OFFICE OF THE SECRETARY OF DEFENSE
1700 DEFENSE PENTAGON
WASHINGTON, DC 20301-1700

OPERATIONAL TEST AND EVALUATION

FEB 27 2012

MEMORANDUM FOR USERS OF THE DOT&E TEMP GUIDEBOOK

SUBJECT: DOT&E TEMP Guidebook

Threat Representation
Guidance
Example for ¶ 1.3.1

My staff has modified this TEMP Guide Annex from the Defense Acquisition Guidebook to assist in preparation of Test and Evaluation Master Plans. In this version of the TEMP Guide, you will find **callouts** like the one to the right with links to policy guidance and examples of particular interest to DOT&E.

The callouts have been placed throughout this TEMP Guide Annex at those locations where DOT&E and other applicable policies apply. Most of the callouts also include links to notional examples to give you an idea of the level of detail expected in the TEMP. Keep in mind that the examples are notional and apply to a specific system, not to every system. In preparing your TEMP, you should follow the policy guidance and not simply copy the examples provided. The examples might not be appropriate for your system. The policy guidance contains additional links to the source policy documents if you wish to further investigate or understand the policy.

You will also find that the structure of the TEMP has been modified slightly to accommodate the latest guidance on information assurance, design of experiments, and reliability growth. The importance of these three topics requires those who develop the test and evaluation strategy to consider them from the very beginning to the end of program development. Accordingly, you should outline the strategy for these topics in the beginning of the TEMP and provide details in subsequent TEMP sections as appropriate. This guide indicates where and how these and other topics should be addressed.

Questions or suggestions about this guidebook should be addressed to Dr. Catherine Warner, Catherine.Warner@osd.mil, 703-697-3655.

J. Michael Gilmore
Director

Attachment:
As stated

Defense Acquisition Guidebook ANNEX

TEST AND EVALUATION MASTER PLAN

FOR

PROGRAM TITLE/SYSTEM NAME

ACRONYM

ACAT Level

Program Elements

Xxxxx

****

SUBMITTED BY

______________________________	__________
Program Manager	DATE

CONCURRENCE

______________________________	__________
Program Executive Officer or Developing Agency (if not under the Program Executive Officer structure)	DATE
______________________________	__________
Operational Test Agency	DATE
______________________________	__________
User's Representative	DATE

DoD COMPONENT APPROVAL

______________________________	__________
DoD Component Test and Evaluation Director	DATE
______________________________	__________
DoD Component Acquisition Executive (Acquisition Category I) Milestone Decision Authority (for less-than-Acquisition Category I)	DATE

Note: For Joint/Multi Service or Agency Programs, each Service or Defense Agency should provide a signature page for parallel staffing through its CAE or Director, and a separate page should be provided for OSD Approval

****

OSD APPROVAL

______________________________	__________
Deputy Assistant Secretary of Defense for Developmental Test and Evaluation	DATE
______________________________	__________
Director, Operational Test and Evaluation	DATE

TABLE OF CONTENTS

1. PART I - INTRODUCTION

1.1. Purpose. State the purpose of the Test and Evaluation Master Plan (TEMP). Identify if this is an initial or updated TEMP. State the Milestone (or other) decision the TEMP supports. Reference and provide hyperlinks to the documentation initiating the TEMP (i.e., Initial Capability Document (ICD), Capability Development Document (CDD), Capability Production Document (CPD), Acquisition Program Baseline (APB), Acquisition Strategy Report (ASR), Concept of Operations (CONOPS)). State the Acquisition Category (ACAT) level, operating command(s), and if listed on the OSD T&E Oversight List (actual or projected)

1.2. Mission Description. Briefly summarize the mission need described in the program capability requirements documents in terms of the capability it will provide to the Joint Forces Commander. Describe the mission to be accomplished by a unit equipped with the system using all applicable CONOPS and Concepts of Employment. Incorporate an OV-1 of the system showing the intended operational environment. Also include the organization in which the system will be integrated as well as significant points from the Life Cycle Sustainment Plan, the Information Support Plan, and Program Protection Plan. Provide links to each document referenced in the introduction. For business systems, include a summary of the business case analysis for the program.

1.3. System Description. Describe the system configuration. Identify key features and subsystems, both hardware and software (such as architecture, system and user interfaces, security levels, and reserves) for the planned increments within the Future Years Defense Program (FYDP).

1.3.1. **System Threat Assessment**. Succinctly summarize the threat environment (**to include cyber-threats**) in which the system will operate. Reference the appropriate DIA or component-validated threat documents for the system.

Threat Representation

Guidance

Example for ¶ 1.3.1

Information Assurance

Guidance

Examples for ¶ 1.3.1

1.3.2. Program Background. Reference the Analysis of Alternatives (AoA), the APB and the materiel development decision to provide background information on the proposed system. Briefly describe the overarching Acquisition Strategy (for space systems, the Integrated Program Summary (IPS)), and the Technology Development Strategy (TDS). Address whether the system will be procured using an incremental development strategy or a single step to full capability. If it is an evolutionary acquisition strategy, briefly discuss planned upgrades, additional features and expanded capabilities of follow-on increments. The main focus must be on the current increment with brief descriptions of the previous and follow-on increments to establish continuity between known increments.

1.3.2.1. Previous Testing. Discuss the results of any previous tests that apply to, or have an effect on, the test strategy.

1.3.3. Key Capabilities. Identify the Key Performance Parameters (KPPs) and Key System Attributes (KSAs) for the system. For each listed parameter, provide the threshold and objective values from the CDD/CPD and reference the paragraph.

1.3.3.1. Key Interfaces. Identify interfaces with existing or planned systems' architectures that are required for mission accomplishment. Address integration and modifications needed for commercial items. Include interoperability with existing and/or planned systems of other Department of Defense (DoD) Components, other Government agencies, or Allies. Provide a diagram of the appropriate DoD Architectural Framework (DoDAF) system operational view from the CDD or CPD.

Information Assurance

Guidance

Examples for ¶ 1.3.3.2

1.3.3.2. Special test or certification requirements. Identify unique system characteristics or support concepts that will generate special test, analysis, and evaluation requirements (e.g., security test and evaluation and **Information Assurance (IA) Certification and Accreditation (C&A)**, post deployment software support, resistance to chemical, biological, nuclear, and radiological effects; resistance to countermeasures; resistance to reverse engineering/exploitation efforts (Anti-Tamper); **development of new threat simulation, simulators, or targets.**

Threat Representation

Guidance

Example for ¶ 1.3.3.2

1.3.3.3. Systems Engineering (SE) Requirements. Reference all SE-based information that will be used to provide additional system evaluation targets driving system development. Examples could include hardware reliability growth sand software maturity growth strategies. The SEP should be referenced in this section and aligned to the TEMP with respect to SE Processes, methods, and tools identified for use during T&E.

2. PART II – TEST PROGRAM MANAGEMENT AND SCHEDULE

2.1 T&E Management. Discuss the test and evaluation responsibilities of all participating organizations (such as developers, testers, evaluators, and users). Describe the role of contractor testing in early system development. Describe the role of government developmental testers to assess and evaluate system performance. Describe the role of the Operational Test Agency (OTA) /operational testers to confirm operational effectiveness, operational suitability and survivability.

2.1.1. T&E Organizational Construct. Identify the organizations or activities (such as the T&E Working-level Integrated Product Team (WIPT) or Service equivalent, LFT&E IPT, etc.) in the T&E management structure, to include the sub-work groups, such as a modeling & simulation, or reliability. Provide sufficient information to adequately understand the functional relationships. Reference the T&E WIPT charter that includes specific responsibilities and deliverable items for detailed explanation of T&E management. These items include TEMPs and Test Resource Plans (TRPs) that are produced collaboratively by member organizations.

2.2. Common T&E Database Requirements. Describe the requirements for and methods of collecting, validating, and sharing data as it becomes available from the contractor, Developmental Test (DT), Operational Test (OT), and oversight organizations, as well as supporting related activities that contribute or use test data (e.g., information assurance C&A, interoperability certification, etc.). Describe how the pedigree of the data will be established and maintained. The pedigree of the data refers to understanding the configuration of the test asset, and the actual test conditions under which the data were obtained for each piece of data. State who will be responsible for maintaining this data.

