ITEA Journal 2011; 32: 39–51

Operational Testing: From Basics to System-of-Systems Capabilities

Col Michael W. Kometer, Ph.D.
School of Advanced Air and Space Studies, Maxwell AFB, Alabama
Kent Burkhart
605th Test and Evaluation Squadron, Hurlburt Field, Florida
Susana V. McKee and Dwight W. Polk, Ph.D.
505th Command and Control Wing, Hurlburt Field, Florida

This article examines the need to develop and implement a System of Systems (SoS) operational test methodology and supporting infrastructure in the U.S. Air Force (USAF). It discusses impacts on the acquisition and decision authority communities at large, with emphasis on operational testing in particular. Beginning with the fundamental requirements for testing, the article presents the realities of the effect of the Information Age on those requirements. The current progress in developing the acquisition business model and the technical capabilities to facilitate SoS test are reviewed. The authors then provide a plausible way ahead that will leverage training and testing capabilities through integrated venues.

Key words: Acquisition/fielding errors; capabilities-based evaluation; Joint Battle Management Command and Control (JBMC2) roadmap; Joint Theatre Air-Ground System capability; network-centric warfare; Red and Green Flag training; scientific method; Single Integrated Air Picture (SIAP); system-of-systems requirements; test and training integration.

During two-thirds of the 100 years the U.S. Army, and subsequently the U.S. Air Force, has been acquiring aircraft and other systems related to air and space warfare, the Department of Defense (DoD) has conducted very little dedicated operational testing to support acquisition or production decisions. With the exception of a 16-year span (1941–1957, Air Proving Ground, Eglin Field, Florida), the vast majority of government-conducted testing, even into the Vietnam era, was what is known today as developmental testing, with well-documented consequences, such as the first deployment of the F-111A to Southeast Asia. In March 1968, six F-111As were sent to Thailand for combat duty. After the loss of three aircraft in less than two months because of malfunctioning horizontal stabilizers, the remaining F-111As were returned stateside (Benson 1992).

Tactical Air Command established test centers at Eglin Air Force Base (AFB; Tactical Air Warfare Center, 1963) and Nellis AFB (Tactical Fighter Weapons Center, 1966) to help rectify the problem. The Air Force Test and Evaluation Center ("Operational" was a later addition to the title) was activated in 1974 to provide operational testing independent of the development, procurement, and user Commands for the largest acquisition programs. Quite naturally, the focus in the ensuing years of operational testing was on the individual system being acquired or fielded within the context of a limited operational environment. In other words, operational testing was effectively an extension of developmental testing. The larger questions of combat capability and mission contribution were left to the Modeling and Simulation (M&S) community. Two notable exceptions were the DoD-sponsored Joint Test and Evaluation (JT&E) program, established in 1972, and Tactics Development and Evaluations (TD&Es) dating to the early years of the Tactical Fighter Weapons Center.

Recent emphasis on capability-based testing takes the evolutionary process a step further. In particular, the test community has been tasked to conduct net-centric (or info-centric) and System-of-Systems (SoS) testing. Net-centric testing requires the investigator to

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **MAR 2011**	2. REPORT TYPE	3. DATES COVERED **00-03-2011 to 00-03-2011**

4. TITLE AND SUBTITLE **Operational Testing: From Basics To System-Of- Systems Capabilities**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **School of Advanced Air and Space Studies,Maxwell AFB,AL,36113**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
ITEA Journal 2011; 32: 39-51

14. ABSTRACT
This article examines the need to develop and implement a System of Systems (SoS) operational test methodology and supporting infrastructure in the U.S. Air Force (USAF). It discusses impacts on the acquisition and decision authority communities at large, with emphasis on operational testing in particular. Beginning with the fundamental requirements for testing, the article presents the realities of the effect of the Information Age on those requirements. The current progress in developing the acquisition business model and the technical capabilities to facilitate SoS test are reviewed. The authors then provide a plausible way ahead that will leverage training and testing capabilities through integrated venues.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**13**	

evaluate systems and their interoperability as part of a Find-Fix-Track-Target-Engage-Assess (F2T2EA) information network. This requires testers to evaluate capability beyond the limits of the particular system-under-test and its interoperability/integration with nearest-neighbor systems, as is currently practiced. SoS testing extends the scope of evaluation beyond that of merely placing a system in the context of a larger multi-system structure; it is the most plausible approach to testing that reaches the level of capability-based evaluation.

This article will examine the basic premise for testing and then show how the Information Age is affecting that purpose. It will then examine efforts to develop a new business model to facilitate testing to SoS requirements versus system-level requirements. It will look at some of the efforts to develop test infrastructure to perform distributed testing with integrated Live, Virtual, and Constructive (LVC) inputs. It will then discuss efforts to integrate testing and training and, finally, develop a concept for bringing all these developments together to accomplish SoS testing.

The primary purpose of testing systems and processes has remained unchanged over the years. Testing is evaluation conducted to mitigate the risks associated with new materiel and non-materiel solutions to warfighting capability needs. Operational testers need to determine as closely as possible the capability's "state of nature" effectiveness and suitability to avoid making the errors of either recommending fielding of a non-value-added capability or recommending not fielding one that could be value-added. The Information Age changes the focus of operational testing by redefining the penalties and benefits associated with the decision processes (or the loss function)—a system should no longer be measured against system-based performance, but against its contribution to overall warfighting capability as measured by SoS-based requirements. The Office of the Secretary of Defense (OSD) has given United States Joint Forces Command (USJFCOM) the job of leading the combatant commands (COCOMs) in defining these requirements by identifying Joint Mission Threads (JMTs) for important Joint missions that cross functional lines. OSD has also begun funding distributed test capabilities that can tie together the Services' test ranges and integrate live, virtual, and constructive (LVC) [simulation] methods to create true Joint environments. Still, assembling the resources for this Joint testing is beyond the constraints of the current fiscal environment, so it will take innovative test and training integration to address SoS requirements. It will also take work at the grass roots level by others in the Testing and Evaluation (T&E) arena (outside USJFCOM) to ensure they are ready to feed into emerging processes and infrastructure. Testers should become adept at using graduated levels of LVC as systems mature through spiral or incremental development to "graduation" in integrated test and training events. The 505th Command and Control Wing (505 CCW) is developing a way to join this effort by creating a (Joint) Theater Air-Ground System (TAGS) capability that can integrate a distributed SoS test capability with training venues such as Red and Green Flag.

The theory behind testing

Military Operational Test and Evaluation (OT&E), at its most basic level, is simply the application of the scientific method to decision requirements for hardware; software; concepts; and tactics, techniques, and procedures. In the larger context, operational testing can be explained in the context of game theory. It is an attempt to ascertain the true state of the capability being tested with respect to fulfillment of certain requirements. Testers sample that capability in a simulated combat environment or in a real operational environment, and based on the observed results, draw conclusions about the true underlying capability of the system, its "state of nature." The recommendations to decision makers follow from these conclusions. Decision-makers use this information to determine their actions based on risk considerations. Thus, the test is a risk-reduction tool for the decision-maker—it gives the best estimate of the state of nature. Appendix A provides a more complete mathematical explanation based on statistical decision theory (Ferguson 1967).

