MLM with a real/virtual SOA environment via parameter values.

Using the commercial BPEL workflow engine such as supported by IBM and Oracle, the configuration can be deployed in a real or virtual SOA environment. Services in a test environment can be real or simulated. Further, services in the real environment can be instrumented via COTS tools to collect service level agreement parameters. These collected measurements can be processed to generate model parameters as an Excel file to MLM.

MLM2SOA process steps

The suggested process steps are:

1) Define mission scope by a
   a) Set of end-to-end mission threads identifying the activities performed via SOA services and humans,
   b) Library of SOA Service Performance Specification,
   c) Set of information exchange events among mission activities, and
   d) Set of mappings of instances of selected SOA services to mission activities. (Multiple mission activities can share a SOA service instance. Mission can use multiple instances of the same service.)
2) Capture OMT via mission level executable model using BPMN standard and SPS
3) Evaluate operational effectiveness and efficiencies by evaluating alternative information flow strategies and alternative activity/service mappings in a net-centric environment.
4) Select information flow strategy and activity/service mapping
5) Perform six sigma analysis for operational performances such as Net-Ready key performance parameters
6) Project expected operational performances for alternative infrastructure implementations
7) Select critical workload, infrastructure configurations to test in the real environment
8) Automatically generate BPEL workflow and needed timing data for real test
9) Develop stimulator and monitor based on the timing data
10) Develop software to extract service performance parameters from service instrumentations
11) Use BPEL workflow engine to set up test configuration by deploying actual and simulated services.
12) Test using simulated services in SOA environment.
13) Incrementally plug and play real and simulated services
14) Test mission with "human-in-the-loop" and collect model parameters from service instrumentations
15) Update the model and re-iterate

MLM2SOA process uses a persistent but logical test harness for testing composition of services as opposed to physical test harness harkened back in the T&E community years ago. The learning curve of the process can be reduced by recruiting M&S experts, who collaborate with test professionals to build an efficient "Model-Test-Model" paradigm. MLM2SOA process can be a part of the Test and Evaluation Master Plan (TEMP) and should use test event OMT to T&E the composition of a set of services.

MLM2SOA process benefits

The proposed MLM2SOA is a middle-out process that integrates top-down mission need refinement with bottom-up composition of SOA services. MLM2SOA process integrates constructive MLM and real/virtual SOA test environments. The MLM2SOA process:

1. Exploits the COTS support of converging BPMN, BPEL, and SOA standards.
2. Builds on commercial best practice of composing atomic processes and atomic services.
3. Reuses test scenarios, data, environment, and results across programs.
4. Uses semi-automated processes to significantly reduce time for changes to mission threads and/or implementation technologies common in DoD programs.
5. Promotes testing early and often.
6. Allows concurrent T&E at multiple interoperability levels: Syntax, Semantics, and Dynamics.

A MLM2SOA prototype

Shading in *Figure 2* indicates the COTS maturity of the current capabilities. Capability to model mission threads in BPMN, as indicated with dark shade, is currently available from multiple vendors under $500/user. The capabilities shown in lighter color boxes are currently supported by few vendors costing in the range of $20,000. Software represented in the lighter boxes is available as Business Process Management Suite from few vendors starting at $250,000. As shown in *Figure 2* above, MITRE is using the iGrafx tool (Pridemore 2006) for BPMN modeling, simulation, analysis, and BPEL generation. MITRE is using MINITAB for stochastic analysis and ActiveBPEL as the BPEL workflow engine.

Missile defense vignette

We illustrate the MLM2SOA process using a missile defense vignette which processes sensor data and displays tracks on Google Earth. The sensor data is stored in a file and is processed by multiple SOA services that were developed by MITRE and constructed using MATLAB. The SOA services include Associate Track, Propagate Track, and Save Track.

Sample track data is read from the file and sent to the services using a JAVA client application. The processed track data is then stored in a back-end database by the services. This stored track data is read asynchronously through additional service interfaces

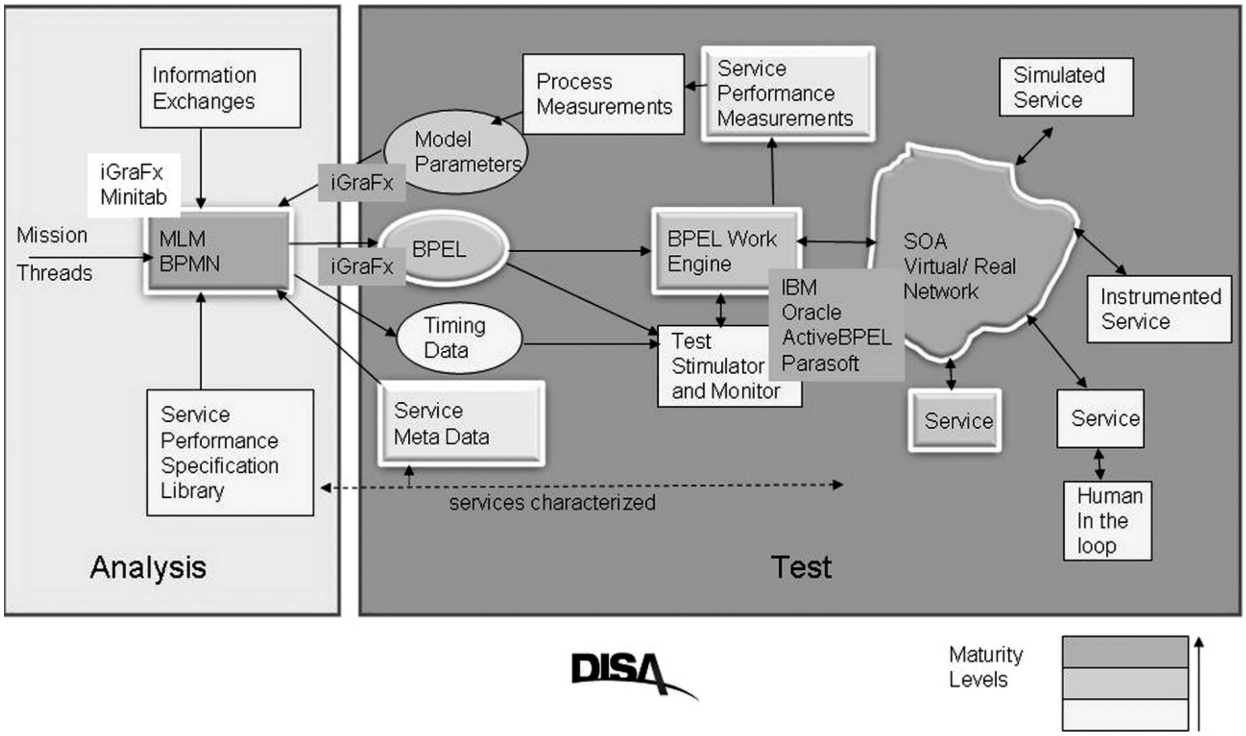


Figure 2. MLM2SOA process commercial-off-the-shelf

which return Keyhole Markup Language data and is displayed using Goggle Earth.

This sample mission thread allows two of the services, Associate Track and Propagate Track, to be invoked sequentially or concurrently. This concurrent composition of services can potentially reduce the end-to-end response times for overall mission performances.

Sequential composition of services

Figure 3 shows a BPMN model for sequential composition of services using the iGrafx tool. The BPMN diagram consists of three swim lanes, each representing an organization providing those services. The *sensor* swim lane generates the track data in this illustration. The *track* process swim lane extracts the