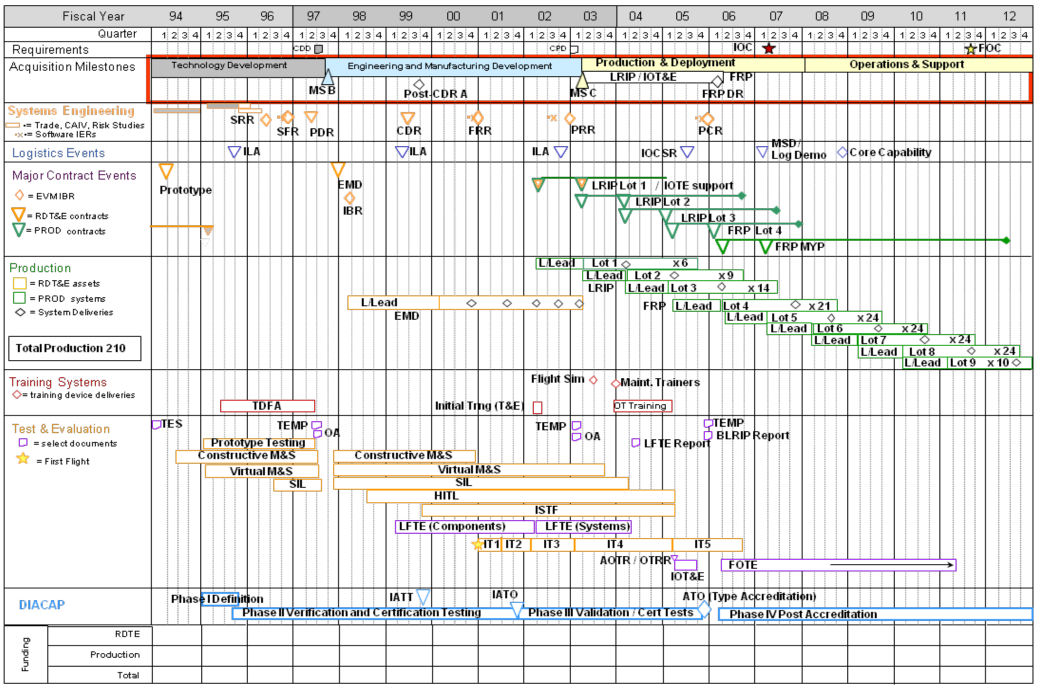
2.3. Deficiency Reporting. Briefly describe the processes for documenting and tracking deficiencies identified during system development and testing. Describe how the information is accessed and shared across the program. The processes should address problems or deficiencies identified during both contractor and government test activities. The processes should also include issues that have not been formally documented as a deficiency (e.g., watch items).

2.4. TEMP updates. Reference instructions for complying with DoDI 5000.02 required updates or identify exceptions to those procedures if determined necessary for more efficient administration of document. Provide guidelines for keeping TEMP information current between updates. For a Joint or Multi-Service TEMP, identify references that will be followed or exceptions as necessary.

2.5. Integrated Test Program Schedule. Display (see Figure 2.1) the overall time sequencing of the major acquisition phases and milestones (as necessary, use the NSS-03-01 time sequencing). Include the test and evaluation major decision points, related activities, and planned cumulative funding expenditures by appropriation by year. Include event dates such as major decision points as defined in DoD Instruction 5000.02, e.g., operational assessments, preliminary and critical design reviews, test article availability; software version releases; appropriate phases of DT&E; LFT&E; Joint Interoperability Test Command (JITC) interoperability testing and certification date to support the MS-C and Full-Rate Production (FRP) Decision Review (DR). Include significant Information Assurance certification and accreditation event sequencing, such as Interim

Authorization to Test (IATT), Interim Authorization to Operate (IATO) and Authorization to Operate (ATO). Also include operational test and evaluation; Low-Rate Initial Production (LRIP) deliveries; Initial Operational Capability (IOC); Full Operational Capability (FOC); and statutorily required reports such as the Live-Fire T&E Report and Beyond Low-Rate Initial Production (B-LRIP) Report. Provide a single schedule for multi-DoD Component or Joint and Capstone TEMPs showing all related DoD Component system event dates.

Figure 2.1 SAMPLE Integrated Program Test Schedule



3. PART III – TEST AND EVALUATION STRATEGY

3.1 T&E Strategy. Introduce the program T&E strategy by briefly describing how it supports the acquisition strategy as described in Section 1.3.2. This section should summarize an effective and efficient approach to the test program. The developmental and operational test objectives are discussed separately below; however this section must also address how the test objectives will be integrated to support the acquisition strategy by evaluating the capabilities to be delivered to the user without compromising the goals of each major kind of test type. Where possible, the discussions should focus on the testing for capabilities, and address testing of subsystems or components where they represent a significant risk to achieving a necessary capability. As the system matures and production representative test articles are available, the strategy should address the conditions for integrating DT and OT tests. Evaluations shall include a **comparison with current mission capabilities** using existing data, so that measurable improvements can be determined. If such evaluation is considered costly relative to the benefits gained, the PM shall propose an alternative evaluation strategy. Describe the strategy for achieving this comparison and for ensuring data are retained and managed for future comparison results of evolutionary increments or future replacement capabilities. To present the

Integrated Testing

Guidance and Best Practices

Information Assurance

Guidance

Examples for ¶ 3.1

Baseline Evaluation

Guidance and Best Practices

program's T&E strategy, briefly describe the relative emphasis on methodologies (e.g., Modeling and Simulation (M&S), Measurement Facility (MF), Systems Integration Laboratory (SIL), Hardware-In-the-Loop Test (HILT), Installed System Test Facility (ISTF), Open Air Range (OAR)).

3.2. **Evaluation Framework**. Describe the overall evaluation approach focusing on key decisions in the system lifecycle and addressing key system risks, program unique Critical Operational Issues (COIs) or Critical Operational Issue Criteria (COIC), and Critical Technical Parameters (CTPs). Specific areas of evaluation to address are related to the:

(1) Development of the system and processes (include maturation of system design)

(2) System performance in the mission context

(3) OTA independent assessments and evaluations

(4) Survivability and/or lethality

(5) Comparison with existing capabilities, and

(6) Maturation of highest risk technologies

(7) **Reliability Growth** This paragraph has been moved forward from paragraph 3.8 of the DAG guidebook. Reliability growth should be integrated into the T&E strategy and explained as part of the Evaluation Framework.

Since reliability is a driver during system development, identify, in tabular form, the amount of operating time being accrued during the each of the tests listed in the Figure 2.1. Table should contain the system configuration, operational concept, etc. Reference and provide hyperlinks to the reliability growth planning document. (Moved from Para 3.8)

(8) **Design of Experiments**. This is a new paragraph added to the DAG guidebook. Design of Experiments is integral to the Evaluation Framework and begins with selection of evaluation metrics in Figure 3.1. In this paragraph, provide an overview of the experimental design and attach Appendix D, with the details of the design. See links at the right for general DOE guidance and examples.

Describe any related systems that will be included as part of the evaluation approach for the system under test (e.g., **data transfer, information exchange requirements,** interoperability requirements, and documentation systems). Also identify any configuration differences between the current system and the system

Mission-Focused Metrics

Guidance

Reliability Growth

Guidance

Reliability Growth Example

Software Tracking Example

DOE

Guidance

TEMP Body Examples

Missile Example Appendix

Artillery Example Appendix

Software Example Appendix

Information Assurance

Guidance

Examples for Figure 3.1

to be fielded. Include mission impacts of the differences and the extent of integration with other systems with which it must be interoperable or compatible. Describe how the system will be evaluated and the sources of the data for that evaluation. The discussion should address the key elements for the evaluations, including major risks or limitations for a complete evaluation of the increment undergoing testing. The reader should be left with an understanding of the value-added of these evaluations in addressing both programmatic and warfighter decisions or concerns. This discussion provides rationale for the major test objectives and the resulting major resource requirements shown in Part IV - Resources.

Include a Top-Level Evaluation Framework matrix that shows the correlation between the KPPs/KSAs, CTPs, key test measures (i.e., Measures of Effectiveness (MOEs) and Measures of Suitability (MOSs)), planned test methods, and key test resources, facility or infrastructure needs. When structured this way, the matrix should describe the most important relationships between the types of testing that will be conducted to evaluate the Joint Capabilities Integration and Development System (JCIDS)-identified KPPs/KSAs, and the program's CTPs. Figure 3.1 shows how the Evaluation Framework could be organized. Equivalent Service-specific formats that identify the same relationships and information may also be used. The matrix may be inserted in Part III if short (less than one page), or as an annex. The evaluation framework matrix should mature as the system matures. Demonstrated values for measures should be included as the acquisition program advances from milestone to milestone and as the TEMP is updated.