For traditional system-level operational testing, the states of nature could be considered dichotomous: the system meets requirements and performs satisfactorily (it is effective and suitable), or it does not. But we can also broaden this to a determination that the system contributes favorably to warfighting capability or has no (or even negative) impact on warfighting capability. The actual observed outcome of the test would be multi-dimensional, representing the level of attainment of objectives identified by the test team. A decision rule would be devised (typically subjectively) to map these potential observed outcomes into an action (recommendation) vector A, say $A = [a_1, a_2, a_3, a_4]$, where

- a_1 is a recommendation to field the system as is;
- a_2 is a recommendation to field the system after identified deficiencies are corrected;
- a_3 is a recommendation not to field the system but to continue development; and
- a_4 is a recommendation not to field the system and to cease development.

Theoretically, a loss function representing the consequences of the ordered pair (the state of nature and the action taken based on the decision rule and observed outcomes) could be calculated. The risk for each state of nature is the statistical expected value of the loss function.

Based on the identified risks for each decision rule, decision makers and the test team could develop a strategy for testing, recommendations based on test results, and subsequent decisions on acquisition and fielding.

Operational testers and decision makers do not define explicitly the loss function or document alternative decision rules—it is highly unlikely they will ever have sufficient data for this. However, intuitively they should be aware of the impacts of fielding immature or deficient systems and of withholding badly needed capability. Testers, either implicitly or explicitly, must account for Type I Errors (failing a mature or well-functioning system) and Type II Errors (accepting an immature or deficient system) and the subsequent impact of their recommendations on decision makers and ultimately the U.S. Air Force's (USAF's) warfighting capability.

Evaluation that extends the operational tester's purview to questions of a system's functioning within an information network or an SoS architecture requires a fresh look at the loss function associated with system capabilities and impact on warfighting capability as well as the overall statistical risk associated with test conduct and recommendations. As we shall see below, as systems become more interdependent for information exchange, the loss function should be based more on a system's impact to overall warfighting capability than on a comparison with system-level requirements.

What is different in the Information Age?

The military services were designed to organize, train, and equip forces to fight in their respective battlespace environments. This organization has its purposes—it ensures the particular needs within these environments are accounted for when developing the capabilities that will allow the military to perform its function. If all the capabilities were independent and did not come into contact with each other, there would be no need to test the systems in an SoS context. But they are not independent, and they do come into contact with each other. Support functions like close air support ensure the different services and functional components have to perform interdependently.

The Information Age has greatly increased the opportunities for integration and interdependence that drive this need to perform as an SoS. The ongoing technological revolution demands an appropriate response from those who wish to remain competitive. Command and control in the business world has demonstrated this revolution. For a century and a half the trend in American business was toward centrally controlling massive corporations. From single-unit, owner-managed enterprises with independent merchant distributors in the early 19th century, the American firm developed into a colossal, centrally managed behemoth in the late 20th century (Chandler 1977). But the Information Age pushed the trend toward decentralization and integration among the lower levels, instead of control from the higher levels. Strategy formerly aimed at controlling the actions of businesses is now instead aimed at constructing relationships among them, coordinating the use of resources so operations can be flexible yet focused. With today's information technology, workers can retrieve all of the information they need at the right time and place to make decisions on the spot, where they are most crucial (Castells 2000). Companies now look for others who have the core expertise to perform parts of their operations for them. They "Interlink" the "value chains" of suppliers, firms, and customers to transform the marketplace (Porter and Millar 1985). It is now more of a system than a pool of competitors.

The DoD is making a similar transformation. For more than a decade, Network Centric Warfare (NCW) prophets have urged this transformation. They propose that the military must prepare to fight NCW, "an emerging mode of conflict (and crime) at societal levels, short of traditional military warfare, in which the protagonists use network forms of organization and related doctrines, strategies, and technologies attuned to the information age" (Arquilla and Ronfeldt 2001). Technology has enabled these new modes because communication is faster, cheaper, and of higher quality. But NCW is not only about technology. It is about the linkages among people—networks, unlike formal hierarchies, are plastic organizations with ties that are constantly being formed, strengthened, or cut (Williams 2001). Most important, these analysts claim that "it takes networks to fight networks" (Arquilla and Ronfeldt 2001).

The U.S. military must capitalize on the current information revolution to transform its organization, doctrine, and strategy. It must retain its Command and Control (C2) capability, while becoming flatter—attaining faster response by eliminating some hierarchical levels in favor of pushing information out to all players at the lower levels. Doctrine should be built around battle swarming, a process of bringing combat power to bear at nearly any time and place based on real-time information (Arquilla and Ronfeldt 1997). The term "NCW" refers to a concept that "translates information superiority into combat power by effectively linking knowledgeable entities in the battlespace" (Alberts, Garstka, and Stein 1999). Its proponents argue that C2 should not be envisioned as a sequential process as it has been in the past—gathering data, analyzing, making a decision, and then implementing it. Instead, sensors, actors, and decision makers should be networked, so that they have a shared awareness of the battlespace. Commanders at the lowest levels will have enough information to take initiative and speed up the response to changing battlefield conditions (Alberts, Garstka, and Stein 1999).

For the acquisition community, the effect of this transformation is that the systems it acquires for the military Services are increasingly required to interoperate with systems of other services. As Admiral Cebrowsky put it, "In reality, what has happened is that a new air-ground system has come into existence where you no longer talk in terms of one being supported and the other supporting. That would be like asking if the lungs are in support of the heart or if the heart is in support of the lungs. It's a single system" (Cebrowsky 2002). The vision of Air Force leadership through the 1990s was that airpower would be able to execute a "kill chain" as rapidly as possible due to the smooth integration of a "system of systems" (U.S. Air Force 2000). The acquisition community has taken strides to facilitate machine-to-machine transfer of data among these systems, using tools like Web services with XML data.

This is the SoS that operational testers must learn to evaluate. In truth, the SoS could take many forms, accomplishing many different operational threads. Close air support, defense against ballistic and cruise missiles, dynamic targeting of mobile ground targets, and construction of a single integrated air picture are all missions that require an SoS to work in an integrated fashion. However, the acquisition community is not structured to consider the needs of the Joint environment in the requirements process (DOD DOT&E, 2004). Recognizing that this emerging network-centric paradigm required a different systems engineering approach, the DoD promulgated guidance in Defense Planning Guidance (DPG) 2003 and 2004 and Strategic Planning Guidance (SPG) 2006. These documents moved the DoD towards a Net-Centric Global Information Grid (GIG) and Network Centric Enterprise Services (NCES) (DOD AT&L, 2004).

If the decision makers were truly to attempt to define the loss function, they would have to consider the impact of the system on the performance of the SoS, not just the comparison of the system to its own isolated requirements. A new gateway may have requirements to forward certain message formats, but its real function is to synchronize the situational awareness of the commanders and troops and allow a more rapid (and effective) transition from information to action. This is the "impact on warfighting capability" discussed earlier. The loss function should be defined based on this broader impact, not on whether it meets the narrower system-based requirements. The loss is positive (or at least non-negative) if the system is fielded but does not increase warfighting capability, or if it is not fielded but could have increased warfighting capability—regardless of whether or not it forwards the required message formats.