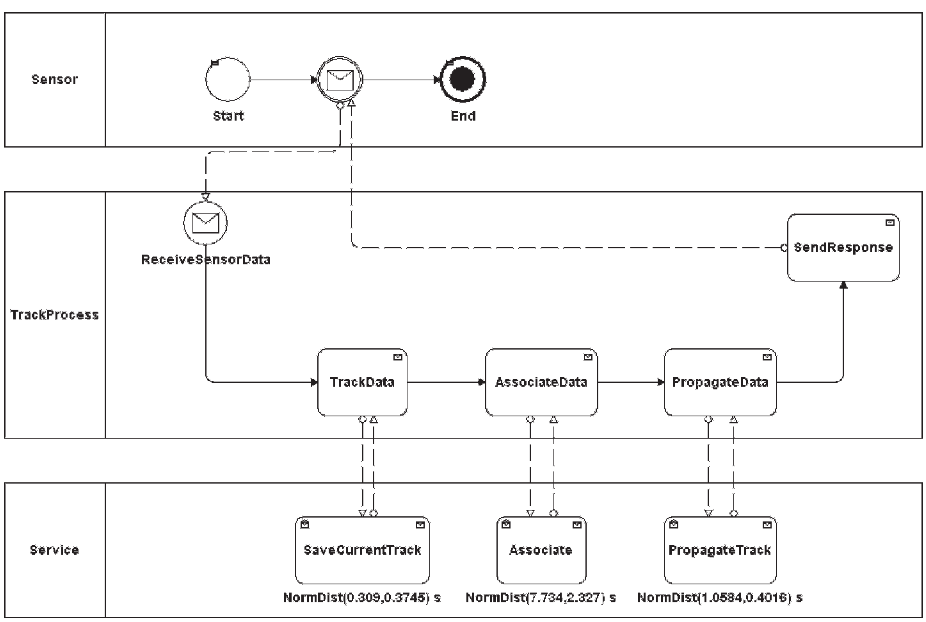


Figure 3. Business process modeling notation sequential service invocation

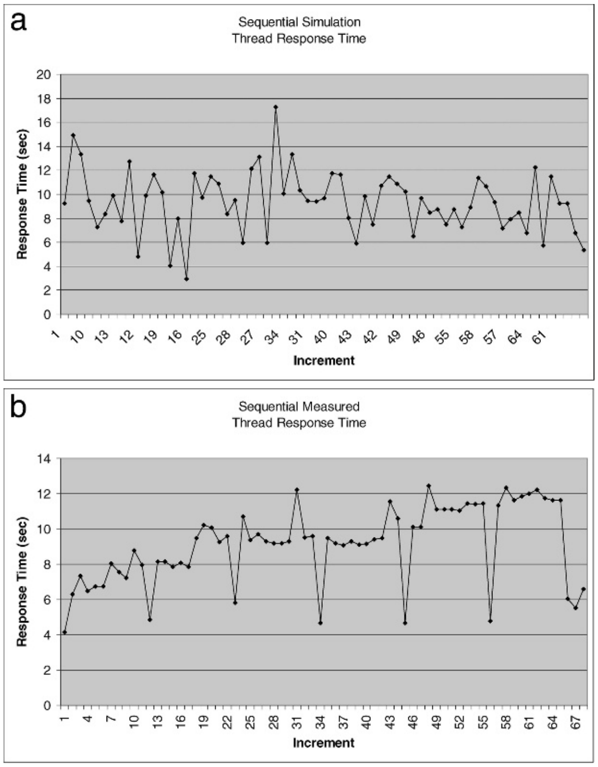


Figure 4. a) Sequential simulation results, b) Sequential measured results

data from the stored file and provides the information to the respective service being executed in the last (*service*) swim lane. BPMN modeling uses messages to communicate among swim lanes as shown via events crossing the swim lanes.

Each of the rectangular boxes represents an operational activity, which can be mapped to human resources or computing resources. In this illustration we map activities to only SOA services. Each of these services can take variable times, which are abstracted via BPMN activity attribute values.

Using the specifications and external events denoted by the start node in the sensor swim lane, mission performances can be projected. *Figure 4a* shows the simulation results for this illustration and *Figure 4b* shows the results from actual service invocations as orchestrated by our BPEL process.

Figure 5 shows ActiveBPEL's graphical representation of the exported BPEL. BPEL is an XML-based language and thus is exported by the iGrafx tool in a readable text file. COTS tools convert the BPEL text file in a proprietary graphical view. The figure shows how the exported BPEL is expressed visually by ActiveBPEL.

Figure 6 shows the configuration used for our demonstration. Servers at MITRE Colorado Springs

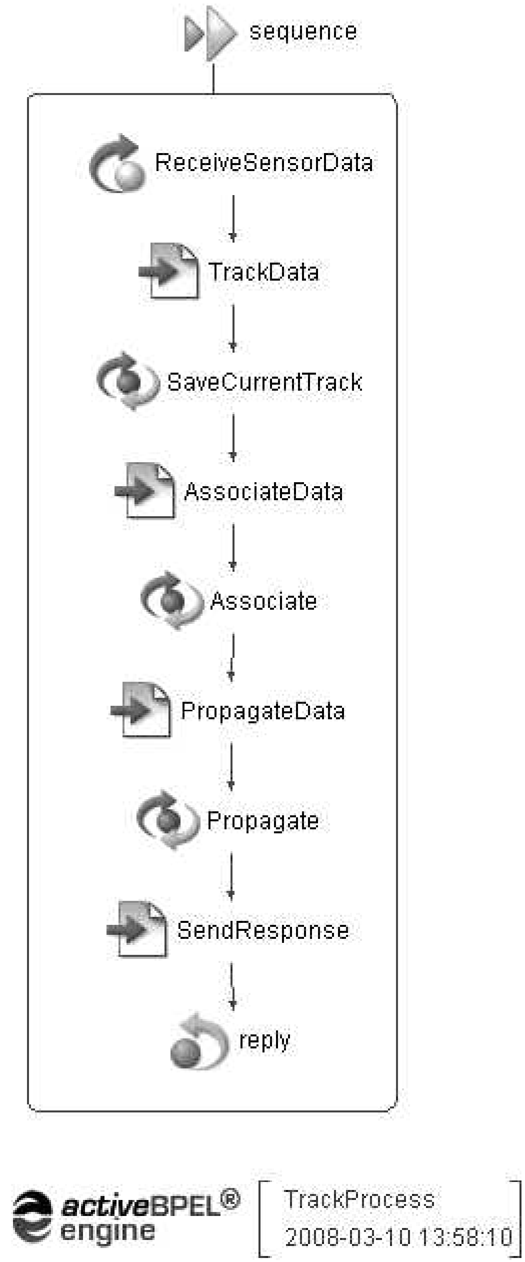


Figure 5. Business process executable language sequential service invocations

hosted the sensor processing services. MITRE Mclean has a PC running the test stimulus and monitoring and another PC to display tracks on Google Earth.

Concurrent composition of services

In contrast to the sequential service invocations shown above, *Figures 7* and *8* are diagrams depicting the concurrent invocation of services in BPMN and BPEL respectively.

Significant mission level performance improvements can be realized by taking advantage of the concurrency

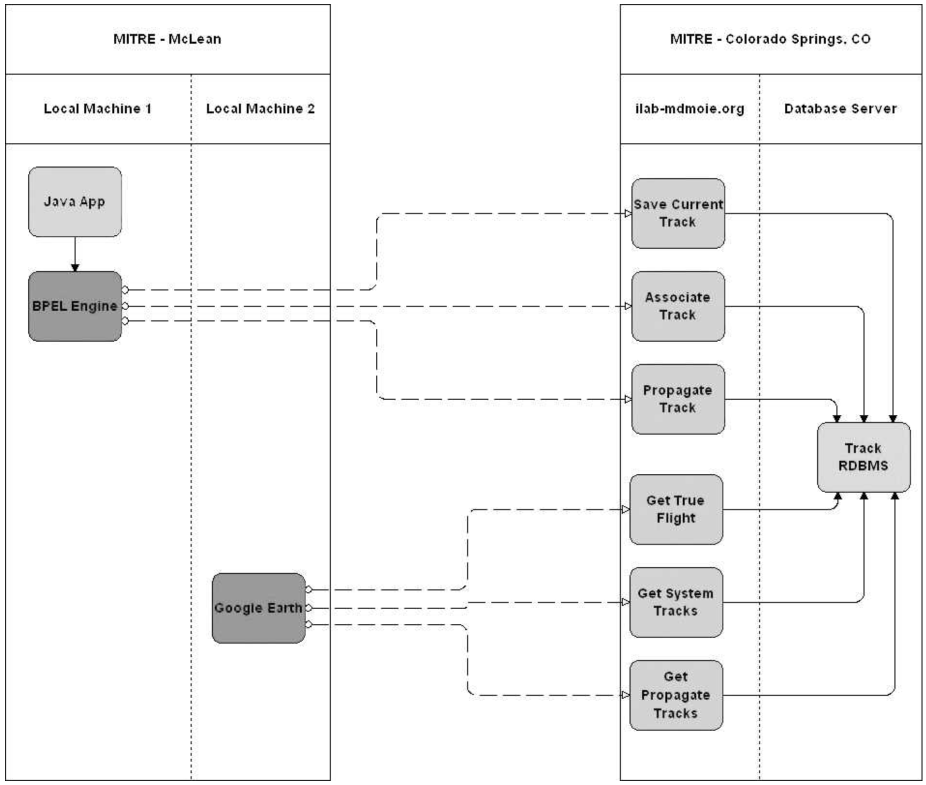


Figure 6. Business process modeling notation model of demo configuration

among process activities and assuring that infrastructure supports concurrent execution of services, mapped to these concurrent process activities. Although our example shows concurrency among process activities, a single computing resource was used to host the services, thus causing a serialization in execution. As such no significant improvements were realized as depicted in *Figures 9a and 9b*.

