ITEA Journal 2011; 32: 15–17

Continuing the Emphasis on Scientific Rigor in Test and Evaluation

Catherine Warner, Ph.D.
Science Advisor, Office of the Director, Operational Test and Evaluation, The Pentagon, Washington, D.C.

The Director, Operational Test & Evaluation has a new Science Advisor for the first time since 1988. Dr. Catherine Warner, formerly a research staff member at the Institute for Defense Analyses, is now the Science Advisor to the Director, Operational Test & Evaluation. She is working closely with the Honorable Dr. Michael Gilmore on his initiatives for Test and Evaluation (T&E). Her goals for the Science Advisor's role include continuing the emphasis on reliability growth and tracking during T&E, increasing the emphasis on scientific test design and statistical rigor in T&E, improving the analytical capabilities of the Department of Defense T&E workforce, developing initiatives to support integrated testing, and collaborating with the newly recreated Developmental Test and Evaluation office to ensure rigorous testing of all systems for our Service members.

This is my first opportunity since being appointed as the Science Advisor to the Director, Operational Test & Evaluation (DOT&E) to address the testing community through *The ITEA Journal*. I look forward to many future discussions, articles, and interactions with the test community through *The ITEA Journal*.

Catherine Warner

Let me begin by introducing myself. I have been involved with operational test and evaluation since 1991, when I became a research staff member at the Institute for Defense Analyses (IDA). At IDA, I performed and directed analysis of operational tests for U.S. Army, Navy, and Air Force systems in support of DOT&E. My initial responsibilities included the evaluation of unmanned aerial vehicle (UAV) systems including the Hunter Short Range UAV system, the Predator Medium Altitude Endurance UAV Advanced Concept Technology Development (ACTD) system, the Global Hawk UAV ACTD system, and the Shadow Tactical UAV system. As part of my work with UAV systems, I was involved with the development of evaluation concepts for intelligence, surveillance, and reconnaissance (ISR) systems. The difficulty in evaluating ISR system effectiveness is quantitatively assessing the product or information the system provides. Our team at IDA developed mission success templates that allowed a quantitative assessment of the timeliness, accuracy, and completeness based on the requestor's needs and ground truth. Also as part of my work on UAV systems, I was involved with the Army's Training and Doctrine Command (TRADOC) Analysis Center at White Sands Missile Range in the development of operational test visualization tools. These computer-based tools allowed the playback of operational test events over terrain maps and could be used for determining line-of-sight between an airborne platform and a ground target of interest. This capability was extremely useful in the analyses of the Joint Surveillance Target Attack Radar (JSTARS) operational testing.

More recently, I was the assistant director of IDA's Operational Evaluation Division and led their air warfare group. My analysis portfolio included major aircraft systems such as the F-22, V-22, F/A-18E/F, and H-1 Upgrades programs. In this role, I worked closely with U.S. Air Force, Navy, and Marine Corps operational evaluators. I coordinated IDA's support for one of DOT&E's largest efforts observing the operational test of the F-22. I also provided analytical support for DOT&E's reports to Congress on the operational test and evaluation of the F/A-18E/F APG-79 (AESA) radar upgrade, the V-22 Osprey (both MV-22 and CV-22 systems), and the H-1 Upgrades (UH-1Y and AH-1Z).

I learned many lessons during 19 years of planning, observing, and evaluating operational tests as an IDA analyst. My experiences have taught me the importance

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

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

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Continuing the Emphasis on Scientific Rigor in Test and Evaluation	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) **Office of the Director,Operational Test and Evaluation,Washington,DC,20301**	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

14. ABSTRACT

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)**	**3**	

of test planning, including the formulation of the evaluation plan and achieving a shared agreement of the plan among the stakeholders—perhaps the primary and most difficult step in the testing process. Next, accurately collecting and documenting data form the foundation for the analyses that follow. The ability to preview data collection tools and databases greatly reduces the risk of ending a test without much documented evidence. Finally, choosing the appropriate evaluation tools helps clarify results for the researcher, decision maker, and end user; using the proper methodology can help find the difference that *makes a difference*. I bring these experiences with me now as the Science Advisor to DOT&E.