Testing only the narrower system-based requirements actually increases the probability of making a Type-II error. Testers could induce loss by recommending fielding of a system that will not have a positive impact on warfighting in today's environment. Warfighters in command and control positions have so much information that simply adding more information does not make them more effective—the information must be added in a way that allows them to do their job more effectively.

This brings up a reality that cannot be overlooked: evaluation of the SoS is decidedly incomplete without consideration of the human interactions involved. All this information must be organized and presented in a way that enables the warfighters to do their jobs effectively and efficiently. The more warfighters are able to cross functional and service lines, the more avenues they have to be innovative in accomplishing the mission. However, having more avenues for innovation also means increased difficulty enforcing global procedures. If the people who make command and control possible are unsure of or drift from global procedures meant to avoid fratricide and other unintended consequences, accidents could occur (Kometer 2007). Tests of NCW capability require

1. assessment of the capability of the SoS as an aid to the people to bring combat power to bear at the right time and place,
2. determination of C2 responsibilities from the lowest tactical level to the strategic level, and
3. development of tasks, techniques, and procedures (TTP) to implement the entire network and SoS.

A business model for SoS requirements

The problem is that testing is requirements driven, and right now requirements are mostly system based (U.S. Air Force 2004a). Currently, DODI 5000.2 requires all systems to undergo interoperability evaluations throughout their life cycles (DOD 2003). For information technology systems with interoperability requirements, the Joint Interoperability Test Command (JITC) must provide certification of critical interfaces throughout the life cycle, regardless of acquisition category (Wiegand 2007).

In fact, Network-Readiness (NR) Key Performance Parameters (KPP) assess NR (which is defined in DODD 4630.5) as follows: "Net-readiness is the continuous capability to interface and interoperate to achieve operationally secure exchanges of information in conformance with enterprise constraints. The NR-KPP assesses net readiness, information assurance controls, and both the technical exchange of information and end-to-end operational effectiveness of that exchange" (DOD 2007).

But most current testing only validates that the systems will be able to pass data to the appropriate interoperable systems. It does not address whether the SoS will function correctly—or, in particular, that the people involved will understand how their role changes when a mission crosses normal functional lines. To go to this level, the systems must be tested within the larger SoS. Indeed, JITC is heavily involved in the push toward end-to-end Joint environment testing to answer the NR-KPP (TRMC 2006, Clarke 2007).

The conceptual issue with tests of this sort stems from the fact that there is no organizational responsibility and, therefore, no resources are available for SoS testing involving multiple command and control, intelligence, surveillance, and reconnaissance (C2ISR) nodes and weapons systems. Systems are currently funded by program, not by capability. Individual program offices fund testing of their system requirements only. Although these requirements now include NR-KPPs, these KPPs deal only with whether the system can receive and provide information in formats required to be interoperable with other nearest-neighbor network systems. They do not specify how the larger SoS should do its job. Yet in the Information Age, a program does not constitute a capability. Capabilities cut across multiple programs, requiring them to interoperate and exchange information.

Tests of capability-based requirements call for a new business model. The Joint Battle Management Command and Control (JBMC2) roadmap is a capabilities-based construct, which lays out the elements of this model. The roadmap relies on a Joint Staff–developed concept of how the Joint force will operate in the future across the range of military operations. This "Joint Operations Concept" leads to Joint Operating Concepts, Joint Functional Concepts, Joint Enabling Concepts, and Integrating Concepts. While its future may be in doubt, the executor of the JBMC2 Roadmap, USJFCOM, currently leads the COCOMs in the development of JMTs—comprehensive descriptions of how the Joint force will execute one of seven warfighting capabilities. The Joint Staff and Joint Requirements Oversight Council (JROC) (or functional capabilities boards, on behalf of the JROC), reviews and validates requirements for development of the JMTs (DOD AT&L 2004). This business model does not change the milestones or purpose of test. Testing of SoS is to be accomplished within the context of existing Developmental Test and Evaluation (DT&E) and Operational Test and Evaluation (OT&E) in the Test and Evaluation Master Plans (TEMPs). But the testing must be done in a Joint environment, using validated requirements for the relevant Joint mission (DOD DOT&E 2004).

DoD is leading an effort to implement such a model for the Single Integrated Air Picture (SIAP)—a "shift in the Department of Defense's traditional focus on an individual combat system's performance to the ensemble performance of a SoS" (JSSEO 2006). The Joint Theater Air and Missile Defense 2010 Operational Concept produced a conceptual template for Joint Integrated Air and Missile Defense that depended on a SIAP (JSSEO 2006). So the Vice Chairman of the Joint Chiefs of Staff (VCJCS), the Under Secretary of Defense for Acquisition, Technology, and Logistics (USD AT&L), and the DoD Chief Information Officer (CIO) chartered the SIAP system engineering task force in 2000. They needed a disciplined, Joint process to resolve interoperability problems to implement a truly Joint data network that could create the SIAP. As the test strategy puts it, "Technically, the SIAP is a state of mutual consistency within a weakly-connected, heterogeneous, decentralized, and distributed system" (JSSEO 2006). Unfortunately, where human organizations and money are concerned, "weakly-connected" and "heterogeneous" do not induce action.

In 2003, the JROC designated JFCOM and a SIAP Acquisition Executive to direct the program and establish funding lines in the Services (JSSEO 2006). The Services will be responsible for ensuring their individual systems conform to an Integrated Architecture Behavior Model (IABM)—an open architecture computer model of prescribed system behavior with bit-level precision that is the product of Joint system engineering (DOD 2003).

Distributed test capabilities for SoS test

Getting the requirements right is just one part of the equation. Another problem with testing Joint, NCW capability is the need to construct the SoS for a test. Command and control of the forces requires a sophisticated network including Internet, landline, satellite, and line-of-sight protocols. This is the SoS under test. But T&E of this SoS requires another SoS to monitor and collect data during the test. On top of this, the test environment may require augmentation of live assets with virtual and constructive Modeling and Simulation (M&S) methods in a Hardware-In-The-Loop (HITL) configuration.

"Testing in a Joint Environment Roadmap for 2006–2011" (DOD DOT&E 2004) presented guidance for developing these capabilities. It addressed the fact that the Services each had disparate test capabilities that were in some cases redundant and in many cases insufficient. It foresaw the need to create a universal "persistent, robust distributed systems engineering and test network that can link the specific remote sets of HITL, M&S, and other resources with the live system in development to accomplish the systems engineering or testing required for a spectrum of transformational initiatives, as well as to support training exercises and experimentation" (DOE DOT&E 2004).