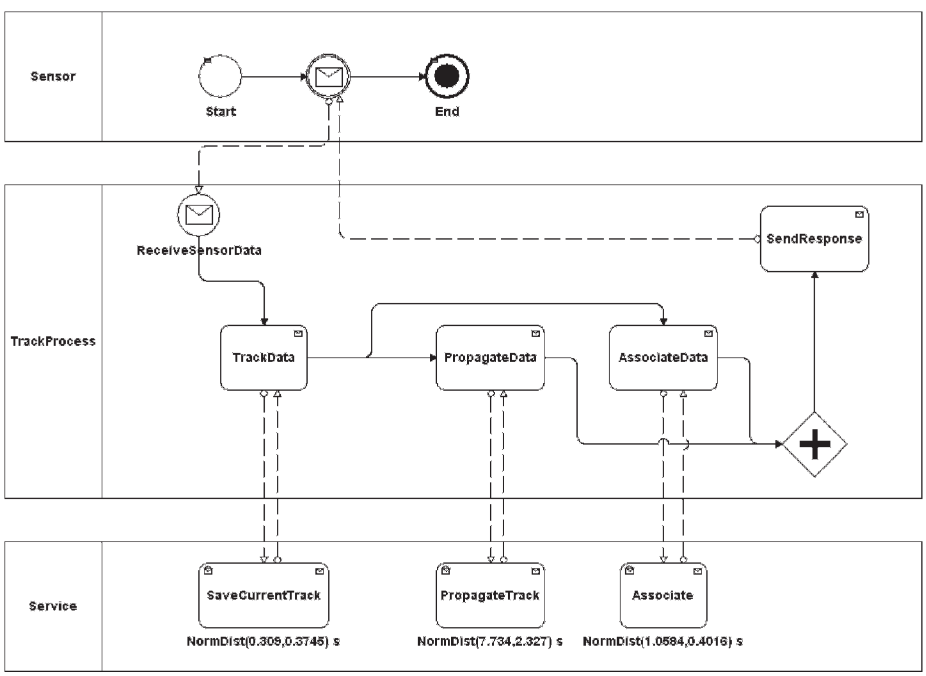


Figure 7. Business process modeling notation concurrent service invocations

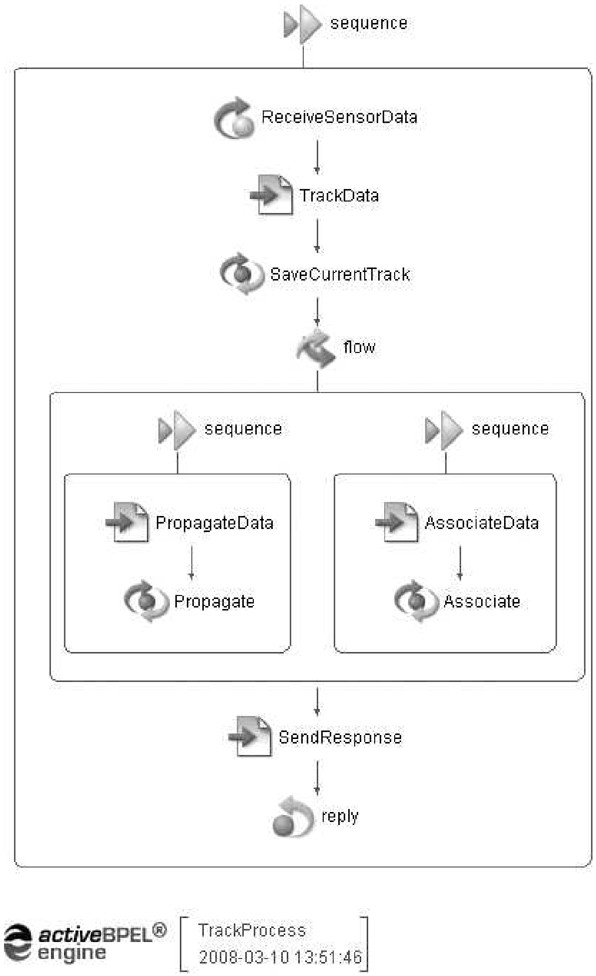


Figure 8. Business process executable language concurrent service invocations

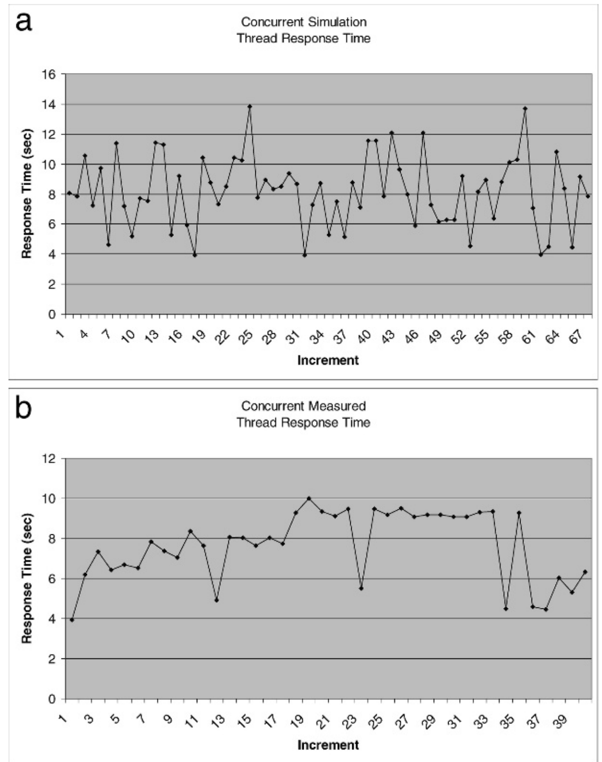


Figure 9. a) Concurrent simulation results, b) Concurrent measured results

MLM2SOA process efficiency

The automated generation of BPEL from BPMN requires a mapping of SOA services to BPMN activities. COTS tools require a mapping file which links the BPMN activities to services and maps BPMN message events to service interfaces defined in WSDLs. This mapping is completely independent of services execution order, e.g., switch from serial execution to concurrent execution of services, does not require a change to the mapping. This provides a separation between the mission thread flow and the activity implementation. This approach is highly desirable for DoD acquisition to provide flexibility for independent and simultaneous evolution of mission threads and SOA services.

However, incorporating new SOA services into the BPMN model is the most time-consuming part of the semi-automated MLM2SOA process. After one month initial learning curve of the COTS tool, initial BPEL representation of the simple BPMN model took only a week. The value of the MLM2SOA process was observed in the subsequent changes for creating new mission threads composing the same SOA services. For example, it took only a day to go from sequential composition to concurrent composition of SOA services. After the initial generation of BPEL, no additional changes to the mapping file were necessary to model new patterns of service invocation. Only the data transformations needed to be updated due to the change in the order of service invocations.

Table 1 captures the development times for various tasks for the simple missile defense vignette example using a few services. It used sequential composition for the initial implementation and concurrent composition for subsequent implementation

In summary, COTS support for automated BPMN to BPEL generation is in the early stages of maturity.

Table 1. Approximate business process modeling notation to business process executable language development times for MLM2SOA prototype

Development activity	Time
Commercial-off-the-shelf learning	20 days
Develop mission level modeling and analyze performance	1 day
Generate initial business process executable language	5 days
Iterate subsequent service composition	1 day

We observed the following restrictions using iGrafx and ActiveBPEL:

1. Not all valid BPMN flows can be translated into the equivalent BPEL.
2. Mapping file may need to be changed to align BPMN attributes to BPEL data types.
3. BPEL uses XPATH expression but BPMN simulation languages are vendor proprietary. Hence the expressions must be converted manually after the BPEL is generated.

ActiveBPEL Designer allows the BPEL process to be simulated and debugged and it allows for manual edits to be validated. Data transformations validation by viewing the data values assigned to variables as one step through the BPEL process is enabled by the simulation capability when in simulation mode. Although actual services are not invoked during simulation mode, output data can be generated based on the service return data types as specified in the WSDL. This output data can then be modified and/or passed to other services as the BPEL process dictates. Service input data types are also validated.

Debug mode was used during the actual running of a BPEL process. Running in debug mode allowed us to step through the execution while inspecting data values returned from the actual service invocations. It also supports manually changing data values at any step in the process to modify its behavior.