The suggested content of the evaluation matrix includes the following:

- **Key requirements & T&E measures** – These are the KPPs and KSAs and the top-level T&E issues and measures for evaluation. The top-level T&E issues would typically include COIs/Critical Operational Issues and Criteria (COICs), CTPs, and key MOEs/MOSs. System-of-Systems and technical review issues should also be included, either in the COI column or inserted as a new column. Each T&E issue and measure should be associated with one or more key requirements. However, there could be T&E measures without an associated key requirement or COI/COIC. Hence, some cells in figure 3.1 may be empty.
- Overview of test methodologies and key resources – These identify test methodologies or key resources necessary to generate data for evaluating the COIs/COICs, key requirements, and T&E measures. The content of this column should indicate the methodologies/resources that will be required and short notes or pointers to indicate major T&E phases or resource names. M&S should be identified with the specific name or acronym.
- Decisions Supported – These are the major design, developmental, manufacturing, programmatic, acquisition, or employment decisions most affected by the knowledge obtained through T&E.

Figure 3.1, Top-Level Evaluation Framework Matrix

Key Requirements and T&E Measures				Test Methodologies/Key Resources (M&S, SIL, MF, ISTF, HITL, OAR)	Decision Supported
Key Reqs	**COIs**	**Key MOEs/ MOSs**	**CTPs & Threshold**		
KPP#1:	**COI #1.** Is the XXX effective for…	**MOE 1.1.**	Engine thrust	Chamber measurement Observation of performance profiles OAR	PDR CDR
	COI #2. Is the XXX suitable for…		Data upload time	Component level replication Stress and Spike testing in SIL	PDR CDR
	COI #3. Can the XXX be…	**MOS 2.1.**			
		MOE 1.3.			FRP
		MOE 1.4.	**Reliability based on growth curve**	Component level stress testing Sample performance on growth curve Sample performance with M&S augmentation	PDR CDR MS-C
KPP #2		**MOS 2.4.**	Data link		MS-C SRR
KPP #3	**COI #4.** Is training….	**MOE 1.2.**		Observation and Survey	MS-C FRP
KSA #3.a	**COI #5.** Documentation	**MOS 2.5.**			MS-C FRP

Reliability Growth

Guidance

Example for Figure 3.1

3.3. Developmental Evaluation Approach. Describe the top-level approach to evaluate system and process maturity, as well as, system capabilities and limitations expected at acquisition milestones and decision review points. The discussion should include logistics, reliability growth, and system performance aspects. Within this section, also discuss:

1) rationale for CTPs (see below for a description of how to derive CTPs),
2) key system or process risks,
3) any certifications required (e.g. weapon safety, interoperability, spectrum approval, **information assurance**),
4) any technology or subsystem that has not demonstrated the expected level of technology maturity at level 6 (or higher), system performance, or has not achieved the desired mission capabilities for this phase of development,
5) degree to which system hardware and software design has stabilized so as to determine manufacturing and production decision uncertainties,
6) key issues and the scope for logistics and sustainment evaluations, and
7) reliability thresholds when the testing is supporting the system's reliability growth curve.

Information Assurance

Guidance

Example for ¶ 3.3

CTPs are measurable critical system characteristics that, if not achieved, preclude the fulfillment of desired operational performance capabilities. While not user requirements, CTPs are technical measures derived from desired user capabilities. Testers use CTPs as reliable indicators that the system is on (or behind) the planned development schedule or will likely (or not likely) achieve an operational capability. Limit the list of CTPs to those that support the COIs. Using the system specification as a reference, the chief engineer on the program should derive the CTPs to be assessed during development.

Mission-Oriented Evaluation

Guidance

Examples

3.3.1. ***Mission-Oriented Approach.*** Describe the approach to evaluate the system performance in a mission context during development in order to influence the design, manage risk, and predict operational effectiveness and operational suitability. A mission context focuses on how the system will be employed. Describe the rationale for the COIs or COICs.

3.3.2. Developmental Test Objectives. Summarize the planned objectives and state the methodology to test the system attributes defined by the applicable capability requirement document (CDD, CPD, CONOPs) and the CTPs that will be addressed during each phase of DT as shown in Figure 3.1, Top-Level Evaluation Framework matrix and the Systems Engineering Plan. Subparagraphs can be used to separate the discussion of each phase. For each DT phase, discuss the key test objectives to address both the contractor and government developmental test concerns and their importance to achieving the exit criteria for the next major program decision point. If a contractor is not yet selected, include the developmental test issues addressed in the Request For Proposals (RFPs) or Statement of Work (SOW). Discuss how developmental testing will reflect the expected operational environment to help ensure developmental testing is planned to integrate with operational testing. Also include key test objectives related to logistics testing. All objectives and CTPs should be traceable in the Top-Level Evaluation Framework matrix to ensure all KPPs/KSAs are addressed, and that the COIs/COICs can be fully answered in operational testing. Summarize the developmental test events, test scenarios, and the test design concept. Quantify the testing sufficiently (e.g., number of test hours, test articles, test events, test firings) to allow a valid cost estimate to be created. Identify and explain how models and simulations, specific threat systems, surrogates, countermeasures, component, or subsystem testing, test beds, and prototypes will be used to determine whether or not developmental test objectives are achieved. Identify the DT&E reports required to support decision points/reviews and OT readiness. Address the system's *reliability growth strategy*, goals, and targets and how they support the Evaluation Framework. Detailed developmental test objectives should be addressed in the **System Test Plans and detailed test plans**.

Modeling and Simulation

Guidance

Examples

3.3.3. **Modeling & Simulation** (M&S). Describe the key models and simulations and their intended use. Include the developmental test objectives to be addressed using M&S to include any approved operational test objectives. Identify data needed and the planned accreditation effort. Identify how the developmental test scenarios will be supplemented with M&S, including how M&S will be used to predict the Sustainment KPP and other sustainment considerations. Identify who will perform M&S verification, validation, and accreditation. Identify developmental M&S resource requirements in Part IV.

DT Test Limitations

Guidance

Example for ¶ 3.3.3

3.3.4. **Test Limitations**. Discuss any developmental test limitations that may significantly affect the evaluator's ability to draw conclusions about the maturity, capabilities, limitations, or readiness for dedicated operational testing. Also address the impact of these limitations, and resolution approaches.

Integrated Survivability Evaluation

Guidance

3.4. **Live Fire Test and Evaluation Approach**. If live fire testing is required, describe the approach to evaluate the survivability/lethality of the system, and (for survivability LFT&E) **personnel survivability** of the system's occupants. Include a description of the overall live fire evaluation strategy to influence the system design (as defined in Title 10 U.S.C. § 2366), critical live fire evaluation issues, and major evaluation limitations. Discuss the management of the LFT&E program, to include the shot selection process, target resource availability, and schedule. Discuss a waiver, if appropriate, from full-up, system-level survivability testing, and the alternative strategy.

Force Protection

Guidance

Test Plan Review and Approval

Guidance

Example

3.4.1. Live Fire Test Objectives. State the key live fire test objectives for realistic survivability or lethality testing of the system. Include a matrix that identifies all tests within the LFT&E strategy, their schedules, the issues they will address, and which **planning documents will be submitted for DOT&E approval and which will be submitted for information and review only.** Quantify the testing sufficiently (e.g., number of test hours, test articles, test events, test firings) to allow a valid cost estimate to be created.

3.4.2. **Modeling & Simulation** (M&S). Describe the key models and simulations and their intended use. Include the LFT&E test objectives to be addressed using M&S to include operational test objectives. Identify data needed and the planned accreditation effort. Identify how the test scenarios will be supplemented with M&S. Identify who will perform M&S verification, validation, and accreditation. Identify M&S resource requirements in Part IV

Modeling and Simulation for LFT&E

Guidance

Examples

3.4.3. **Test Limitations**. Discuss any test limitations that may significantly affect the ability to assess the system's vulnerability and survivability. Also address the impact of these limitations, and resolution approaches.