The roadmap set the stage for the Joint Mission Environment Test Capability (JMETC). The JMETC program office falls under the Test Resource Management Center (TRMC), which reports to USD AT&L. It collaborates with the Central Test and Evaluation Investment Program (CTEIP) to fund the programs that will provide a corporate approach to integrating distributed LVC capabilities, solving the problems inherent with the Service-specific capabilities, multiple networks, and various different standards that exist. It was therefore meant to reduce duplication of effort, provide readily available security agreements, and facilitate Joint testing and integrated test and training. JMETC establishes persistent connectivity via a Virtual Private Network (VPN) on the SECRET Defense Research and Engineering Network (SDREN), adopts the Test and Training Enabling Architecture (TENA) middleware and standard interface definitions, collaborates with CTEIP to adopt distributed test support tools, and provides data management solutions and a reuse repository (Ferguson 2007). The distributed test support tools are being developed as part of a project called the Joint C4ISR Interoperability Test and Evaluation Capability (InterTEC) and include communications control, test control, instrumentation and analysis tools, synthetic battlespace environment, and simulation/emulation gateways (JITC 2008). As the CTEIP 2006 report puts it, "the envisioned end-state is a seamlessly linked, but geographically separated, network of test facilities and ranges in which the most modern and technologically advanced defense systems can be tested to the full extent of their capabilities" (TRMC 2006).

These capabilities are still in the maturing phase. The development of Joint Close Air Support (JCAS) mission threads was the impetus for several events during 2007. The 46th Test Squadron (TS), Eglin AFB, conducted a baseline assessment of Link 16 and Situation Awareness Data Link (SADL) to answer questions about the capability of Joint Terminal Air Controllers (JTACs) to send digital 9-line messages directly to the cockpit (46th Test Squadron 2007).

That same year, the Simulation and Analysis Facility (SIMAF) at Wright-Patterson AFB sponsored the Air Force Integrated Collaborative Environment event "Integral Fire 07." This event developed a distributed test environment to satisfy three different test customers: JFCOM's Joint Systems Integration Center, the DoD Joint Test and Evaluation Methodology JT&E program, and the Warplan-Warfighter Forwarder initiative, sponsored by the USAF Command and Control and Intelligence, Surveillance and Reconnaissance Battlelab. This was the inaugural use of the JMETC to tie together three separate enclaves—15 total locations—with an aggregation router. The sites used the TENA gateways to exchange simulation or instrumentation information via TENA protocols (TENA 2008).

More recently (2009–2010), JMETC has participated in JEFX events to provide the network backbone to share and exchange near-real-time tactical information among participants and monitoring test agencies. Two examples are the use of the Guided Weapons Evaluation Facility (GWEF) to generate simulated Net-Enabled Weapons (NEW) for management and control by Air Operations Center personnel (JEFX 09) and F-15 and JSTARS Operational Facility (OPFAC) simulation of manned aircraft interacting with live systems to "round out" testing (JEFX 09, JEFX 10).

In September 2007, USJFCOM J85 conducted an Advanced Concept Technology Demonstration (ACTD) called BOLD QUEST that provided another opportunity to develop distributed test capability. Expanding on the JCAS JMT, JFCOM decided to look at interoperability of coalition fighters with three Joint terminal air controller suites: Tactical Air Control Party Close Air Support System (TACP-CASS); Battlefield Air Operations (BAO) Kit; and Target Location, Designation, and Hand-off Kit. BOLD QUEST had been designed to assess Coalition Combat Identification (CID) but was deemed the right venue and timing for USJFCOM's Joint Fires Interoperability and Integration Team (JFIIT) to conduct the JCAS assessment as well (JFIIT 2007). The 46 TS and 640th Electronic Systems Squadron (Hanscom AFB, Massachusetts) deployed three mobile data link facilities (called "Winnies") to Angels Peak, Antelope Peak, and the National Training Center (NTC). These Winnies supported both data collection and the tactical infrastructure capability for the event.

In this way the 46 TS demonstrated the capability for remote collections with local, centralized, analysis. The 46 TS established network connectivity to Winnies at Angels Peak, Antelope Peak, NTC, and a JFIIT-provided range-instrumentation data-stream at Nellis AFB. Additionally, each mountain top Winnie established connections to the Joint Range Extension (JRE) at the Joint Interface Control Cell (JICC) Combined Air and Space Operations Center (CAOC) and JFIIT Gateway Manager (GM). These configurations provided dedicated data to 46 TS, JFIIT, and the JICO (Cebrowski 2002). Remote data collection and analysis are, of course, important components of distributed testing. JMETC has provided additional networking capability during BOLD QUEST 08/09/10 to employ 46 TS OPFACs in testing the CID server's capability to determine receipt of fighter Link-16 J12.6 messages asking for five closest friendly positions near intended target area before employing the CID server in live-fly environment. In addition, it supported testing and verification of CID server J3.5 responses to fighter queries.

The challenge of executing SoS tests

SoS testing is not achieved typically because the cost is prohibitive and dedicated access to key assets is frequently limited or nonexistent. Some test organizations made progress toward objectives in the development of warfighting capabilities or processes, but not in proportion to the cost. Typically, with the exception of BOLD QUEST, these events were underwritten with the intent of developing the capacity to perform this type of testing in the future. Until SoS test environments are consistently—or even persistently—available, it will not be feasible for the test customer to test this way in the absence of sponsorship (46th Test Squadron 2007). Even after the test SoS is available, testing in a Joint environment typically requires significant live assets, more coordinated planning, and a robust scenario. ACTDs like BOLD QUEST and other specific projects may have the budget to accomplish this, but most will not.

Getting all of the assets together at the same time and appropriate place is also difficult. A robust command, control, communications, computers, intelligence, surveillance, and reconnaissance (C4ISR), air, space, and cyber network and appropriate-sized strike force could include hundreds (or thousands) of entities integrated with real-world communications and other operational infrastructure. Operational assets such as airborne C2 platforms, control and reporting centers, and tactical air control parties are in high demand for deployments and exercises. Scheduling them and getting them to operate simultaneously in the same dedicated test battlespace is extremely challenging for many reasons (DOD DOT&E 2004).

One solution being investigated is integrated test and training. Even though the two venues differ in their objectives, test and training share common resources and analytical methodologies to some extent (DOD DOT&E 2004). Exercises and experiments already attract most of the assets required for end-to-end tests. Green Flags present an air-ground war scenario, while Red Flags are centered on preparing for the air war in general air warfare. Atlantic Strike prepares ground troops for coordination with air support in the war on terrorism. The added fidelity and infrastructure of assets usually identified with testing would be welcomed in many cases. For this reason, OSD has directed the integration of test and training in a memo to the services (Krieg, Chu, and McQuery 2006).

But the two disciplines do not mix easily. Any testing added to the exercise venues likely would require transparency to training audiences and have no negative impact on training objectives. Testers would have to participate throughout the planning process to develop appropriate scenarios and identify required operator actions to satisfy their data requirements. The primacy of training objectives may preclude capturing all the data necessary to complete test objectives. In addition, testers might be restricted from repeating events where the conditions were not right for test purposes. For these reasons, the debate about whether test and training integration will work routinely in practice rages on.