1. Potential to Improve T&E Efficiency

MLM2SOA process integrates mission level modeling with SOA environments consisting of real, instrumented and simulated services. The process implemented on evolving COTS solutions provides DoD an excellent opportunity to improve T&E efficiency. Once mission threads requirements are locked in, measures such as MOP, MOE, and KPP documented, and test infrastructure is established, the MLM2SOA process enables fielding of net-centric capabilities in weeks and months with high confidence.

Faster

1. Reduce response time to mission and implementation changes by automating the integration of MLM with real environment
2. Semi-automatically create operational test event configuration, stimulus and monitoring environment
3. Enable to test KPP and service integration early
4. Debug hard timing problems using MLM, instrumented, and simulated services
5. Concurrently test at syntax, semantic, and dynamic levels

Better

1. Operationally driven test environment
2. Gain in test confidence from complementary data generated from MLM and operational events
3. Use of instrumented service measured data from real environment in MLM

Cheaper

1. Improved collaboration in developing operational mission threads using MLM graphical environment
2. Reducing dependence on big bang test by early verification using MLM and simulated services in MLM and SOA environment
3. Reuse test assets across programs

Persistence SOA test environment configurable by allocating different services to computational nodes provides a high degree of reuse for SOA real test. Mission threads captured in standard BPMN can be reused in subsequent MLM development.

2. MLM2SOA Potential

The proposed process has potential beyond what is discussed in this paper and is summarized below:

1. The proposed process has the potential to influence the total life cycle of capability-based DoD 5002 acquisition process. The MLM if developed at JCIDS level will streamline the** overall acquisition process from JCIDS to OT. MLM used in Joint Combat Capability Developer process provides an opportunity to understand the mission behavior via simulation and hence can help in defining the performance metrics such as KPP, MOP, and MOE. MLM at this level can also guide the development of a test plan identifying the needed test event OMT for improved test efficiency. MLM also enables to verify and assess mission effectiveness for new information flow strategic operational concepts in a constrained infrastructure environment via simulation.
2. Graph theory analysis of MLM can provide mission thread coverage and service interoperability coverage. This analysis can provide the needed data to program managers to assess test cost versus its benefits and needed data for risk mitigation analysis.
3. MLM can be analyzed to identify and verify the specification of gateways for cross security domain architecture.
4. T&E process in commercial industry for SOA-based systems is developed by each project and is not repeatable. DoD has an opportunity to provide the leadership to the commercial industry to define a repeatable T&E process. COTS solutions implementing the process can reduce DoD T&E cost.

5. MLM can provide insight into the design constraints, which must be imposed to ease distributed test and debugging and to improve T&E confidence in designs evolving via compose-ability of validated designs.

We propose that the MLM2SOA process initially be used by the T&E community to test composition of services to meet a new operational need specified as a Test Event OMT. This capability can later evolve to the JCID process to set the joint program requirements. Test Event OMT can then be derived from the MLM developed during the JCID process. ❑

DR. PREM JAIN is leading the development of a new "Mission Level Modeling (MLM) and Analysis" capability at MITRE Corporation. He supports the Joint Interoperability Test Command (JITC) to research and develop new methods and technologies needed to support GIG Test and Evaluation (T&E). He joined MITRE in October 2005 from BAE Systems "Center for Transformation" experimenting with computer-aided analysis technologies for Family of Systems. He has spent over 20 years in the Electronic Design Automation industry and was the CEO of "Cynergy System Design/CAE Plus", which developed "system-on-a-chip" architecture analysis and architectural synthesis tools. His patents include graphical executable language, architecture synthesis algorithms, and synchronous time sliced simulation. During 1988–1992, he was a faculty member of the computer science department of the University of Texas in Austin. He received his Bachelor of technology degree in electrical engineering from the Indian Institute of Technology, Kanpur, India. His Masters degree and Ph.D. are in electrical engineering from Rensselaer Polytechnic Institute, Troy, New York. E-mail: pjain@mitre.org

BEN KING is a lead simulation modeling engineer in the Advanced Architecture Technologies and Solutions department at the MITRE Corporation. He has supported DISA, USD(I), National Intelligence community, DHS and U.S. Army for performance engineering, security engineering, and systems engineering. He is leading the prototype development of the MLM2SOA process. King joined MITRE in 1990, and has spent the last 18 years supporting/developing DOD training simulations and supporting technologies and analyzing performance of large DOD information technology systems. He received a bachelors of science degree in electrical engineering and a bachelors of science degree in computer science from the University of South Alabama, Mobile, Alabama, and a masters degree in computer and information sciences from George Mason University, Fairfax, Virginia.

References

DoD Instruction 5000. 2. 2003. "Operation of the Defense Acquisition System," May 12, 2003. Available at http://akss.dau.mil/dag/DOD5002/PDFs/DOD1%205000.2signed%20(may%2012%202003).pdf.

OASIS (Organization for the Advancement of Structured Information Standards), 2007. Web Services Business Process Execution Language (WSBPEL) Version 2.0. OASIS Standard, April 11, 2007. Available at http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-os.pdf. Accessed May 7, 2007.

OMG (Object Management Group), 2006. Business Process Modeling Notation (BPMN) Specification Version 1.1. Available at http://www.omg.org/spec/BPMN/1.1/PDF.

Pridemore, B. and Jain, P. 2007. "MITRE Report Net Centric Mission Level Modeling & Analysis Experimentation" March 2007. Abstract 5519. Available at http://proceedings.nda.org/8870/5519.pdf.

Tolk, A. and Maguira, J. 2003. "The Levels of Conceptual Interoperability Model (LCIM)." Fall Interoperability workshop.

ITEA Journal 2008; 29: 190–200

Assessing Risk Levels of Verification, Validation, and Accreditation of Models and Simulations

James N. Elele, Ph.D
U.S. Naval Air Warfare Center (NAVAIR), Patuxent River, Maryland

This article presents a new methodology for assessing the risks associated with the level of verification, validation, and accreditation (VV&A) of a given model and/or simulation when used in support of major decisions. As stated by Department of Defense (DoD) Instruction 5000.61 (DOD, 2003),[1] "It is DoD policy that: models and simulations (M&S) used to support major DoD decision-making organizations and processes shall be accredited for that specific purpose by the DoD component M&S application sponsor." Instruction 5000.61 applies to "All models and simulations developed, used, or managed by the DoD components after the effective date of this instruction." The requirements cited above have set the need for VV&A of M&S at the forefront of concerns for DoD and DoD-Component acquisition personnel. When an acquisition program involves a large number of models, cost associated with VV&A can become enormous. There is a need therefore to have a systematic approach for assessing and prioritizing the risks associated with the level to which individual models have been verified, validated, and accredited. To provide decision-makers with a judicious way for determining the risks associated with using a given M&S, and the extent to which VV&A work will be needed to meet these requirements, a methodology is developed for assessing the risks associated with the level of VV&A of a given M&S when used to support decision-making. This approach parallels the formal DoD Risk Assessment procedure, but with application to the use of M&S, as it relates to VV&A.

Key words: Accreditation; consequence of error; likelihood of error; modeling & simulation credibility; risk assessment; validation; verification.

Modeling & Simulation (M&S) have become an integral part of the defense acquisition process in the United States. The value of M&S in various aspects of the defense acquisition process will continue to increase as advancement in computer technology and new modeling techniques emerge. Since resources are limited, there is a need for a logical way to guide and prioritize the investment in M&S while making sure that models are credible, particularly when such M&S are used to support very important decisions. Moreover, M&S saves money by reducing the cost associated with system design, development, and testing. But, in addition to cost savings, there is a less emphasized but very valuable role of M&S in military acquisitions: M&S allow system developers to reduce and/or avoid risks associated with human involvement when investigating the boundaries of applicability/survivability of systems. With M&S playing such critical roles in war fighter systems development, there is a critical need to devise verification, validation, and accreditation (VV&A) approaches that use available resources efficiently without compromising the needed credibility of the M&S.

Formal VV&A can be very expensive. How much (and what kind of) evidence is required for establishing confidence in, and reaching an accreditation decision for, a particular M&S determines the amount of resources required. For existing M&S (e.g., legacy M&S that have been used in the past without formal VV&A), current methodologies provide the users with little, if any, way to assess the risks associated with accepting and using such a model or a simulation when its credibility is in question; and the latter is often the case even though the model may have a history of extensive use. What is usually done is to accept and use the M&S with or without VV&A. But when the model result deviates seriously from what is expected, the user may have to start from scratch and reengineer the model development process in order to generate the data and artifacts needed for establishing credibility.