LFT&E Limitations

Guidance

Example for ¶ 3.4.3

3.5. **Certification for Initial Operational Test and Evaluation (IOT&E).** Explain how and when the system will be certified safe and ready for IOT&E. Explain who is responsible for certification and which decision reviews will be supported using the lead Service's certification of safety and system materiel readiness process. List the DT&E information (i.e., reports, briefings, or summaries) that provides predictive analyses of expected system performance against specific COIs and the key system attributes - MOEs/MOSs. Discuss the entry criteria for IOT&E and how the DT&E program will address those criteria.

IOT&E Entrance Criteria

Guidance

Examples

Test Plan Review and Approval

Guidance

Examples

3.6. **Operational Evaluation Approach**. Describe the approach to conduct the independent evaluation of the system. Identify the periods during integrated testing that may be useful for operational assessments and evaluations. Outline the approach to conduct the dedicated IOT&E and resolution of the COIs. COIs must be relevant to the required capabilities and of key importance to the system being operationally effective, operationally suitable and survivable, and represent a significant risk if not satisfactorily resolved. A COI/COIC is typically phrased as a question that must be answered in the affirmative to properly evaluate operational effectiveness (e.g., "Will the system detect the threat in a combat environment at adequate range to allow successful engagement?") and operational suitability (e.g., "Will the system be safe to operate in a combat environment?"). COIs/COICs are critical elements or operational mission objectives that must be examined. COIs/COICs should be few in number and reflect total operational mission concerns. Use existing documents such as capability requirements documents, Business Case Analysis, AoA, APB, war fighting doctrine, validated threat assessments and CONOPS to develop the COIs/COICs. COIs/COICs must be formulated as early as possible to ensure developmental testers can incorporate mission context into DT&E. If every COI is resolved favorably, the system should be operationally effective and operationally suitable when employed in its intended environment by typical users.

3.6.1. **Operational Test Objectives**. State the key MOEs/MOSs that support the COIs/COICs. Ensure the operational tests can be identified in a way that allows efficient DOT&E approval of the overall OT&E effort in accordance with Title 10 U.S.C. § 139(d). Describe the scope of the operational test by identifying the test mission scenarios and the resources that will be used to conduct the test. Summarize the operational test events, key threat simulators and/or simulation(s) and targets to be employed, and the type of representative personnel who will operate and maintain the system. Identify planned sources of information (e.g., developmental testing, testing of related systems, modeling, simulation) that may be used to supplement operational test and evaluation. Quantify the testing sufficiently (e.g., number of test hours, test articles, test events, test firings) to fallow a valid cost estimate to be created.

End-to-End Testing

Guidance

Examples

Information Assurance

Guidance

Examples for ¶ 3.6.1

Realistic Operational Test Conditions

Guidance

Example

Production-Representative Test Articles

Guidance

Example for ¶ 3.6.1

Threat Representation

Guidance

Example for ¶ 3.6.1

3.6.2. **Modeling & Simulation** (M&S). Describe the key models and simulations and their intended use. Include the operational test objectives to be addressed using M&S. Identify data needed and the planned accreditation effort. Identify how the operational test scenarios will be supplemented with M&S. Identify who will perform the M&S verification, validation, and accreditation. Identify operational M&S resource requirements in Part IV.

Modeling and Simulation
Guidance
Examples

3.6.3. **Test Limitations*.*** Discuss test limitations including threat realism, resource availability, limited operational (military; climatic; Chemical, Biological, Nuclear, and Radiological (CBNR), etc.) environments, limited support environment, maturity of tested systems or subsystems, safety, that may impact the resolution of affected COIs. Describe measures taken to mitigate limitations. Indicate if any system contractor involvement or support is required, the nature of that support, and steps taken to ensure the impartiality of the contractor providing the support according to Title 10 U.S.C. §2399. Indicate the impact of test limitations on the ability to resolve COIs and the ability to formulate conclusions regarding operational effectiveness and operational suitability. Indicate the COIs affected in parenthesis after each limitation.

OT Test Limitations
Guidance
Example for ¶ 3.6.3

Information Assurance
Guidance
Examples for ¶ 3.7

3.7. Other Certifications. Identify key testing prerequisites and entrance criteria, such as required certifications (e.g. ***DoD Information Assurance Certification and Accreditation Process (DIACAP) Authorization to Operate***, Weapon Systems Explosive Safety Review Board (WSERB), flight certification, etc.)

*3.8***.** Reliability growth**.** Content moved to Paragraph 3.2.

3.9. Future Test and Evaluation - Summarize all remaining significant T&E that has not been discussed yet, extending through the system life cycle. Significant T&E is that T&E requiring procurement of test assets or other unique test resources that need to be captured in the Resource section. Significant T&E can also be any additional questions or issues that need to be resolved for future decisions. Do not include any T&E in this section that has been previously discussed in this part of the TEMP.

Adequate Test Resources

Guidance

Resource Example

Funding Example

4. PART IV-RESOURCE SUMMARY

4.1. Introduction. In this section, specify the resources necessary to accomplish the T&E program. Testing will be planned and conducted to take full advantage of existing DoD investment in ranges, facilities, and other resources wherever practical. Provide a list in a table format (see Table 4.1) including schedule (Note: ensure list is consistent with figure 2.1 schedule) of all key test and evaluation resources, both government and contractor, that will be used during the course of the current increment. Include long-lead items for the next increment if known. Specifically, identify the following test resources and identify any shortfalls, impact on planned testing, and plan to resolve shortfalls.

Production-Representative Test Articles

Guidance

Example for ¶ 4.1.1

4.1.1. **Test Articles**. Identify the actual number of and timing requirements for all test articles, including key support equipment and technical information required for testing in each phase of DT&E, LFT&E, and OT&E. If key subsystems (components, assemblies, subassemblies or software modules) are to be tested individually, before being tested in the final system configuration, identify each subsystem in the TEMP and the quantity required. Specifically identify when prototype, engineering development, or production models will be used.

4.1.2. Test Sites and **Instrumentation**. Identify the specific test ranges/facilities and schedule to be used for each type of testing. Compare the requirements for test ranges/facilities dictated by the scope and content of planned testing with existing and programmed test range/facility capability. Identify instrumentation that must be acquired specifically to conduct the planned test program.

Instrumentation

Guidance and Best Practices

4.1.3. Test Support Equipment. Identify test support equipment and schedule specifically required to conduct the test program. Anticipate all test locations that will require some form of test support equipment. This may include test measurement and diagnostic equipment, calibration equipment, frequency monitoring devices, software test drivers, emulators, or other test support devices that are not included under the instrumentation requirements.

Threat Representation

Guidance

Example for ¶ 4.1.4

4.1.4. **Threat Representation**. Identify the type, number, availability, fidelity requirements, and schedule for all representations of the threat (to include threat targets) to be used in testing. Include the quantities and types of units and systems required for each of the test phases. Appropriate threat command and control elements may be required and utilized in both live and virtual environments. The scope of the T&E event will determine final threat inventory.

4.1.5. Test Targets and Expendables. Specify the type, number, availability, and schedule for all test targets and expendables, (e.g. targets, weapons, flares, chaff, sonobuoys, smoke generators, countermeasures) required for each phase of testing. Identify known shortfalls and associated evaluation risks. Include threat targets for LFT&E lethality testing and threat munitions for vulnerability testing.

4.1.6. Operational Force Test Support. For each test and evaluation phase, specify the type and timing of aircraft flying hours, ship steaming days, and on-orbit satellite contacts/coverage, and other operational force support required. Include supported/supporting systems that the system under test must interoperate with if testing a system-of-systems or family-of-systems. Include size, location, and type unit required.

4.1.7. Models, Simulations, and Testbeds. For each test and evaluation phase, specify the models and simulations to be used, including computer-driven simulation models and hardware/software-in-the-loop test beds. Identify opportunities to simulate any of the required support. Identify the resources required to validate and accredit their usage, responsible agency and timeframe.

4.1.8. Joint Mission Environment. Describe the live, virtual, or constructive components or assets necessary to create an acceptable environment to evaluate system performance against stated joint requirements. Describe how both DT and OT testing will utilize these assets and components.