Some test and training integration efforts have demonstrated the potential for success. JFIIT conducts accreditation of C2 procedures at the National Training Facility and at Avon Park during Atlantic Strike exercises for ground troops headed for the Central Command theatre (USJFCOM 2010). Along with this training role, the unit frequently accomplishes test activity using the same resources. For example, at the Atlantic Strike exercise in November 2007, JFIIT teamed with 605 Test and Evaluation Squadron (TES) and 46 TS to develop the architecture for and demonstrate the performance of the new Air Support Operations Center (ASOC) gateway. This was not a formal test, but it reduced risk for the upcoming operational test of TACP-CASS 1.4.2 and ASOC gateway by solidifying the concept of employment (CONEMP) and TTPs. JT&E organizations like Joint Datalink Information Combat Execution (JDICE) and Joint Command and Control of Network Enabled Weapons (JC2NEW) have successfully used training venues to accomplish their objectives. JDICE conducted a quick reaction test of Joint Integration of

Nationally Derived Information at Valiant Shield 07 and several real-world events. JDICE also validated Link 16 TTPs and architectures to enhance the kill chain and to filter and de-conflict the targeting picture for ground troops in Red Flag 04, Valiant Shield 06, and Red Flag 06.[1] The effort validated architectures and TTPs for the national intelligence community to provide information rapidly to tactical and operational level warfighters. It should be noted, however, that JT&Es focus on tactics, processes, and procedures—not new systems.

On the whole, the test world is still struggling to develop the methodology for integrated testing and training in a way that brings the promised efficiencies. The transition to capability-based requirements and testing is ponderous. New systems reaching operational testing are still, for the most part, subject to the classic requirements process. This is true to an even greater degree for capabilities in the sustainment phase. For Air Force systems in sustainment (and new ones for which Air Force Operational Test and Evaluation Center [AFOTEC] non-involves) Major Command (MAJCOM) test organizations often assume responsibility for operational testing (U.S. Air Force 2004a).

505 CCW concepts for SoS test

As the Air Combat Command (ACC) focal point for C2 operational testing, 505 CCW is somewhat trapped in the systems-based requirements process. The 605 TES tests only those capabilities that are in sustainment or for which AFOTEC waives involvement. The squadron tests programs based primarily on ACC or other requesting agency requirements, and usually not JROC-approved JMT requirements, as mentioned earlier. To get to true NCW, testers will have to adopt an SoS methodology for new systems. Upgrades to fielded systems must also be tested for their contribution to warfighter capabilities, for consistency with testing to this standard before and during initial operational test and evaluation (IOT&E). MAJCOM testers have not been involved in the SoS requirements development process to this point. Testers do not (indeed should not) develop requirements except to the extent that they can influence those who do to make the requirements "testable," and program offices are not inclined to fund testing beyond that which determines the extent of system-level compliance with requirements.

TD&Es are an alternative avenue for addressing mission-level testing. When warfighters need additional capability, new materiel solutions provide one way to fill the shortfall. Non-materiel solutions are another possibility. In Air Combat Command, wings submit tactics improvement proposals (TIPs), which are then prioritized at an annual tactics review board in January and subsequent overall combat air forces (CAF) test prioritization process. Predictably, these TIPs often call for development of TTPs for dealing with cross-functional problems such as close air support, nontraditional ISR, or sensor fusion. C2 TD&Es typically are not accomplished because of a lack of advocacy at sufficiently high levels and inadequate funding. In spite of being the CAF's lead for C2 operational testing, including TD&Es, the 605 TES is largely funded by charging program offices (or other requesting agencies) for level-of-effort support and travel expenses on specific test projects. ACC has never formalized TD&E funding even though some TD&Es are among the highest prioritized CAF test projects each year (Kometer 2007). Piggy-backing TD&Es on training exercises may be the only way to accomplish them on a recurring, systematic basis, at least for those involving multiple and varying platforms/systems.

To gain access to major training exercises for test purposes, the test construct must be transparent (or nearly so), and preferably beneficial to the training community. The largest training exercises, the Red and Green Flags and Weapon School Mission Employment Phase, are flown out of Nellis AFB, where the 422d Test and Evaluation Squadron and the 57th Wing are located. These exercises already use CAOC-Nellis (CAOC-N) to provide a limited operational and tactical C2 experience to enhance the excellent tactical level training traditionally associated with these events.

The 505 CCW, in collaboration with 46 TS and SIMAF, is developing a test and training capability to begin bridging the gap to net-centric and SOS testing. Starting in 2007, the 605 TES earned OSD Resource Enhancement Program money to fill operational testing shortfalls for Theater Battle Management Core Systems, datalink, TACP-CASS, and network-centric collaborative targeting. In implementing their shortfall solution, the team was able to include a gateway to access the distributed architecture discussed above. The wing has successfully obtained CTEIP funding to implement infrastructure to enable the conduct of integrated LVC end-to-end testing in conjunction with existing training events to support Joint warfighting customers by leveraging the Western Range Complex, LVC experts, and distributed architectures. OSD is providing additional support toward developing capability with the goal of creating a simulated TAGS anchored around CAOC-N that can be integrated with Air Force, Joint, and coalition training events. It will be fully meshed with JMETC and InterTEC to ensure it can be an integral part of the distributed testing architecture discussed earlier (Wiegand 2007). The 505 CCW is reorganizing to focus greater attention on integrated testing of air, space, and cyber capabilities in this emerging environment. The new organizational structure will facilitate better mutual support among the Wing's geographically separated testing, simulation, and training components. It will also provide greater opportunity for leveraging the connectivity provided by its Distributed Mission Operations Center at Kirtland AFB to other DoD developmental, simulation, and testing components needed for robust SoS testing. Additionally, the addition of the JDICE mission and expertise to the Wing will tremendously enhance the capability to test and train in the emerging net-centric Joint range environment.

Although the initial drivers for developing the TAGS environment were TD&Es, the project was undertaken because of the foresight needed to meet SoS testing requirements at every level. Talks with the aircraft and weapons testers in the 53d Wing helped stimulate this developmental effort since both ACC test organizations realized they would have to interface for net-enabled weapons testing. As C2 and weapons systems become more interdependent they become elements within an integrated SoS architecture and must be tested as such. Other C2 capabilities already demand SoS test methods.

The proposed concept for transitioning to SoS testing is based in part on evolutionary acquisition (the sequential release of increments and versions of capability), the currently favored paradigm for acquisition of many C2 systems. This paradigm comes from the software development industry and has often been associated interchangeably with "spiral development." Software developers realized long ago that first attempts to deliver a coded product often would reveal significant problems. Upon seeing early deliveries, users would find glitches and possibly even realize that originally stated requirements did not adequately describe their needs. But software-intensive systems are often easier and less expensive to change than hardware acquired from production lines and, thus, are more amenable to concurrent and sequential development, correction, and enhancement. Thus, software developers evolved the best practice of developing capability in successive iterations called "spirals," each of which provided greater/improved capability while more fully satisfying customer needs. When, during the spiral process, the capability was finally acceptable, it could be delivered. The current acquisition model for software-intensive C4ISR systems is built around this approach. DODI 5000.2 (DOD 2003), AFI 99-103 (U.S. Air Force 2004a), and AFI 63-101 (U.S. Air Force 2004b) tell us that capability is to be delivered in "increments," each of which is to undergo this type of spiral development.