The methodology we present here uses information from various aspects of the development of the M&S — its VV&A history, knowledge from subject matter experts (SMEs), the role the M&S will play in the decision-making, and the consequence if the M&S is wrong — to arrive at a categorized risk level (high, medium, or low risk) expected with accepting such an M&S for an intended use. Determining this risk allows the user to choose to mitigate the identified risk by conducting formal VV&A, or to accept the risk, accredit the M&S with limitations, and continue to use it as is. This approach becomes very useful when a large number of models and simulations are needed to support an acquisition program, so there is a need to prioritize and allocate M&S-related funds and resources efficiently.

The issue is credibility of M&S for an intended use

The goal of VV&A is to generate, maintain, gather, and apply information about a given M&S to support the decision to use that model. The problem has been that VV&A was frequently discussed rather than practiced; and when implemented, it was usually done as an afterthought. One thing no one argues against is that it is very important for M&S to be credible and suitable for the specific intended use. But credibility and appropriateness of M&S for intended use are best established through the VV&A process. Thus M&S users need to recognize that VV&A is necessary for risk reduction and critical for establishing credibility. Because VV&A is frequently tailored to specific needs, a formal definition of VV&A is presented. The definitions given below are congruent to those of DoD, Navy, and other Services; and are particularly suited for practical applications and the pedagogical purposes we pursue here.

- **Verification:** Verification is the process of determining that a model implementation and its associated data accurately represent the developer's conceptual description and specifications. The practical question answered by verification is "Is the model relatively error free, and does it do what the originator intended?"
- **Validation:** Validation is the process of determining the degree to which a model and its associated data are an accurate representation of the real world from the perspective of the intended uses of the model. Note that M&S validation is not the same as software validation. The practical question answered by verification is "Do model results match real world data well enough for the user's needs?"
- **Accreditation:** Accreditation is the official determination and certification that a model, simulation, or a federation of models and simulations and its associated data are acceptable for use for a specific purpose. The practical questions answered by accreditation are "Does the accreditation authority have adequate documented evidence to be confident that a model and its input data are credible for a particular use; and is there enough documented information to show that this M&S is fit for this purpose?"

It is reemphasized here that VV&A, in principle, is a process for reducing risk; in that sense VV&A provides a way for establishing whether a particular M&S and its input data are suitable and credible for a particular use. According to *M&S VV&A Implementation Handbook*, Volume 1,[2] risk management as related to VV&A is simply answering the question "What is the risk of using a model compared to using the real world system or methodology?" VV&A provides the answer to this question through processes and procedures designed to mitigate the risks while providing objective evidence as needed to establish the credibility of the M&S.

Statistical basis for risk-based VV&A

Risk-based VV&A is based on the statistical principle of hypothesis testing. Given any M&S, VV&A has become the standard accepted way to answer the question "Is this M&S credible for the intended use?" In statistical language, this problem is stated as testing the "Null Hypothesis," which asserts **Ho:** ***The M&S is credible for the intended use.*** (One usually must also state an "Alternate Hypothesis"; e.g., **Ha:** ***The M&S is not credible for the intended use.***) To complete such a test, one would set up an experiment, collect data, and test the above hypotheses for acceptance or rejection. The data so collected is considered as sample points taken from a known or assumed probability distribution for the events occurring in the experiment. Since this type of testing is based on sampled data, there is always the possibility of making errors; and the two categories of possible errors that result are, namely, Type I and type II errors:

- A **Type I error** denoted by α; occurs when one rejects H_0 when it is true.
- A **Type II error** denoted by β; occurs when one accepts H_0 when it is false.

The individual conducting the test must then specify the maximum allowable probability of making a Type I error (α, called the level of significance). The problem is that, in general, the experimenter is not able to control the probability of making a Type II error. From the M&S perspective, making Type II errors is equivalent to accepting that M&S is credible for the intended use when it is not. Because one is not able to control this type of error in the experiment, there is not much one can do (in theory) to control the probability of Type II error where M&S is concerned. In practice, however, the acceptance or rejection of a model for an intended use is only done through formal VV&A assessment of the M&S under consideration.

Typically, statisticians try to avoid the risk of making a Type II error by using an expression such as "**cannot reject** H_0" based on the evidence/data, and by actually avoiding to say, "**accept** H_0" whenever possible An alternative approach uses the sampled data to compute a statistic called a ***P* value**, (which is the probability of obtaining a sample result that is at least as unlikely as the data that was observed). If the ***P* value** is less than the level of significance α, the tester rejects the null hypothesis, and accepts it otherwise.

It is the above statistical test structure that gives rise to the risk associated with using M&S and actually forms the theoretical basis for the risk-based VV&A approach.

Obviously, once a decision to use M&S has been made, one automatically becomes subjected to making these types of errors. Type I error is made when a user rejects a model or simulation that is credible for the intended use. But making Type I error is not considered very critical in real life (though this type of error can become very important from a cost-savings perspective). This fact is because the main loss for a user who makes an M&S-related Type I error (i.e. rejecting to use an M&S as not credible when it really is) is opportunity lost/cost. However, if a user decides to apply M&S, that user has accepted the chance of committing Type II error, and the consequences of such an error can be substantial. Thus it is emphasized that the principal risk associated with using M&S in decision-making is defined by Type II error. Moreover, the likelihood of making such an error and the associated consequence determine the risk associated with using that M&S in decision-making. VV&A therefore provide the accepted practical method for reducing the risk associated with using the M&S and establishing confidence in the M&S. The consequences of the model being wrong, and the level of risk one can accept, drive the amount of effort required to establish an acceptable credibility level for the M&S.

Definition of risk in the risk-based VV&A paradigm

The Defense Modeling and Simulation Office (DMSO) defines *risk* as "the potential realization of undesirable consequences from hazards arising for a possible event".[3] Of course, risk is a factor in any kind of decision-making in which imperfect evidence is used to help make the decision. Though VV&A is oftentimes more concerned with operational risk (i.e., the risk associated with using the M&S), the approach presented here can also be used, and has been used, to assess the risks associated with M&S development (i.e., the risk associated with completing the M&S development on time and within budget). This practical approach allows the System Safety community to define risk as the product of the likelihood of error and the consequences associated with such an error.

DEFINITION: Risk = (Likelihood of Error) * (Consequence of Error)

In this definition, ***likelihood of error*** is the probability that the M&S and/or its input data are incorrect or inappropriate for the intended use. The ***consequence*** is defined as the impact if the M&S output is wrong and the user accepts and uses it as correct. Thus to reduce risk, one either has to reduce the likelihood that something will go wrong or reduce the severity of the consequence(s) or impact/effect that will result when something goes wrong.

The consequence or impact when a model is wrong depends on the role the M&S will play and how important this role is in the decision-making process. One can choose not to use M&S (for example, by choosing to only accept actual physical measurements in decision-making), thus avoiding the risk associated with M&S use. Alternatively, one can choose to reduce the risk associated with using M&S by limiting the role M&S will play in the decision process. However, the reality is that DoD, the Navy, and other Services have mandated the use of M&S (ubiquitously) in the acquisition process, so that what is really needed is a method that will give M&S users the ability to assess the risks associated with the use of M&S, while providing them with the capability to mitigate such risks. This is what the current method tries to do.

Risk associated with M&S may be viewed differently from different managerial points of view. For example, an M&S developer/program manager may be more interested in the risk associated with the delivery of a simulation on time (schedule) and on budget (a form of development risk). Verification and Validation (V&V) is concerned with mitigating development risk by collecting objective evidence as required to demonstrate the capability, relative freedom from defects, level of fidelity, and the accuracy of representation of reality. From this perspective, part of what V&V is concerned with is providing the artifacts needed for accreditation. Providing (or requiring more) evidence needed to certify usability, limitations, and fitness for intended use is the goal of accreditation. Consequently, accreditation is more concerned with exposing operational risk associated with M&S use.

Problem

As precise as the above definition of risk is, it has problems. To determine the risk, one must be able to determine both the likelihood of error and its consequence precisely, and then be able to multiply these two values to compute the value of the risk. But for M&S, one is not usually able to determine the likelihood of error precisely. Even when one is able to assess the consequence(s) of such an error, the data required to determine the likelihood of error precisely is usually lacking (recall that statisticians do not have the ability to control Type II errors). How then can one multiply something not determined? One cannot!