4.1.9. Special Requirements. Identify requirements and schedule for any necessary non-instrumentation capabilities and resources such as: special data processing/data bases, unique mapping/charting/geodesy products, extreme physical environmental conditions or restricted/special use air/sea/landscapes. Briefly list any items impacting the T&E strategy or government test plans that must be put on contract or which are required by statute or regulation. These are typically derived from the JCIDS requirement (i.e., Programmatic Environment, Safety and Occupational Health Evaluation (PESHE) or Environment, Safety and Occupational Health (ESOH)). Include key statements describing the top-level T&E activities the contractor is responsible for and the kinds of support that must be provided to government testers.

4.2. Federal, State, and Local Requirements. All T&E efforts must comply with federal, state, and local environmental regulations. Current permits and appropriate agency notifications will be maintained regarding all test efforts. Specify any National Environmental Policy Act documentation needed to address specific test activities that must be completed prior to testing and include any known issues that require mitigations to address significant environmental impacts. Describe how environmental compliance requirements will be met.

4.3. Manpower/Personnel and Training. Specify manpower/personnel and training requirements and ***limitations*** that affect test and evaluation execution. Identify how much training will be conducted with M&S.

4.4. Test Funding Summary. Summarize cost of testing by FY separated by major events or phases and within each Fiscal Year (FY) DT and OT dollars. When costs cannot be estimated, identify the date when the estimates will be derived.

Table 4.1 Test Sites and ***Instrumentation*** Example

Fiscal Year	06	07	08	09	10	11	12	TBD
TEST EVENT / TEST RESOURCE	IT-B1	IT-B2	IT-B2 / IT-C1	IT-C1	IT-C1	IT-C2	OT-C1	OT-D1
Integration Lab	X	X	X	X	X	X		
Radar Integration Lab	X	X	X	X	X	X		
Loads (flights)								
Operating Area #1 (flights)		X(1)	X(1)				X (1)	X (2)
Operating Area #2 (flights)		50(1)	132(1)	60	100	140	X (1)	X (2)
Northeast CONUS Overland (flights)		10					X (1)	X (2)
SOCAL Operating Areas (flights)				X		X		
Shielded Hangar (hours)			160			160		
Electromagnetic Radiation Facility (hours)			40			40		
Arresting Gear (Mk 7 Mod 3)(events)				10		10		
NAS Fallon				5	5	A/R	X (1)	X (2)
Link-16 Lab, Eglin AFB							X	
NAWCAD WD, China Lake Range							X	
Eglin AFB ESM Range							X	

1. Explanations as required.
2. Enter the date the funding will be available.

Threat Representation – Guidance

Guidance

In operational testing, threats should be adequately represented to assist in evaluation of the system under test in a realistic operational environment. The goal for threat presentation is to match the envisioned threat to the system under test (SUT), based on Defense Intelligence Agency (DIA) or Service intelligence threat assessments. Particular emphasis should be placed on adequate representation of threats that are most relevant to the evaluation of the system under test. Threat systems serve as targets for demonstration of SUT performance and as threats to SUT survivability.

The TEMP should illustrate that threats will be adequately represented in testing by including plans to:

- Section 1.3.1: Identify the threats of most interest to evaluation of the system under test (Example)
- Section 1.3.3.2: If necessary, describe the development of special threat or target systems (Example)
- Section 3.6.1: Describe the necessary capabilities (weapons, tactics, command and control, etc.), physical and kinematic attributes (signatures, speed, attack profile, maneuverability, size and shape, etc.), or the necessary fidelity of the proposed threats for IOT&E (Example)
- Section 3.6.3: Identify projected critical/severe or major test limitations stemming from inadequate threat representation, and plans to mitigate those limitations (Example)
- Section 4.1.4: Identify the necessary quantity (numbers of troops, attack aircraft, surface-to-air missiles, torpedoes, tanks, etc.) of threat systems necessary for all test events (Example)

Identification and description of certain threats in the Service or DIA threat assessments may lead to an early conclusion (that should be flagged as early as Milestone A TEMPs) that a credible, threat-representative surrogate does not exist and may require development to achieve an adequate IOT&E.

Thorough Service-sponsored technical and operational comparisons (validation) must be made between the threat and candidate surrogates. Validation culminates in a report documenting validation results. DOT&E monitors the validation and approves the Service-validated reports. Furthermore, careful consideration and documented explanation (OTA accreditation) of threat-to-surrogate differences and their projected impacts on the system under test must be made to determine surrogate adequacy.

References

Defense Acquisition Guidebook, Chapter 9

Threat Representation – Examples

1.3.1 System Threat Assessment

The Dakota Threat Assessment Report (STAR) prepared by the Intelligence Division, U.S. Army Aviation and Missile Command, contains the Defense Intelligence Agency-validated threat to Dakota. The Dakota STAR was validated in April of 2010. The following is an unclassified summary of the STAR"s key points.

Most of the regional powers will field large armored forces supported by fixed- and rotary-wing aircraft, mobile artillery, longer-range antiaircraft artillery, surface-to-air missiles, antitank guided missiles, communications and non-communications electronic warfare systems, ground-based and airborne reconnaissance, surveillance, target acquisition systems operating in various regions of the electromagnetic spectrum, and a sophisticated command, control, and communications (C3) system. Modern major weapon systems will be acquired mainly from Russia, China, or the West. Most of the regional powers will field camouflage, concealment, and deception and various countermeasures equipment designed to degrade or negate the effectiveness of enemy sensors and precision-guided munitions. A few technologically advanced countries are exploring the feasibility of high-energy laser or high-powered microwave devices that could evolve into weapons development programs and eventually proliferate. Most of the regional powers will be capable of offensive chemical and biological warfare and some will acquire or improve the capability to conduct tactical nuclear warfare. Some of the more technologically advanced countries will develop a limited capability to conduct information operations. As in the past, the ability to effectively employ modern warfighting concepts and deploy and maintain sophisticated equipment will vary from country to country.

In the wake of a major regional conflict, or at the outset of a low-intensity conflict, an asymmetric threat will exist from dispersed light forces that will employ tactics and techniques that will be difficult for U.S. forces to counter. Generally, asymmetric combatants will exploit complex terrain, particularly highly populated urban terrain, for concealment as well as political advantage, exploiting the indigenous environment and its inhabitants for surprise, escape routes, and shielding while negating a conventionally armed adversary's strength in numbers, equipment, and firepower. Asymmetric combatants will be armed with infantry small arms, rocket propelled grenades, light artillery and antiaircraft machineguns, man-portable antitank and surface-to-air missiles, and night vision devices, either inherited from the old regime or acquired from outside suppliers, as well as various improvised weapons produced locally. Some adversaries could acquire weapons and equipment incorporating relatively sophisticated technology that nonetheless is suitable for small unit operations, such as man-portable ground surveillance radar, unmanned aircraft systems, low-energy laser blinders, anti-helicopter mines, Global Positioning System (GPS) jammers, or expendable radio frequency weapons. Asymmetric adversaries will employ commercial communications equipment such as cell phones, as well as portable military radios, for C2.

Information Assurance - Guidance

General Guidance

The TEMP should describe the operational test strategy for evaluation of information assurance for all oversight programs.

As weapons systems become increasingly reliant on information technologies, it has become necessary to consider information assurance for all programs. Acquisition programs are required to protect information systems during all phases of the acquisition, including initial design, development, testing, fielding, and operation. Operational test and evaluation should seek to evaluate, in realistic operational environments, an acquisition system's (or a system-equipped unit's) ability to defend against, detect, and react to penetrations and exploitations of information systems, and to restore data and information if necessary. DOT&E procedures provide a framework for information assurance evaluations that defines issues and measures, encourages leveraging of accreditation and developmental test data by the Operational Test Agency, and suggests a six-step process for determining an operational test strategy (Procedures for Operational Test and Evaluation of Information Assurance in Acquisition Programs, January 21, 2009; Clarification memo, November 4, 2010).

Information assurance information for inclusion in the TEMP

Portions of the information assurance strategy should appear throughout the TEMP in the following paragraphs:

- Paragraph 1.3. System Description. Provide the Mission Assurance Category (MAC) and Confidentiality Level for the system (DoD Instruction 8500.2).
- Paragraph 1.3. System Description. Describe any previous information assurance certifications or accreditations.
- Paragraph 1.3.1. System Threat Assessment. Identify and cite an appropriate threat assessment, such as the Defense Intelligence Agency Capstone Information Operations Threat Document.
- Paragraph 1.3.3.2. Special Test or Certification Requirements. State that the information assurance testing will be performed in accordance with policies and requirements established by Department of Defense Regulation 5000.02-R, DoD Directive 8500.1, DoD Instruction 8500.2, and DOT&E Procedures for the Operational Test and Evaluation of Information Assurance.