Marrying the spiral development model with the constraint on limited open-air test time leads us to consider a graduated approach to testing, using LVC infrastructure and capabilities as much as possible. The increasingly robust and progressively more operationally relevant electronic warfare test process, defined in AFMAN 99-112 (U.S. Air Force 1995), could be a starting point for this approach. This process recommends the use of six types of test support/environments—M&S, system integration laboratories, HITL, measurement facilities, installed system test facilities, and open air ranges—each of which tests a system in a different state of maturity. Program managers are encouraged to select the appropriate level of testing at the appropriate time in the life cycle of a system to reduce risks for the acquisition program and for future tests.

C4ISR systems could use a similar progression, using varying levels of LVC testing at the appropriate time in the process. The test strategy should include integration into the larger C4ISR SoS within which the individual system will function. In early development, this could be accomplished by using a simulated environment such as that provided by the JMETC architecture, essentially a distributed HITL network. As the system matures, real systems could be added to this HITL architecture structure to ensure that the system can perform satisfactorily in an operationally realistic environment (while retaining the operationally representative stimuli provided by the simulated environment). During testing with progressively more live-entity involvement, testers could examine the system's performance, effectiveness, and suitability within the SoS and the impact on the TTP employed by the operators within the SoS. Finally, open-air/field exercises could be used to verify system capabilities and TTP developed through earlier spirals, in a "graduation event" for OT&E.

For example, TACP-CASS version 1.2.5 test results showed the need for this type of approach. The system passed Developmental Testing (DT) and proceeded to operational testing in 2005. One system requirement was that it successfully interface with the Army Battle Command System to receive friendly ground force situational awareness messages. However, in the operational environment, it bogged down when the message load approached that of a division. It also had problems reconciling messages from two Army units in close proximity to each other. The problems were so significant that TACP-CASS was not fielded—solutions were postponed to later versions (DiFronzo 2006). These problems were not observed during DT because of the constrained DT test environment. The anomalies appeared once the system encountered operationally representative information traffic. Testing that requires a dedicated large Army force and live aircraft is too expensive to be accomplished throughout development of a system. The proposed alternative would subject the developing system to a sequence of progressively more robust and demanding simulated environments, culminating in multiple systems in a division-sized configuration. Had this been done, the deficiencies would have surfaced as the scale and relative positions of real-world forces evolved. These deficiencies, in turn, could have been corrected prior to operational testing.

Similarly, during 2009 testing of TACP-CASS 1.4.2, the 605 TES established an operationally realistic Joint air request High-Frequency (HF) radio network to evaluate the capability of contractor off-the-shelf (COTS) equipment designed to provide an alternative to the fielded satellite communications (SATCOM) system. The radios performed satisfactorily in a limited network designed during DT but failed miserably in operational field testing. The cause of the failure is still being investigated, and the system was fielded without the radios. While TACP units can accomplish their mission via SATCOM, the opportunity to relieve the demand on SATCOM support was lost. A persistent network, readily available to the test community, could have provided the proper environment to detect the deficiency in early TACP-CASS 1.4.2 testing rather than in final operational testing.

Conclusions

Right now, it is not clear who will be responsible ultimately for SoS testing. It is quite likely that for the near future, Service operational test agencies and MAJCOM operational testers will continue to test systems to system-level requirements. However, as these systems are increasingly seen as families of systems that affect the JMTs, they eventually need to be tested in Joint SoS environments. That makes it imperative that the initial testing lead to and be guided by the eventual requirements for SoS testing, including common measures and objectives where possible, and by common LVC-enhanced environments wherever feasible.

As the capability to accomplish SoS testing matures and becomes persistent and universally available, smaller test organizations, like those of the 505 CCW and other MAJCOM testers, will be able to access the services of JMETC and M&S providers more easily. Operational test organizations will work with DT organizations to ensure M&S adds operational reality and robust environments into DT (possibly via operational assessments) that will reduce the risk for eventual OT&E during an integrated test and training exercise. The data collection infrastructure employed simultaneously with training exercises will facilitate data collection and management for both test and exercise agents.

However, the test community has not progressed to that point yet. Distributed test capabilities are not universally available yet, much less embraced by all in the greater acquisition and user communities. Indeed, it will take both greater incentives and support from OSD and increased initiative from the grass roots level to lead the two communities in that direction. Test units, user commands, and program offices will have to agree to lobby for the addition of testing in the Joint environment, using LVC simulation early on while leading to robust SoS tests. Test units will have to become creative and develop approaches for testing and training integration acceptable to all responsible organizations and at the same time champion funding the investments required for testing at the network and SoS levels.

The transformation of DoD testing will require a change in culture. Testing in conjunction with training will require testers to collect data over multiple events to accomplish their objectives; they will no longer be able to rely on planning a single dedicated opportunity to satisfy all objectives. This effort may even lead to more opportunities to acquire reliability, availability, and maintainability data as systems undergo testing over several events instead of just one. The outcome could be serendipitous to developers and testers trying to respond to DOT&E's direction (DOT&E Memo, 30 June 2010) for greater rigor in addressing suitability issues. Of course, this could wreak havoc with success-oriented program schedules. Testers will have to work with decision makers to determine what level of uncertainty is acceptable to demonstrate the impact of the new capability on the warfighting SoS. Seamless verification will include seamless transition from M&S-based evaluation to testing in a robust Joint LVC environment. Unless the acquisition and test communities are willing to embrace creative means like those discussed, and they receive the necessary support from service and OSD leadership, decision makers will not be able to gauge effectively the risks involved with their decisions for fielding new capability. Capability-based acquisition is mandated by OSD, it is gaining momentum, and it is the right thing to do for our nation's security. ❑

COL. MICHAEL W. KOMETER is currently professor of security studies at the School of Advanced Air and Space Studies (SAASS), Maxwell AFB, Alabama. Commissioned at the Air Force Academy (Colorado Springs, Colorado) in 1988, he is an electronic warfare officer with combat time in the AC-130H and experience as a planner and strategist in multiple conflicts. He has performed

developmental and operational test, including commanding the 605th Test and Evaluation Squadron from 2006 to 2008. Col. Kometer earned master's degrees from Georgia Tech (Atlanta, Georgia), Air Command and Staff College, and SAASS and a doctor of philosophy from Massachusetts Institute of Technology (Cambridge, Massachusetts). E-mail: michael.kometer@maxwell.af.mil

Lyle Kent Burkhart is an operational test contractor (Sentel) for the 605th Test and Evaluation Squadron, Hurlburt Field, Florida, conducting operational T&E of C2 systems. He has over 17 years of T&E experience and 20 years active duty in the USAF, 8 years in C2. Mr. Burkhart has led initiatives in support of test and training interdependencies and continues to support these efforts as a system tester. He received his bachelor's degree in general studies at Louisiana Tech University, Ruston, Louisiana, and is completing his master's degree in systems engineering with Regis University, Denver, Colorado. E-mail: Lyle.Burkhart.tr@.af.mil