As Science Advisor to DOT&E, I provide technical and policy advice on all matters of operational test and evaluation within the Department of Defense (DoD). I will identify future trends and needs and implement new test and analysis techniques—continually guiding and shaping the scientific focus of DOT&E policies and interests. I am a technical resource for the DOT&E staff and a liaison between DOT&E and the T&E community. I look forward to picking up where my predecessor, Dr. Ernest Seglie, left off.

In November 2009, Dr. Michael Gilmore, Director, Operational Test & Evaluation, authored a memorandum outlining his initiatives for Operational Test and Evaluation. He published these initiatives in the June 2010 *ITEA Journal*. They are as follows:

1. Field New Capability Rapidly;
2. Engage Early to Improve Requirements;
3. Integrate Developmental, Live Fire, and Operational Testing; and
4. Substantially Improve Suitability Before Initial Operational Test and Evaluation (IOT&E).

As the Science Advisor to DOT&E, I am already working to support these initiatives. Going forward, I will provide technical expertise for the programmatic and technical areas upon which DOT&E should focus to ensure proper support for future testing requirements. In particular, my initial priorities are to improve reliability planning, tracking, and assessment methodologies, integrate developmental, live fire, and operational testing, and increase scientific-based test design within the DoD.

The importance of reliability and maintainability on life cycle costs has gained traction throughout DoD. My predecessor commissioned many studies on the subject, and these studies indicate that there are solid returns on investments made for reliability from about seven to one and up to approximately fifty to one if done early in the life cycle. For this reason, DOT&E has placed a great deal of emphasis on reliability in operational testing and test planning. Over the past two years, DOT&E and the Undersecretary of Defense for Acquisition, Technology and Logistics (USD[AT&L]) have worked together to lead improvements in system reliability and reliability policy. Examples include the following:

- updated policy in DoDI 5000.02 to require reliability growth;
- approved a new industry standard for best practices for reliability, the ANSI/GEIA-STD-0009;
- published sample Request For Proposal (RFP) and contract language to assure reliability growth is incorporated in system design and development contracts;
- updated the DoD *Reliability, Availability, Maintainability & Cost (RAM-C) Manual;*
- sponsored development of the Reliability Investment Model; and
- began drafting the Reliability Program Handbook, HB-0009.

The Weapon Systems Acquisition Reform Act (WSARA) of 2009 also added emphasis on reliability by specifying reliability responsibilities for the Director of Systems Engineering and the Director of Developmental Test and Evaluation (DDT&E). However, more remains to be done. To further improve reliability, the Director and I are working with the USD(AT&L) to strengthen and clarify Department policy. I plan to work with the Department to upgrade educational programs and to introduce more rigor and objectivity into planning reliability test programs. I am working with the Army Test and Evaluation Command (ATEC) and Material Systems Analysis Activity (AMSAA) to make their extensive efforts in reliability growth and monitoring available to the T&E community at large. We have participated together on a Government Accountability Office panel on reliability. I have invited AMSAA to provide in-house training to DOT&E and DT&E action officers. I hope that by extending this knowledge to the T&E Integrated Product Teams (IPTs) and Program Offices at large that reliability planning and tracking will continue to increase.

Another of the Director's initiatives, which was echoed in the WSARA 2009, is to use integrated testing. Integrated testing encourages all testers (contractor, developmental, operational, and live fire) to plan together the test program, seeking an efficient continuum—eliminating redundant processes and products. Although each test type has a different

objective, data from each test can provide insight for others. Our goal is to have an efficient and adequate test program that is not duplicative. We want to test early in the mission context and in realistic operational environments, even for component testing, to discover problems early. Evaluators must plan to use all test data to support their evaluations to the extent possible. But dedicated OT is still required.