What is done is to classify each of these parameters (risk, likelihood, and consequence) into discrete scales or levels sometimes called baskets or bins: e.g., high, medium, and low for risk; frequent, probable, occasional, remote, and improbable for likelihood; and some form of numerical rankings for consequence. Doing this allows the user to apply the above scales in a systematic way, in an attempt to get a handle on the risks associated with using M&S to support decision-making. Understanding how this works forms the practical basis for the risk-based VV&A procedure, and this approach also makes it possible for an M&S user to determine how much effort to put into VV&A, the appropriate and extent of the review process, the level of independence in V&V review, as well as the appropriate level of accreditation authority. More details are provided on how this methodology can be used to support the accreditation process in a companion article.

In practice

We now turn attention to the practical applications, "The How To" of this method. This is necessary to allow more users to use this method, and because the more people there are who understand and use this methodology the better it will become. The method is broad enough that it can be adapted for specific applications. To start, some of the specific tools needed to make the method work are identified.

Where to begin

When faced with a VV&A question about a given model, the ease or difficulty of the validation process depends on whether a version of the M&S system currently exists or if it is being built for the first time. Designing the VV&A processes into a new model while it is being built is the better and preferred approach. In any case, the first task is usually to determine the scope of the effort necessary to accomplish the needed V&V, and the attendant accreditation for the specific intended use.

This problem requires that a cost-effective VV&A plan be devised. Such a plan needs to include a well-articulated Intended Use Statement (IUS). The IUS will help in defining the role the M&S will play in the decision process. The IUS may have already been written as part of the M&S requirement document; however, the role of the M&S in decision-making may best be determined through the knowledge of such things as available technology, planned activities, cost and schedule, and program management priorities. This requires one to answer such questions as "Will M&S be the only tool that will be used to generate the data needed for a major decision, or will other activities such as laboratory/actual test and evaluation (T&E) data be used with data from M&S to support the decision?" At a minimum, one must answer the question, "Will this M&S be used to support a minor, medium, or major acquisition decision?"

Sample scales and tables

This procedure will require the implementer to develop or adapt/adopt various scales and tables for use as classification buckets. The Battlespace Verification, Validation, and Accreditation Support Office (BVVASO)/Joint Accreditation Support Activity (JASA) has developed a series of tables and scales based on years of practical implementation of VV&A procedures for various acquisition programs (BVVASO, October, 1997).[4] Our experience from working with acquisition programs indicates that some of these tables would be invaluable, at least, as a starting point for practical implementation of the methodology. We provide a list of the tables and scales that a user might need when implementing the method in subsequent sections.

Table 1 gives a sample confidence/likelihood scale that is based on BVVASO's experience and guidelines in *DMSO VV&A Recommended Practice Guide (RPG)*.[5] It is recommended that one level be included in this table for either low or unknown level of confidence to allow for a minimal effort option, and to cover emergency or low consequence situations.

Table 2 provides (from BVVASO experience) a possible way to scale likelihood of error and/or confidence level. The table summarizes what has been found to be the information necessary to support an accreditation assessment.

- The higher the likelihood of error (or the lower the confidence in the M&S), the more need there

is for rigorous oversight and review through documentation, code and logic verification, configuration management, review of model development approach, data and code validation, to drive down likelihood of error.
- As likelihood of error goes down, confidence in model results goes up. Using the table below, one is able to assess the likelihood of error in an existing model or even determine what further evidence or level of accreditation is needed.

Table 2 is based on BVVASO's rules which have been adopted by the *DMSO VV&A RPG*[5]. See "Role of Accreditation Agent in VV&A of Legacy Simulations" for more details (available at www.vva.dmso.mil). *Table 3* is an example scheme quantifying likelihood.

Table 4 gives a sample scale for identifying and categorizing the importance of the role M&S will play in the decision-making. For example if M&S will be the only tool for making a decision, as is often the case in survivability studies, then the impact of this decision being wrong can be catastrophic, hence it is important to consider the risk associated with the level of VV&A of any such M&S.

Table 5 presents a scale for categorizing the level of importance of a decision based on the intended use of any model or simulation. This table provides a scale for accounting for the risk associated with errors resulting from M&S influence on such decisions. *Table 6* gives an example of the level of consequence on performance (cost, schedule, and technical quality) if the decision is a poor one. Criticality measure is determined from the level of reliance on M&S and importance of the decision. Criticality measure drives the nature and amount of information and effort applied to VV&A of this model *(Figure 1)*.

Table 7 gives a very simplified example of consequence scale with four broadly defined levels. Typically, consequence scales are defined with five levels for ease of application. This is because the standard risk chart has normally been defined with five levels of consequence. In practice, the choice of levels may best be determined by the circumstance being addressed by the user.

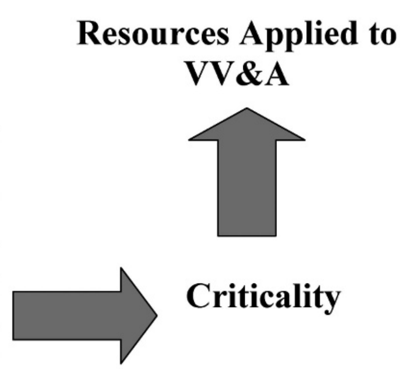
Table 8 gives a more advanced example of consequence scale with four broadly defined levels (Catastrophic, Critical, Marginal, and Negligible). Typically, Consequence scales are defined with five levels for ease of application. This is because the standard risk chart has normally been defined with five levels of consequence. In practice, the choice of levels may best be determined by the circumstance being addressed by the user.

Rules for associating/combining scales and tables

Once the scales and tables have been assembled, the user will also need to use combination/association rules to determine the risk associated with M&S. For example, the following rules were used in practical applications:

- **Rule 1:** **Value of "role (or level of reliance) of M&S (gray color)" in decision-making & value of "the importance of decision (gray color)" → "consequence of model error (green color)."** Determine the level of consequence if the model is in error by associating the role of M&S in decision-making and the level of importance of decision.
- **Rule 2:** **Value of the "level of reliance (gray color)" & the "confidence level (gray color)" → "likelihood of model error (gray color)."** Then

Importance of Decision	Level of Reliance on M&S			
	4	3	2	1
4	4	4 or 3	3 or 2	2
3	3	3	2	2 or 1
2	2	2	2	1
1	1	1	1	1



Source: DD(X) Verification, Validation, and Accreditation Overview by Charles Hays of Northrop Grumman Corporation. Presented at NMSO V V & A TWG. Salt Lake City, UT, February 16, 2005.

Figure 1. DD(X) criticality measure

Table 1. Levels of confidence/likelihood of M&S error

Likelihood of error	Confidence level	Description
1	4	Very high confidence based upon extensive documented V&V relevant to intended use.
2	3	High confidence based on face validation by SMEs.
3	2	Moderate confidence based upon previous usage history.
4 (high)	1	Low or unknown level of confidence. M&S appears to have the functionality required but credibility is unknown.

M&S, models & simulations; SME, subject matter experts; V&V, verification & validation.

Table 2. Evidence required to support likelihood of error and accreditation requirements

Likelihood of error	Confidence level	Description
1	4 (high)	Level 3 + Extensive body of documented verification and validation and extensive disciplined M&S development including history of technical and managerial review over time.
2	3	Level 2 + SME face validation relevant to current intended use + Evidence of effective configuration management.
3	2	Level 1 + Usage history + Known V&V history
4 (high)	1	Comparison of M&S requirement derived from intended use with capabilities and limitations of candidate simulation.

M&S, models & simulations; SME, subject matter experts; V&V, verification & validation.

Table 3. An example scheme for "quantifying" likelihood

Likelihood description	Likelihood of occurrence over lifetime of an item	Likelihood of occurrence per number of items
Frequent	Likely to occur frequently	Widely experienced
Probable	Will occur several times in life of item	Will occur frequently
Occasional	Likely to occur some time in life of item	Will occur several times
Remote	Unlikely but possible to occur in life of item	Unlikely but can reasonably be expected to occur
Improbable	So unlikely, it can be assumed occurrence may not be experienced	Unlikely to occur, but possible

Table 4. Critical analysis: importance of the role of M&S

Role level	Definition
4	M&S will be the **only method** employed to make a decision.
3	M&S will be the **primary method,** employed with other non-M&S methods.
2	M&S will be a **secondary method,** employed with other non-M&S methods, and will **provided significant data unavailable through other means.**
1	M&S will be a **supplemental method**, employed with other non-M&S methods, and will **provide supplemental data already available through other means.**

M&S, models & simulations.