- Paragraph 3.1. T&E Strategy. Integrate testing and assessment of system information assurance into appropriate integrated tests to identify risk and potential vulnerabilities. Complete an initial system-level information assurance assessment in conjunction with the DoD Information Assurance Certification and Accreditation Process (DIACAP).
- Paragraph 3.2. Evaluation Framework. Integrate information assurance into the overarching system evaluation. Identify issues and measures for the information assurance assessment. Identify the key interfaces required for end-to-end information assurance testing.
- Paragraph 3.3. Developmental Evaluation Approach. Identify test events (during DT or DT/OT) that will assess information assurance Measures of Performance appropriate for the MAC and CL assigned to the system.
- Paragraph 3.5. IOT&E Entrance Criteria. Consider information assurance when defining test entrance criteria, such as IA certifications, authorities to operate, and completion of vulnerability assessments.
- Paragraph 3.6.1 Operational Test Objectives. Identify events and organizations for completing the Step 4 (operational vulnerability evaluation) and Step 5 (Red Team penetration testing) of the DOT&E Procedures. Specify an appropriate threat level to be portrayed, and ensure that the test duration is adequate for that level. For force-on-force testing, consider appropriate integration of information assurance activities by the Red Team into the opposing forces to make information assurance testing more representative and mission-focused. Describe the plan for assessing continuity of operations.
- Paragraph 4.4 Test Funding for information assurance testing and analyses. Identify information assurance test resources and funding, including identification of organizations supporting the planned certification, testing, and evaluation activities.

Definitions

Information Assurance: Measures that protect and defend information and information systems by ensuring their availability, integrity, authentication, confidentiality, and nonrepudiation. This includes providing for restoration of information systems by incorporating protection, detection, and reaction capabilities (Reference: Joint Publication 1-02).

Information Systems: Any equipment or interconnected system or subsystem of equipment that is used in the automatic acquisition, storage, manipulation, management, movement, control, display, switching, interchange, transmission or reception of data or information by the DoD Component (Reference: Joint Publication 1-02).

Information Assurance – Examples

1.3.1 SYSTEM THREAT ASSESSMENT

Example

The system will operate in the full spectrum of threat environments that DoD is expected to face. These environments may range from peacetime or peacekeeping environments to Major Theater War environments. The primary threats may come from electronic warfare measures such as electronic support (intercept or direction finding) or electronic attack (jamming), or as offensive information warfare. Computer Network Operations threats to the information systems (including insiders, distributed denial of service, malicious code, and unauthorized users) exist throughout the entire system lifecycle. Potential threats also include collateral blast and fragmentation from small arms; direct fire weapons; indirect fire weapons including mortars, artillery, rockets, and guided and unguided missiles; conventional and guided bombs; rocket propelled grenades and guided missiles; upset or damage from radio frequency directed energy weapons; and effects of chemical, biological, radiological, and nuclear weapons. These weapons may be employed by any combination of irregular forces, infantry, field artillery, mechanized or armored forces, unmanned vehicles (ground, airborne, or waterborne), ships, and aircraft.

The system specific threats are addressed in the Defense Intelligence Agency (DIA) validated System Threat Assessment Report; Enterprise Threat Assessment Report; DIA Information Operations Capstone Threat Assessment Volume 1-8 and 10-14, 5th Edition, DI-1577-33-06 January 2006, (SECRET//NOFORN//20300804); DIA Information Operations Capstone Threat Assessment Volume 9, 6th Edition, DI-1577-37-07 April 2007 (SECRET//FGI//NOFORN//20311018).

Information Assurance – Examples

1.3.3.2 SPECIAL TEST OR CERTIFICATION REQUIREMENTS

(See examples below)

3.7 OTHER CERTIFICATIONS

(If not already covered in Paragraph 1.3.3.2)

Example 1

Weapon System Explosive Safety Review Board approval is obtained through the process laid out in reference (h). Information Assurance Interim Authority to Operate and Authority to Operate are obtained in accordance with the DIACAP process laid out in reference (i) and the processes put forth in reference (j).

Example 2

The radio sets will comply with DIACAP. A Platform Information Technology determination request has been submitted for the small form fit sets and DIACAP will be conducted at the host platform level for the these sets. DIACAP is a dynamic certification and accreditation process that supports the net-centric, Global Information Grid-based environment. It is the DoD process for identifying, implementing, validating, certifying, and managing information assurance capabilities and services, expressed as information assurance controls, and for authorizing the operation of DoD Information Systems, including testing in a live environment, in accordance with statutory, Federal, and DoD requirements. The Joint Staff (J6), as the Principal Accrediting Authority (PAA) for the Warfighting Mission Area, appointed the Naval Chief Information Officer (CIO) as the Designated Approving Authority (DAA). Navy CIO has delegated DAA responsibilities to the Navy Network Warfare Command (NNWC).

Example 3

Per the DoDI 5000.2, this system is designated as Mission Critical. Development of the system information assurance requirements throughout the system life cycle is in accordance with the Department of Defense Instruction (DoDI) 8510.01, DoD Information Assurance Certification and Accreditation Process (DIACAP). The Security Test and Evaluation (ST&E) will explicitly address testing each of the required information assurance control measures and report each information assurance control to the Designated Approval Authority (DAA). DAA is the Chief Information Officer for DISA; the Certification Authority is the DISA Chief Field Security Operation Division; and the information assurance manager is the system program information assurance manager. The Operational Test Agencies will verify the operational aspects of the information assurance control measures as defined in Director, Operational Test and Evaluation (DOT&E) guidance. The program management office will use the DIACAP Knowledge Service and Enterprise Mission Assurance Support Service. The Acquisition Information Assurance Strategy details the implementation of information assurance across the program lifecycle.

Integrated Testing – Guidance

Guidance

DOT&E and AT&L directives require the seamless integration of developmental and operational testing throughout the life cycle of a system under test. In their joint memo of 25 April 2008 DOT&E and AT&L defined integrated testing as follows:

> Integrated testing is the collaborative planning and collaborative execution of test phases and events to provide shared data in support of independent analysis, evaluation and reporting by all stakeholders particularly the developmental (both contractor and government) and operational test and evaluation communities.

Background

If planned and executed appropriately, integrated testing allows for a faster and more cost effective T&E process that ultimately provides the Services with more capable systems sooner and at a reduced cost as compared to sequential testing. As noted by DOT&E on 24 November 2009, integrated testing will never do away with the need for a dedicated operational test to confirm that systems will work in combat. The legal requirement (USC 139, USC 2399) for a dedicated operational test is also clear. Nonetheless, separation of developmental and operational testing has caused difficulties in the development process that have been documented by the Defense Science Board and the National Academies.

Integrated testing may come about in two ways: a developmental test is made into an integrated test by changing the manner in which it is executed, or a developmental test provides adequate data for an operational evaluation regardless of how it is executed. The latter type of integrated test is relatively easy to plan and execute because it only requires that the metrics being measured be invariant under the developmental and operational test conditions. Which is to say, the measured value of the metric being tested is the same under the conditions of a development test and an operational test.

Developmental Evaluation Strategy Section of the TEMP

The Developmental Evaluation section of the TEMP (paragraph 3.3) should list each developmental test event that will be used as an integrated test event. In addition to other relevant details of the test (i.e., when will the test be conducted, where will the test be conducted, and who will conduct the test), it should include details about each developmental test's objectives and the corresponding operational test's objectives along with some justification why the two sets of objectives may be satisfied by a single test event. Additionally, the text should describe the operational conditions necessary for the integrated test and should explain why any deviation from operationally realistic conditions, if any, is acceptable.

Best Practices

Good examples of metrics that do not usually depend on the conditions of the test are cargo and storage capacity requirements common to amphibious ship programs. These requirements require the ship to provide a specified, cubic foot amount of cargo storage or square feet of vehicle storage. The amount of space available does not depend on the conditions of the test, so a developmental test that measures the space should provide adequate data for an operational test. Developmental tests that are often used to provide this data are Marine Corps Certification Exercises or Navy In-service Inspections.