Our test resources are limited, in many cases we do not have the forces available to field a complete capability for test purposes, and sometimes our test ranges are not large enough to contain the full-scale test. We have limited time available for testing and few test articles, either because of cost or the time to produce them. To overcome these constraints, we must use statistical tools. Stochastic simulations provide synthetic forces to supplement operational units and also supplement field tests for conditions that cannot be replicated in the field. Statistical methods also facilitate the performance assessment of systems when only small samples are available.

To deal with many of the foregoing testing constraints, we are promoting the use of Design of Experiments. Design of Experiments is a structured, rigorous statistical tool for test planning and analysis. We are working to make Design of Experiments a common tool for test planning, execution, and evaluation. In May 2009, DOT&E and the OT agencies signed a joint letter endorsing the application of Design of Experiments. While this subject has been extensively written about in an academic setting, there are still many questions about how it applies to T&E within DoD. DOT&E and DDT&E have already engaged in joint training on Design of Experiments and are discussing approaches to Developmental Testing – Operational Testing (DT-OT) integration. By the time this article appears in *The ITEA Journal*, I will have traveled to all of the operational test agencies to discuss these issues. With DT&E, the operational test agencies, and the Service academies, I am forming a steering group to develop a roadmap for the institutionalization of scientific test design and statistical rigor. The steering group will assess the current state of analytic capabilities within each of the Services and the Office of the Secretary of Defense (OSD). This group will also develop options for training and other support that Services and agencies will need to increase the rigor of test design and analytic capabilities to desired levels. In addition to the steering group, I also plan to form a permanent advisory board to act as a resource on future methods for incorporating statistical rigor and test science.

Parallel to the formation of a permanent advisory board, I am working to form research relationships with academic institutions that have excelled in the fields of statistics, experimental design, T&E, and systems engineering. We have some initial work currently ongoing at the Naval Postgraduate School and hope to further collaborate with the Air Force Institute of Technology and other institutions such as Arizona State University and Virginia Tech.

In summary, my role as the Science Advisor at DOT&E is to help ensure that our Service members have high confidence in their systems, know what their systems can and cannot do, and know that they will work when needed. In addition to rigorously and objectively evaluating these systems, we must make all of our OT&E results more readily available to the end users of the equipment. DOT&E reports are now available through Defense Technical Information Center and the Congressional Research Service. We also have established points of contact within each combatant command, and we have a classified Web site that is accessible throughout DoD. However, I am exploring other options to ensure that the information is accessible not only to the decision makers but to the fighting forces as well.

I look forward to working with the test community in continuing the emphasis on scientific rigor in T&E and ensuring we have the best methods in place, so that the acquisition community can provide the best possible systems to our Service members. ❑

CATHERINE WARNER, PH.D., became the Science Advisor for DOT&E on September 13, 2010. She serves as the technical advisor to the Director on all matters of testing and evaluation in the DoD. Previously, Dr. Warner was an assistant director and head of the Air Warfare group for the Operational Evaluation Division at the Institute for Defense Analyses (IDA). She managed a team of project leaders supporting the DOT&E Deputy Director for Air Warfare and provided technical support as needed to the Director, OT&E for special interest items. Her analysis portfolio included major aircraft systems such as the F-22, F/A-18E/F, V-22, and H-1 Upgrades. She also evaluated unmanned aerial vehicles such as the Global Hawk, Predator, Shadow, and Hunter UAV systems. Earlier, Dr. Warner worked at the Lawrence Livermore National Laboratory in the laser materials group and as a research chemist at IBM Corporation in San Jose, California. Dr. Warner grew up in Albuquerque, New Mexico, attended the University of New Mexico as an undergraduate, and earned both bachelor of science and master of science degrees in chemistry from San Jose State University. She earned both master of arts and doctor of philosophy degrees in chemistry from Princeton University. E-mail: catherine.warner@osd.mil