Table 5. Critical analysis: importance of decisions

Level	Description
4	Intended use addresses **multiple areas** of significant program risk, key program reviews and test events, key system performance analysis, primary test objectives and test article design, system requirements definition, and/or high software criticality; used to make a technical or managerial decision.
3	Intended use addresses an **area of significant program risk** .
2	Intended use addresses **medium or low program risk**, other program reviews and test events, secondary test objectives and test article design, other system requirements and system performance analysis, and medium or low S/W criticality used to make technical or managerial decisions.
1	Intended use addresses **program objectives or analysis that is not a significant factor** in the technical or managerial decision-making process.

Table 6. Levels of consequences on (cost, schedule & technical) performance if decision is poor

Level	Technical performance	Schedule	Cost
5	Severe degradation in technical performance; cannot meet KPP or key technical/supportability threshold; will jeopardize program success; no workarounds.	Cannot meet key program milestones.	Exceeds APBA threshold.
		Slip ≥months.	>10% of budget.
4	Significant degradation in technical performance or major shortfall in supportability; may jeopardize program success; workarounds may not be available or may have negative consequences.	Program critical path affected, all schedule float associated with key milestone exhausted.	Budget increase or unit production cost increases.
		Slip ≤months	< 10% of budget.
3	Moderate reduction in technical performance or supportability with limited impact on program objectives; workarounds available.	Minor schedule slip, no impact to key milestones.	Budget increase or unit production cost increases.
		Slip <month(s) of critical path. Subsystem slip ≥months.	<5% of budget.
2	Minor reduction in technical performance or supportability; can be tolerated with little or no impact on program; same approach retained.	Additional activities required; able to meet key dates.	Budget increase or unit production cost increases.
		Slip ≤month(s).	<1% of budget.
1	Minimal or no impact.	Minimal or no impact.	Minimal or no impact.

Table 7. Simplified levels of consequence

Consequence level	Definition
High	Major disruption to program. Different approach required. Priority management attention and resource allocation required immediately.
Moderately high	Significant disruption to program. Different approach required. Priority management attention required.
Moderate	Noticeable disruption. Different approach may be required. Additional management attention may be needed.
Low	Minimal impact. Minimum oversight needed to ensure risk remains low.

Table 8. A more advanced scheme for ''quantifying'' consequence (impact) of poor decision

	Impact level			
Impact categories	Catastrophic	Critical	Marginal	Negligible
Personnel safety	Death	Severe injury	Minor injury	<Minor injury
Equipment safety	Major equipment loss broad-scale major damage	Small-scale major damage	Broad-scale minor damage	Small-scale minor damage
Environmental damage	Severe (Chernobyl)	Major (Love Canal)	Minor	Some trivial
Occupational illness	Severe & broad	Severe or broad	Minor and small scale	Minor or small scale
Cost	Loss of program funds: 100% cost growth	Funds reduction: 50% to 100% cost growth	20% to 50% cost growth	<20% cost growth
Schedule	Slip reduces DoD capabilities	Slip causes cost impact	Slip causes internal turmoil	Republish schedules
Political	National or international (Watergate)	Significant (Tailhook)	Embarrassment ($200 hammer)	Local
Operational	Widespread additional combat deaths	Limited additional combat deaths	Moderate additional casualties	Minimal additional casualties

Level	What is the likelihood the risk event will happen?
E (High)	Near certainty
D	Highly likely
C	Likely
B	Unlikely
A	Remote

Likelihood					
E	M	M	H	H	H
D	L	M	M	H	H
C	L	L	M	M	H
B	L	L	L	M	M
A	L	L	L	L	M
	A	B	C	D	E
	Consequence				

	Assigned Risk Level
R	**High**- Unacceptable, Major disruption likely. Different approach reqd. Priority mgmt attention reqd.
Y	**Moderate** - Some disruption. Different approach may be reqd. Addl Mgmt attention may be needed.
G	**Low** - Minimum impact. Minimum oversight needed to ensure risk remains low.

Level	Technical Performance	And/or	Performance	And/or	Cost	And/or	Impact on Other Teams
A	Minimal or no impact		Minimal or no impact		Minimal or no impact		None
B	Acceptable, some reduction in margin		Additional resources reqd; able to meet need dates		< 5%		Some impact
C	Acceptable, significant reduction in margin		Minor slip in key milestones; not able to meet need date		5 - 7%		Moderate impact
D	Acceptable, no remaining margin		Major slip in key milestones or critical path impacted		7 - 10%		Major impact
E (High)	Unacceptable		Can't achieve key team or major program milestones		>10%		Unacceptable

Figure 2. Example use of standard risk chart

the likelihood of model error is determined by *associating* the level of reliance on the M&S and the level of confidence in the M&S.

- **Rule 3: Value of "consequence of model error" & "likelihood of model error" → risk.** The risk level is finally determined by ***associating*** the consequence of model error with the likelihood of model error.

We point out to the user that it is in defining these association rules that the tailoring of this process can best be done. Here is where experience, technical skills, thorough understanding of what is being modeled, how it is to be used, the associated costs and schedules, etc., come in to play. Depending on the scenario, the definition of the association rules can become a serious task. It may become necessary to

Likelihood

Level	Likelihood the risk event will happen?	Probability of occurrence
5 (High)	Near certainty	~90%
4	Highly likely	~70%
	Likely	~50%
2	Low likelihood	~30%
1	Not likely	~10%

5	L	M	H	H	H
4	L	M	M	H	H
3	L	L	M	M	H
2	L	L	L	M	M
1	L	L	L	L	M
	1	2	3	4	5

Level of Risk:

High, Med, or Low

Consequence

Level	Technical Performance	Schedule	Cost
5 (High)	Severe degradation in technical performance: cannot meet KPP or key technical/supportability threshold; will jeopardize program success; no workarounds available	Connot meet key program milestone Slip > ___months	Exceed APBA threshold > (10% of budget)
4	Significant degradation in technical performance or major shortfalls in supportability; may jeapardize program success; workarounds may not be available or may have negative consequences	Program critical path affected, all schedule float associated with key milestone exhausted Slip <___months	Budget inceases or unit production cost increases <(10% of budget)
3	Moderate reduction in technical performance or supportability with limited impact on program objectives; workarounds available	Minor schedule slip, no impact to key milestones Slip<month(s) of critical path Sub-system slip >___month(s)	Budget increase or unit cost increases < (5% of budget)
2	Minor reduction in technical performance or supportability, can be tolerated with little or no impact on program; same approach retained	Additional activities required, able to meet key dates Slip <___months (s)	Budget increase or unit cost increases < (1% of budget)
1	Minimal or no consequences to technical performance	Minimal or no impact	Minimal or no impact

**Source is Defense Acquisition University Continuous Learnig Course CLE 003*

Figure 3. Example of program risk reporting

Probability	Level of Impact			
	Catastrophic	Critical	Marginal	Negligible
Frequent	High	High	Medium	Low
Probable	High	High	Medium	Low
Occasional	High	Medium	Low	Low
Remote	Medium	Low	Low	Low
Impossible	Low	Low	Low	Low

Indicates risk level

Risk level values are:

- Subjective
- Based on MIL-STD-882C/D (US Military System Safety Standard)
- Tailorable to each problem

Source: MIL-STD 882D, Department of Defense Standard Practice for System Safety, February 10, Available at www:safetycenter.navy.mil/instructions/osh/milstd882d.pdf

Figure 4. Combination scheme for reduced number of risk levels

brainstorm a bit to determine what is suitable, and it may be necessary to get help from the experts and SMEs.

Risk determination procedure

The standard risk chart *(Figures 2–4)* requires the user to input values of likelihood and consequence. Thus, the use of any combination of the scales given above must result in the values of likelihood and consequence for the final step in the risk determination procedure. The association rules given above are meant to guide the user in defining a combination of schemes that will lead to this final step. A step-by-step recipe is provided with a simplified example to illustrate the process in what follows. No matter which scheme the user chooses, consideration must be given to consequences of the varying nature, including cost, schedule, personnel safety, political (including ridicule), and operational setting. The user must also accommodate all of the ways the model output could be wrong (e.g., erroneously overestimated or underestimated performance for which the consequences might be different in each case).