Susana V. McKee has 24 years of T&E experience. She is the 505th Command and Control Wing's chief of capabilities development and integration at Hurlburt Field, Florida. Previously, Ms. McKee was director of test for the 605th Test and Evaluation Squadron C4ISR testing efforts. At Eglin AFB, she served as chief of analysis for the 53rd Wing's Air-to Ground Weapon System Evaluation Program and was an analyst for operational flight program upgrades. At the 46th Test Squadron, she conducted safe escape and ballistics analyses. Ms. McKee earned undergraduate and graduate degrees from University of West Florida, earning mathematics and political science degrees. E-mail: Susana.McKee@Hurlburt.af.mil

Dr. Dwight Wilson Polk is technical advisor and chief of the Test Office, 505th Command and Control Wing, Hurlburt Field, Florida. Dr. Polk is the wing's technical authority on combat air forces command and control T&E. He also provides technical integration for the wing with the USAF Warfare Center and Air Combat Command. Dr. Polk earned bachelor of science and master of science degrees in mathematics, and a doctor of philosophy degree in engineering management from Clemson University, Clemson, South Carolina. He has been associated with T&E in some form during 34 of his 40 years of combined active duty and civil service experience. E-mail: Dwight.Polk@Hurlburt.af.mil

Endnotes

[1]Suh, Theresa L., Maj., USAF. "Combining Test and Training." Briefing received via e-mail from Michael Tamburo, Major, USAF to (M.W.K.) 7 January 2008.

References

Alberts, David S., John J. Garstka, and Frederick P. Stein. 1999. *NetworkCentric warfare: Developing and leveraging information superiority*. Washington, D.C.: DOD C4ISR Cooperative Research Program.

Arquilla, John, and David Ronfeldt. 1997. "Looking ahead: Preparing for information-age conflict." In *In Athena's camp: Preparing for conflict in the information age*, ed. John Arquilla and David Ronfeldt. Santa Monica, CA: Rand.

Arquilla, John, and David Ronfeldt. 2001. "The advent of netwar (revisited)." In *Networks and netwars*, ed. John Arquilla and David Ronfeldt, 1–25. Washington, D.C.: Rand.

Benson, Lawrence R. 1992. "History of Air Force Operational Test and Evaluation (OT&E) mission, organization, and policy." December 1992. Kirtland, AFB, NM: Air Force Operational Test and Evaluation Center.

Castells, Manuel. 2000. *The rise of the network society*. Malden, MA: Blackwell.

Cebrowsky, Arthur. 2002. Vice Adm. (retired), U.S. Navy. Quoted in *Army Times*, 25 November 2002.

Chandler, Alfred T. Jr. 1977. *The visible hand: The managerial revolution in American business*. Cambridge, MA: Harvard University Press.

Clarke, Rich. 2007. "JITC NR-KPP Testing and Certification Presentation to ITEA." Briefing, April 2007.

DiFronzo, Vincent P. 2006. "Memorandum for USAF WC/CG, ACC/A3, and ACC/A8. Subject: Do Not Field Memo for the TACP-CASS (v) 1.2.5." February 23, 2006.

DOD. 2003. *Operation of the defense acquisition system.* Department of Defense Instruction Number 5000.2, 12 May 2003. Washington, D.C.: U.S. Department of Defense, http://www.dtic.mil/whs/directives/corres/pdf/500002p.pdf (accessed December 14, 2010).

DOD. 2007. *Interoperability and supportability of Information Technology (IT) and National Security Systems (NSS)*. Assistant Secretary of Defense (Networks and Information Integration). Department of Defense Directive (DODD) 4630.05, 23 April 2007. Washington, D.C.: U.S. Department of Defense. http://www.dtic.mil/whs/directives/corres/pdf/463005p.pdf. (accessed December 14, 2010).

DOD AT&L. 2004. Under Secretary of Defense Acquisition, Technology, and Logistics (USD AT&L) and United States Joint Forces Command (USJFCOM). "Joint Battle Management Command and Control (JBMC2) roadmap, Version 1.0." 22 April 2004. Washington, D.C.: U.S. Department of Defense.

DOD DOT&E. 2004. Director of Operational Test and Evaluation (DOT&E). "Testing in a Joint Environment Roadmap: Strategic Planning Guidance, Fiscal Years 2006–2011." Final Report. 12 November

2004. Washington, D.C.: U.S. Department of Defense.

DoD DOT&E. 2010. Director of Operational Test and Evaluation (DOT&E). "Memorandum for Principal Deputy Under Secretary of Defense (Acquisition, Technology, and Logistics). Subject: State of Reliability." June 30, 2010.

Ferguson, C., L. Zimmermann, and F. Barone. 2007. "Joint Mission Environment Test Capability (JMETC) briefing for the Joint Strike Fighter PMO," 17 April 2007. Washington, D.C.: Test Resource Management Center (TRMC).

Ferguson, Thomas S. 1967. *Mathematical statistics: A decision theoretic approach*. New York: Academic Press.

46th Test Squadron. 2007. 46th Test Squadron Data Link Test Facility. "System of Systems Activity Report for FY 07." Draft. Unpublished. 7 December 2007.

JFIIT. 2007. Joint Fires Interoperability and Integration Team (JFIIT). "Coalition Combat Identification (CCID) Advanced Concept Technology Demonstration (ACTD) Bold Quest Close Air Support (CAS) assessment report." TR-07-06. October 2007. Elgin AFB, FL: JFIIT.

JITC. 2008. Joint C4ISR Interoperability Test & Evaluation (T&E) Capability (InterTEC) Web site. http://jitc.fu.disa.mil/intertec/index.html (accessed January 18, 2008).

JSSEO. 2006. Joint Single Integrated Air Picture System Engineering Organization (JSSEO). "Single Integrated Air Picture (SIAP) Test and Evaluation Strategy (TES)." Technical Report 2006-018. 3 May 2006. Arlington, VA: JSSEO.

Kometer, Michael W., Lt. Col., USAF. 2007. *Command in air war: Centralized vs. decentralized control of combat airpower*. Maxwell Air Force Base, Alabama: Air University Press.

Krieg, Kenneth J., David S. Chu, and Charles E. McQueary. 2006. Memorandum for Service Secretaries. Subject: Test and Training Interdependency Initiative. 7 September 2006. Washington, D.C.: DoD.

Porter, M., and V. Millar. 1985. How information gives you competitive advantage: The information revolution is transforming the nature of competition. *Harvard Business Review* 63 (4), 149–160.

TENA. 2008. Test and Training Enabling Architecture (TENA) Fact Sheet. https://www.tena-sda.org/download/attachments/6750/TENAFactSheet-IF07-2007-10-16.pdf?version=1 (accessed 18 January 2008).

TRMC. 2006. Test Resource Management Center (TRMC), Central Test and Evaluation Investment Program (CTEIP). *2006 Annual Report: Improving Test and Evaluation Capabilities*. http://www.acq.osd.mil/trmc/jipp/cteip/ (accessed December 14, 2010).