The former type of integrated test – a test where a developmental test is conducted under operationally realistic conditions – requires great care to ensure that the developmental test goals do not interfere with the operational test goals, and to ensure that the test is executed under operationally realistic conditions.

Air Warfare Ship Self-Defense test events, particularly those conducted on the remote control Self-Defense Test Ship,[1] are good examples of integrated tests where a developmental test is executed under conditions that are sufficiently operationally realistic. During Self-Defense Test Ship events, aerial targets are flown directly at the test ship. The combat system elements of the ship are operated by civilian experts via remote control. As a developmental test platform, the test ship provides a highly controlled environment for testing specific system metrics. By ensuring the aerial targets are representative of actual anti-ship cruise missile threats, and by ensuring the flight profile of the target is the same as the threat, the developmental test can be used as an integrated test.

---

[1] The Self-Defense Test Ship is a former Spruance Class Destroyer that has been equipped with multiple modern-day anti-air warfare combat systems. The ship and its combat systems are both capable of being operated by remote control, thereby reducing the risk of mishap when engaging anti-ship cruise missiles and aerial targets.

Identifying and planning integrated tests is usually the responsibility of the Integrated Test Team. The Integrated Test Team is responsible for identifying potential integrated tests, ensuring that the test objectives for the developmental and operational tests are sufficiently compatible, and ensuring that the developmental test is executed under operationally realistic conditions. The Integrated Test Team should include representatives from the Operational Test Agency, DOT&E, AT&L DT&E, and the program office.

Baseline Evaluation – Guidance

Summary

The primary objective of Defense acquisition is to acquire quality products that satisfy user needs with measurable improvements to mission capability and operational support, in a timely manner, and at a fair and reasonable price.

One way to determine "measurable improvements" is through comparative or baseline evaluation, which compares unit mission accomplishment when equipped with the new system to unit mission accomplishment when equipped with current force capabilities. This comparison is in addition to assessing a new system's achievement of its required performance characteristics.

Typically, many uncontrollable variables are present during operational testing, especially in force-on-force exercises. Areas where commonality should be sought between trials in order to enable valid comparisons include: the mission to be accomplished; the size, organization, and capability of the enemy force; the terrain (or environment) where the test is conducted; the size, organization, and capability of the Blue forces; and time available to accomplish the mission (referred to as Mission, Enemy, Terrain, Troops available and Time, or METT-T in Army parlance).

Best Practices

There are several ways to gather data on a current force unit's mission accomplishment for baseline purposes. One way is to conduct a side-by-side operational test, as during the Stryker IOT&E, with a current force unit and a unit equipped with the new system. In the M2A3 Bradley IOT&E, the M2A3 Bradley unit conducted operations against a normal Bradley unit for a head-to-head comparison. Current force field training exercises can also be used as a source of baseline data. The Task Force XXI Advanced Warfighting Experiment at the National Training Center used three NTC rotations to establish a baseline for normal unit performance. The use of the Analysis of Alternatives can be helpful in determining the factors and levels described above for cases to be examined, and also for predicting what the baseline force performance will be in actual field trials.

The Navy made effective use of hardware-in-the-loop (HWIL) M&S to support the evaluation of heavyweight torpedoes. The OT objective was to assess a form-fit-functional replacement of the weapon's Guidance and Control section running a rehosted version of the tactical software. The HWIL simulation allowed testers to run both the legacy and upgraded systems through a series of identical scenarios and then compare the results. While a limited number of in-water trials were conducted to validate the model and verify system suitability, this M&S approach was able to provide a large, well-controlled data sample to compare the performance of the two variants.

References

Test and Evaluation Policy Revisions, DOT&E, December 22, 2007

DoDD 5000.01

Mission Focused Metrics – Guidance

General Guidance

TEMPs should include quantitative mission-oriented metrics (also referred to as response variables) for effectiveness and suitability. Evaluation metrics are key to good test designs; poorly-chosen or poorly-defined measures will result in a poorly designed test, and can lead to unnecessary costs or ambiguous test results that are not relevant to the operational needs of the user.

Choosing Metrics

The selection of evaluation metrics is a critical part of test design effort, and should occur as test planning begins. Step 1 is to identify the critical operational issues (COIs): what capability is this system intended to provide? Once this is known, testers should select appropriate metrics that provide a means to measure performance and provide data for answering the COIs. Ideally, the metrics will provide a determination of mission capability, lend well to good experimental design (DOE), and encapsulate the reasons for procuring the system.

Evaluation metrics are typically selected from key performance parameters, measures of effectiveness, measures of suitability, critical technical parameters, key system attributes, and/or measures of performance already documented in requirements documents. Although many metrics can be used to characterize system performance in a given mission, it is desirable that one or two primary metrics be identified to be the focus of test design and used in concert with design of experiments methodologies. Additional secondary metrics are encouraged, and are necessary to characterize other aspects of system performance. For example, for test design, the hit success rate may be identified as the primary variable, even though other metrics to characterize success in the dependent portions of the kill chain are valuable (e.g., detection, localization).

Exceptions to using CDD/CPD-defined Metrics

The primary metric identified for test design need not be the KPP(s). Often KPPs, while important, are insufficient for measuring the operational performance of the system; this is especially true when KPPs detail gross static requirements for a system such as maximum size/weight, total number of weapons loads, or frequency coverage of a sensor. While important attributes of the system, such metrics do not characterize the intended system performance in an operational environment; a more operationally relevant and mission-related metric should be selected in order to plan the test program. Examples of mission-focused metrics that enable mission-focused test design include detection/classification range, miss distance, probability of hit, search rate, time to accomplish a successful mission, and probability of successful intercept.

Many CDDs define KPPs/MOEs for the technical characteristics of a system: e.g., signature requirements such as radar cross section or radiated noise. These requirements, if selected as the primary metric for test planning, would lead to a structured test program to precisely measure these quantities under controlled conditions (necessary for developmental testing but usually inappropriate for operational testing). The selection of a more operational metric in lieu of the KPP (e.g., counterdetection range) enables testers to design a test that examines an operationally meaningful question under a variety of realistic conditions and scenarios.

When testers select these primary metrics, the resultant test design should ensure that adequate data will be collected to accomplish several goals:

- Provide adequate data to evaluate CDD requirements (even if the response variable selected is not explicitly defined in the CDD)
- Provide a meaningful measure of system performance across the operational envelope
- Provide sufficient data for the secondary metrics needed to characterize system performance.

Types of Metrics

Response variables can be continuous or discrete. Examples of continuous responses include time to detect, miss distance, and range of engagement. Examples of discrete responses include hit/miss, message complete/not complete, and detect/not detect. A continuous response variable is preferred to a discrete one, since it will almost always require a smaller sample size and fewer test resources for the risk levels chosen (confidence and power).

Continuous variables also often contain more information regarding the performance of the system, whereas a corresponding discrete variable will throw away information. For example, measuring detect/not detect provides no information about how close the sensor approached. Using the range at which detection occurred in concert with the closest point of approach in cases where no detection occurred provides a better characterization of sensor performance. The probability of detection over all ranges is the only quantity that can be calculated with the discrete data, but if the continuous variable (range) is measured, one can determine both the mean range of detection as well as the probability of detection as a function of range.

Definitions of Metrics

The metric chosen must also be well-defined and meaningful. Evaluators should be encouraged to consider example operational scenarios to ensure that the metric can be unambiguously measured (scored) and calculated in all cases. The following principles are critical:

- Formulas for the metric should not be ambiguous – TEMPs should provide amplifying information (explicit formulas and/or scoring criteria) if the CDD requirement is unclear
- Metrics should be testable and not require unsafe or unexecutable test constructs or cost-prohibitive instrumentation
- Metrics should accurately represent the desired performance of the system – Good scores should correspond to desired operational performance
- Metrics should not lead to non-production representative modifications to the system or unrealistic tactics.

Metric Selection for Survey Data and Expert Panels

In operationally focused testing, the use of operator surveys and subject matter expert panels are needed and useful to aid in the characterization of system performance. This is particularly true when quantitative data is scarce due to expensive field testing or low sample sizes. Additionally, many important aspects of operational suitability are best addressed by survey data (e.g., human machine interface, operator workload). Ideally, survey data and subject matter expert panels should be used in concert with objective quantitative data.