Conclusion

This article presents a methodology for assessing and prioritizing the risk associated with using M&S to support acquisition decisions. DoD, the Navy, and other Services mandate the use of M&S to support acquisition. Oftentimes limitations of resources make it hard to formally conduct VV&A of these models as needed to establish their credibility. However, using models and simulations whose credibility is in question can involve serious risk. The procedures presented here provide decision-makers with judicious ways for determining the risks associated with using a given M&S, and the extent to which VV&A work will be needed to reduce such risks. The approach presented parallels the formal DoD risk assessment procedure, but with application to the use of M&S as it relates to VV&A. The final output is a ranking of the risk associated with the models as either high, medium, or low risk. With this ranking the decision-maker is now able to decide whether or not to accept the risk or to invest the resources needed to mitigate the risk. An important outcome of this approach is that, through this method, there is the ability to determine what level of authority is needed to serve as the accreditation authority. ❑

JAMES N. ELELE, Ph.D., was born in Nigeria, West Africa, and immigrated to the United States 32 years ago. He earned a Bachelor of Science degree in chemical engineering (1980), a Master of Science degree (1985), and a Ph.D. in Applied Mathematics (1988), all from the University of Arizona, Tucson, Arizona.

He was employed as an engineer at IBM from 1980 to 1981 and at General Electric from 1981 to 1983. He was then employed by the U.S. government at the Electronics Proving Ground, Fort Huachuca, Arizona in 1988. He transferred to Patuxent River, Maryland, in 1997 and has continued there to the present.

Dr. Elele has extensive experience in modeling, simulation, and VV&A. He was part of the team that created the Army's Mobile Subscriber Equipment Performance Model (MSEPAM) and oversaw its VV&A. He served as the data specialist for the Army's Extended Air Defense Test Bed (EADTB) and completed a developmental assignment with the Army Test & Evaluation Command, which was instrumental in defining and introducing the Virtual Proving Ground (VPG).

Dr. Elele currently provides VV&A support to various acquisition programs, while serving as the M&S lead for T&E IPT in support of the development of the Marine Corps Heavy Lift Helicopter (CH-53K) replacement program. He also serves as the acting director of the BVVASO also known as JASA and as an adjunct professor for Strayer University and (formerly for) the Florida Institute of Technology. He has published 24 technical papers in journals & conferences, and he is the owner of two U.S. patents. E-mail: james.elele@navy.mil.

Endnotes

[1]DOD, 2003. Department of Defense Instruction 5000.61, DoD Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A), 12 May 2003.

[2]U.S. Navy, 2004. Department of Navy. "Modeling and Simulation Verification, Validation and Accreditation Implementation Handbook, Volume I: VV&A Framework." March 30, 2004.

[3]DMSO, 2000. Defense Modeling and Simulation Office (DMSO). Risk and Its Impact on VV&A. DMSO VV&A Recommended Practices Guide Special Topic Paper, November 30, 2000.

[4]BVVASO, (1997). VV&A from A to Z, Joint Accreditation Support Activity The Battlespace Verification, Validation, and Accreditation Support Office (BVVASO)/JASA publication on practical Work Breakdown on How to conduct VV&A.

[5]DMSO, 1996. Department of Defense "Verification, Validation, and Accreditation (VV&A) Recommended Practices Guide," November 1996. Available at: www.vva.dmso.mil.

References

Commander Operational Test Force, US Navy. "Navy Modeling and Simulation (M&S) Management." Chief of Naval Operations Instruction 5200.34, May 28, 2002.

Commander Operational Test Force, US Navy. "Use of Modeling and Simulation (M&S) in Operational Testing." COMOPTEVFORINST 5000.1A, September 5, 1995.

Commander Operational Test Force, US Navy. "Use of Modeling and Simulation (M&S) in Operational Testing." COMOPTEVFORINST 5000.1, September 9, 2004.

Commander Operational Test Force, US Navy. "COMOPTEVFOR Modeling and Simulation Accreditation Documentation." COMOPTEVFOR Policy and Information Notice 99-01, May 13, 1999.

Commander Operational Test Force, US Navy. "Accreditation of Modeling and Simulation in Support of Operational Test and Evaluation." COMOPTEVFOR Letter 3980 Ser 00T/312, April 29, 1998.

US Army. "Management Verification, Validation, and Accreditation of Army Models and Simulations." Department of the Army Pamphlet 5–11. September 30, 1999.

US Army. "Management Simulation Support Planning and Plans." Department of the Army Pamphlet 5–12, March 2, 2005.

US Department of Defense. "DoD Modeling and Simulation (M&S) Management." Department of Defense Directive 5000.59, January 4, 1994.

US Department of Defense. "DoD Modeling and Simulation (M&S) Management." DoD Directive 5000.59, January 4, 1994.

US Department of Defense. "DoD Glossary of M&S Terms." DoD 5000.59-M, December 1997.

US Department of Defense. "Modeling and Simulation (M&S) Master Plan." DoD 5000.59-P, October 1995.

US Department of Defense. "DoD Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A)." DoD INST 5000.61, April 29, 1996.

US Department of Defense. "Introduction to VV&A Standards and Practices for DoD M&S" (Draft), September 30, 1995.

Elele, J. E. 2007. "Assessing and Prioritizing the Risks Associated with the Level of Verification, Validation and Accreditation." *Proceedings of the NDIA 10th Annual Systems Engineering Conference*. San Diego, California, September 2007.

Elele, J. E. 2007. "A Methodology for Assessing and Prioritizing the Risks Associated with the Level of Verification, Validation and Accreditation of Models and Simulations." *Proceedings Annual ITEA Conference*. Kauai, Hawaii, November 2007.

Kilikauskas, M. and Hall, D. H. YEAR "The Use of M&S VV&A as a Risk Mitigation Strategy in Defense Acquisition," *JDMS*, 2, (4), p. 215.

Kilikauskas, M., Hall, D. H. and Weeks, J. 2003. "Comparison and Analysis of DoD and Service Policy and Guidance on Modeling and Simulation (M&S) Verification, Validation and Accreditation." Joint Aircraft Survivability Program Office Task M-7-03, December 2003.

"Department of Defense Standard Practice for System Safety." MIL-STD 882D, February 10, 2000, www.safetycenter.navy.mil/instructions/osh/milstd882d.pdf

MIL-STD-882 Rev D of February 10, 2000 is current issue, Rev E dated 1 February 1, 2000 is out in draft.

"Operational Risk Management Policy." NAVTESTWINGLANT/NAVTESTWINGPAC INSTRUCTION 3500-1, December 11, 2002.

NMSO. "Modeling and Simulation, Verification, Validation, and Accreditation Implementation Handbook, Volume I, VV&A Framework."

US Navy. "Navy Modeling and Simulation Master Plan." February 21, 1997.

US Department of Defense. "Risk Assessment and Its Impact on VV&A." DMSO RPG, November 30, 2000.

Defense Acquisition University. "Risk Management Guide for DoD Acquisition, 6th Edition," August 2006, Defense Acquisition University.

US Navy. "Department of the Navy Modeling and Simulation Program." SECNAVINST 5200.38, October 18, 1994.

US Navy. "VV&A of Models and Simulations." SECNAVINST 5200.40, April 19, 1999.

US Navy. "Department of the Navy Modeling and Simulation Management." Secretary of the Navy Instruction 5200.38A, February 28, 2002.

US Navy. "Verification, Validation, and Accreditation (VV&A) of Models and Simulations." Secretary of the Navy Instruction 5200.40A (Draft), November 3, 2003.

U.S. Air Force. "Operations Support: Verification, Validation and Accreditation (VV&A)." Air Force Instruction AFI 16-1001. June 1, 1996.

U.S. Department of Defense. "DoD Modeling and Simulation (M&S) Management." DoD Directive 5000.59. (Administrative Change 1 to the original document dated January 4, 1994). January 20, 1998.

U.S. Department of Defense. "Operation of the Defense Acquisition System." DoD Instruction 5000.2. May 12, 2003.

U.S. Department of Defense. "DoD Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A)." DoD Instruction 5000.61. May 13, 2003.

U.S. Navy. "Verification, Validation and Accreditation (VV&A) of Models and Simulations." Secretary of the Navy Draft Instruction SECNAVINST 5200.40A (in coordination), January 20, 2005.

"Verification, Validation and Accreditation of Joint Models and Simulations." Joint Staff Instruction JSI 8104.01A, January 26, 2004.

Acknowledgments

The author wishes to thank PMA-261 and its T&E/M&S Working Group (T&EMSWG) team members for providing the largest and most comprehensive, heretofore, practical application of the risk-based VV&A methodology. The team applied this methodology on over 150 models being used by the PMA, and the result is that they helped in ironing out some of the practical application issues associated with this methodology. I also wish to thank the IBST M&S Division (NAVAIR 5.4.2) for supporting and making the writing of this article possible. The knowledge and experience from previous work done by the Battlespace Verification, Validation, and Accreditation Support Office (BVVASO), also known as the Joint Accreditation Support Activity (JASA) is hereby acknowledged as invaluable as a source of reference for this article.