U.S. Air Force. 1995. AFMAN 99-112. *Electronic warfare test and evaluation process—direction and methodology for EW testing.* Air Force Manual. 27 March 1995. Washington, D.C.: Office of the Secretary of the Air Force.

U.S. Air Force. 2000. Chief of Staff. *Global vigilance, reach, and power: America's Air Force vision 2020*. Washington, D.C.: HQ USAF.

U.S. Air Force. 2004a. *Capabilities based test and evaluation.* Air Force Instruction (AFI) 99-103. 6 August 2004. Washington, D.C.: U.S. Department of Defense, http://www.af.mil/shared/media/epubs/AFI99-103.pdf (accessed December 14, 2010).

U.S. Air Force. 2004b. *Operation of the capabilities-based acquisition system.* Interim Guidance 63-101. Washington, D.C.: Office of the Assistant Secretary of the Air Force.

USJFCOM. 2010. United States Joint Forces Command (USJFCOM). Joint Fires Interoperability and Integration Team (JFIIT) Web page. https://jfiit.eglin.af.mil/public/index.php (accessed December 14, 2010).

Wiegand, Ronald C., Col., USAF. 2007. "USAFWC Capability for Net-centric Test and Training." Briefing to Test Resource Management Center. Washington, D.C., 19 December 2007.

Williams Phil. 2001. "Transnational Criminal Networks." In *Networks and Netwars*, ed. John Arquilla and David Ronfeldt, 61–97. Washington, D.C.: Rand.

Acknowledgments

The authors wish to thank Mr. Keith Poch (JMETC User Support Technical Team), Mr. Joe R. Siebenberg (605 TES), Mr. Robert M. Summitt (JFCOM JFIIT/J3), and Maj. Michael "Ponch" Tamburo (JDICE) for their collaboration in researching current efforts to develop SoS testing capabilities. Thanks also to Mr. Paul E. Witt, Air Force Special Operations Command/History Office (AFSOC/HO) for his help in researching the history of operational testing and to M.Sgt. Michelle A. Copenhaver (505 CCW/TA) for her editorial and proofreading support

Appendix – Decision theory and testing

A good text on the underlying mathematical theory is that of Thomas S. Ferguson, *Mathematical Statistics, A Decision Theoretic Approach* (Ferguson 1967).

The following are basic definitions:

- $\mathbf{\Theta} = \{\theta: \theta \text{ is a is a state of nature}\}$
- $\mathbf{A} = \{a: a \text{ is an available action}\}$

- X is a random variable representing an outcome of the statistical experiment and 1 is a realization of X.
- **S** = {s:s is a possible outcome of the statistical experiment}
- *d*(X) is a nonrandomized decision rule mapping **S** into **A**.
- L(θ,*d*(X)) is the loss function for the product space **Θ**x**A**.
- R(θ,*d*) = E [L(θ,*d*(X))] is the risk function, which is assumed to exist and be finite for all θ ε **Θ**, and the expected value is taken over X.
- $R(\theta,d) = E[L(\theta,d(X))] = \int L(\theta, d(x))dP_\theta(x)$

For traditional system-level operational testing, the states of nature could be considered dichotomous: the system meets requirements and performs satisfactorily (it is effective and suitable), or it does not. An alternative might be a determination that the system would contribute favorably to warfighting capability (with possible qualifications) versus one in which its impact on warfighting capability would not be favorable. The actual observed outcome of the test is an entry in a multi-dimensional matrix **S** representing the level of attainment of objectives identified by the test team. The random variable *X* is a function that maps **S** into the real line. A decision rule is devised (typically subjectively) to map *X* into an action (recommendation) vector ***A***, say ***A*** = [a_1, a_2, a_3, a_4], where

- a_1 is a recommendation to field the system as is,
- a_2 is a recommendation to field the system after identified deficiencies are corrected,
- a_3 is a recommendation not to field the system but to continue development, and
- a_4 is a recommendation not to field the system and to cease development.

The loss function represents the consequences of the ordered pair: the state of nature and the action taken based on the decision rule and observed value for *X*. The risk is the statistical expected value of the loss function in the form of a Lebesgue integral. For the purposes of this paper, the reader can consider $P_\theta(x)$ to be the cumulative distribution function for the random variable *X* when the state of nature is θ. A simplified example follows, where

- θ_1 = the state that the system essentially meets requirements;
- θ_2 = the state that the system is significantly deficient;
- x_1 = 1 (the system is observed to meet essential requirements during testing);
- x_2 = 0 (the system is observed to have significant deficiencies during testing);
- a_1 = action to field the system;
- a_2 = action to withhold fielding the system;

$$d_1(x): x_1 \to a_1 \quad d_2(x): x_1 \to a_1$$
$$x_2 \to a_2 \qquad\qquad x_2 \to a_1$$
$$d_3(x): x_1 \to a_2 \quad d_4(x): x_1 \to a_2$$
$$x_2 \to a_1 \qquad\qquad x_2 \to a_2$$

- $p_1(x_1) = P[X = x_1 \mid \theta = \theta_1]$;
- $p_1(x_2) = P[X = x_2 \mid \theta = \theta_1]$;
- $p_2(x_1) = P[X = x_1 \mid \theta = \theta_2]$; and
- $p_2(x_2) = P[X = x_2 \mid \theta = \theta_2]$.

Assume that

$$p_1(x_1) = 2/3 \quad \text{and} \quad p_1(x_2) = 1/3 \quad \text{when} \quad \theta = \theta_1;$$
$$p_2(x_1) = 1/4 \quad \text{and} \quad p_2(x_2) = 3/4 \quad \text{when} \quad \theta = \theta_2$$

$$L(\theta_1, a_1) = -600, L(\theta_1, a_2) = 720, L(\theta_2, a_1) = 900,$$
$$L(\theta_2, a_2) = -300$$

(Note that $L(\theta_1, a_2)$ is the opportunity loss corresponding to a Type I Error in a statistical test of hypothesis, and $L(\theta_2, a_1)$ is the loss associated with a Type II Error in a statistical test of hypothesis.) The tester is assumed to make recommendations based solely on the decision rule *d*, not on exogenous considerations. Choose $d(x) = d_1(x)$.

Then,

$$\begin{aligned} R(\theta_1, d_1) &= E[L(\theta_1, d_1(X))] \\ &= L(\theta_1,a_1)p_1(x_1) + L(\theta_1,a_2)p_1(x_2) \\ &= -160 \end{aligned}$$

$$\begin{aligned} R(\theta_2, d_1) &= E[L(\theta_2, d_1(X))] \\ &= L(\theta_2,a_1)p_2(x_1) + L(\theta_2,a_2)p_2(x_2) \\ &= 0 \end{aligned}$$

Determination of $R(\theta,d)$ for $d_2(x)$, $d_3(x)$, and $d_4(x)$ is straightforward. Based on a complete evaluation of $R(\theta,d)$, the investigator can develop an appropriate strategy for the decision problem.