Survey use should follow best practices, such as:

- Clearly identify survey objectives: TEMP should indicate which COIs will be addressed by survey data
- Surveys should be tested on an appropriate group to reveal if questions are confusing or if information is missing
- Survey questions should be clear and unbiased (e.g., no leading questions)
- Surveys should use quantitative (e.g., Likert-scale) and qualitative responses (open ended questions); quantitative data should be coded, compiled and summarized using statistical methods to aid in system characterization in concert with the metrics employed in field testing.

References

Reporting of Operational Test and Evaluation (OT&E) Results, DOT&E, January 6, 2010

Test and Evaluation Policy Revisions, DOT&E, December 22, 2007

Guidance on the Use of Design of Experiments (DOE), DOT&E, October 19, 2010

Reliability Growth – Guidance

Summary

The majority of life cycle costs for DoD systems reside in the Operations and Sustainment (O&S) phase, while the single greatest driver of O&S costs is reliability. The more reliable the system, the less it costs to operate and sustain in the field. With today's highly complex systems, a small decrease in reliability can mean additional, substantial cost, but a small investment in reliability growth can significantly decrease O&S costs.

From the start, a program should formulate and document a comprehensive reliability, availability, and maintainability (RAM) program using an appropriate reliability growth strategy to improve RAM performance until RAM requirements are satisfied. Systems engineering and design activities should be included in the Test and Evaluation Master Plan (TEMP). The TEMP must also describe the reliability growth strategy and specify how reliability will be tested and evaluated during each acquisition phase. To be effective, the reliability growth strategy must include opportunities to implement corrective actions and test events to confirm effectiveness of those actions. Reliability growth curves (RGC) included in the TEMP must reflect the reliability growth strategy. The TEMP resources section must identify resources necessary to support the strategy.

A program should monitor reliability growth throughout the acquisition process and report the status of reliability objectives at all formal reviews. When necessary, reliability growth should continue after the full-rate production decision (FRP) and fielding until RAM requirements are met. Provisions should be made to monitor reliability even after requirements are met.

Reliability information for inclusion in the TEMP

At all milestones, the TEMP should include an overview of the RAM program and testing needed to assess and monitor reliability growth, including design for reliability test and evaluation (T&E) activities. The test schedule should be event-driven and allow adequate time to support pre-test predictions; testing; post-test analysis, evaluation, and reporting; and adequate time to support execution of corrective actions in response to discovered deficiencies. The TEMP T&E resources section must identify resource requirements (including test articles and expendables) that reflect the best estimate for conducting all reliability T&E activities.

Reliability growth should be measured, monitored, and reported throughout the acquisition process. Programs using mean time between failure (MTBF) metrics should calculate the reliability growth potential (the maximum MTBF that can be attained with the current management strategy) to ensure that reliability thresholds are achievable. RGCs are employed to report reliability growth status at Defense Acquisition Executive System reviews. PMs must continue to track reliability after FRP, regardless of whether reliability requirements have been met.

A comprehensive RAM program consists of engineering activities, including: RAM allocations to components and subsystems, block diagrams and predictions; failure definitions and scoring criteria (FDSC); failure mode, effects and criticality analysis (FMECA); maintainability and built-in test demonstrations; reliability growth testing at the system and subsystem level; and a failure reporting and corrective action system (FRACAS) maintained through design, development, production, and sustainment.

Reliability Growth Curves reflect the reliability growth strategy and are employed to plan, predict, illustrate, and report reliability growth. A RGC must be included in the SEP at Milestone A, and updated in the TEMP beginning at Milestone B. RGC should be stated in a series of intermediate goals and tracked through fully integrated, system-level test and evaluation events until the reliability threshold is achieved. For complex systems and systems-of-systems, a single curve will likely not be adequate to describe overall system reliability. A separate growth curve should be provided for each critical subsystem with rationale for its selection.

At Milestone C, RGCs are updated with current information (including the current reliability estimate) and an updated reliability growth plan. The reliability growth plan is updated to form a comprehensive RAM program to improve RAM performance throughout the life of the program. The TEMP should characterize key failure modes and their disposition. Post-Milestone C TEMPs must be updated as needed to continue reliability monitoring and reliability growth after fielding until terminated by the receiving Service.

Software-intensive Programs

Programs consisting entirely of software and commercial off the shelf hardware do not have to include a reliability growth curve in the TEMP. However, they must still address how they plan to track software reliability across contractor, developmental, and operational testing; categorize problems; conduct regression tests; provide forecasts for system reliability at critical decision points; and close out problem reports as usually done with other programs.

Additional software reliability and maturity metrics, along with best practices in software coding, quality control, peer review, and other aspects of the Capability Maturity Model Integration (CMMI), can be found at the Software Engineering Institute website (www.sei.cmu.edu). The National Institute of Standards and Technology (NIST) Special Publication 500-209, Software Error Analysis, dated March 1993 is another useful reference for software defect tracking and reporting.

References

Directive-Type Memorandum (DTM) 11-003 – Reliability Analysis, Planning, Tracking, and Reporting, March 21, 2011.

DoD Instruction 5000.02

Examples

Reliability Metrics for Table 3.1 Example

Reliability Growth Example

Software Tracking Example

Information Assurance – Examples

3.1 TEST AND EVALUATION STRATEGY

Example 1

1. Information Assurance (IA). IA testing will be integrated throughout the test process. The IA effort is closely tied to and coordinated with the OT effort, but takes advantage of other test activities, including PVT and program sponsored security test and evaluation events. The effort will utilize document reviews, participation in and/or observation of events, and other relevant events to collect data in support of the IA effort.

2. Information Assurance BLUE and RED Teams. The Director, Operational Test and Evaluation (DOT&E) Policy mandates that a Red Team Assessment be conducted for all MAC I and MAC II systems assigned a CL of Classified or Sensitive. The Red Team capabilities must be commensurate with the threat and expected risks for the program. One of the best ways to prepare for a cyber threat is through the use of an information assurance Red Team, which is an independent, interdisciplinary, simulated enemy force. After proper safeguards are established, the Red Team uses active and passive techniques to expose and exploit information assurance vulnerabilities of friendly forces. The results are used as a means to improve those forces' readiness. The threat capabilities should be based upon an IO CTA, CPD, or equivalent. The Blue Team incorporates both technical and non-technical assessments to identify system vulnerabilities. The Blue Team data must be completed at least a month prior to the Red Team event. The Red Team (vulnerability assessments and penetration testing) will be based upon threat capabilities validated by the Defense Intelligence Agency (DIA).

Reliability Growth – Example

3.2 Reliability Growth

Dakota reliability growth will consist of positive improvement through systematic removal of failure mechanisms by way of positive changes in design, material, or manufacturing. Dakota reliability growth will begin at program initiation and continue through production. Reliability growth will be achieved not only through lab and flight testing, but also by way of design analysis, production experience, and operational experience.

The reliability growth test program will accomplish its goals by: (1) finding reliability problems through testing, (2) establishing a Failure Reporting, Analysis, and Corrective Action System (FRACAS) to identify root causes of failure and corrective actions, (3) incorporating corrective actions when timely or appropriate, and (4) continual monitoring of corrective actions and the system's reliability throughout all test phases.

Dakota Reliability, Availability, and Maintainability (RAM) performance will be continuously assessed using data from development flight testing, logistics demonstration, and operational testing. Dakota reliability growth will be tracked against a reliability growth curve that estimates reliability thresholds associated with program decision points. The focus of the Dakota reliability growth program will be on identification of new and existing failure modes and correction of hardware and software failures. A failure review board consisting of Government and contractor elements will convene monthly to discuss the FRACAS data and evaluate the root cause determination, proposed corrective actions, and the verification methodology. Once corrective actions are verified and incorporated, the corrective action will continue to be monitored for fix effectiveness to assess its impact on reliability growth.

RAM Scoring Conferences will be held quarterly. All RAM data will be scored using the approved Dakota Failure Definition/Scoring Criteria. The RAM Scoring Conference voting members are the materiel developer, the combat developer, and the evaluator. Testers and technical support personnel may support the Scoring Conferences in an advisory capacity.

The goal for the reliability growth program is to demonstrate the 17-hour MTBF Full Rate Production requirement with 80 percent confidence using data from IOT&E. To provide evidence at Milestone C that reliability the reliability growth goal is achievable,

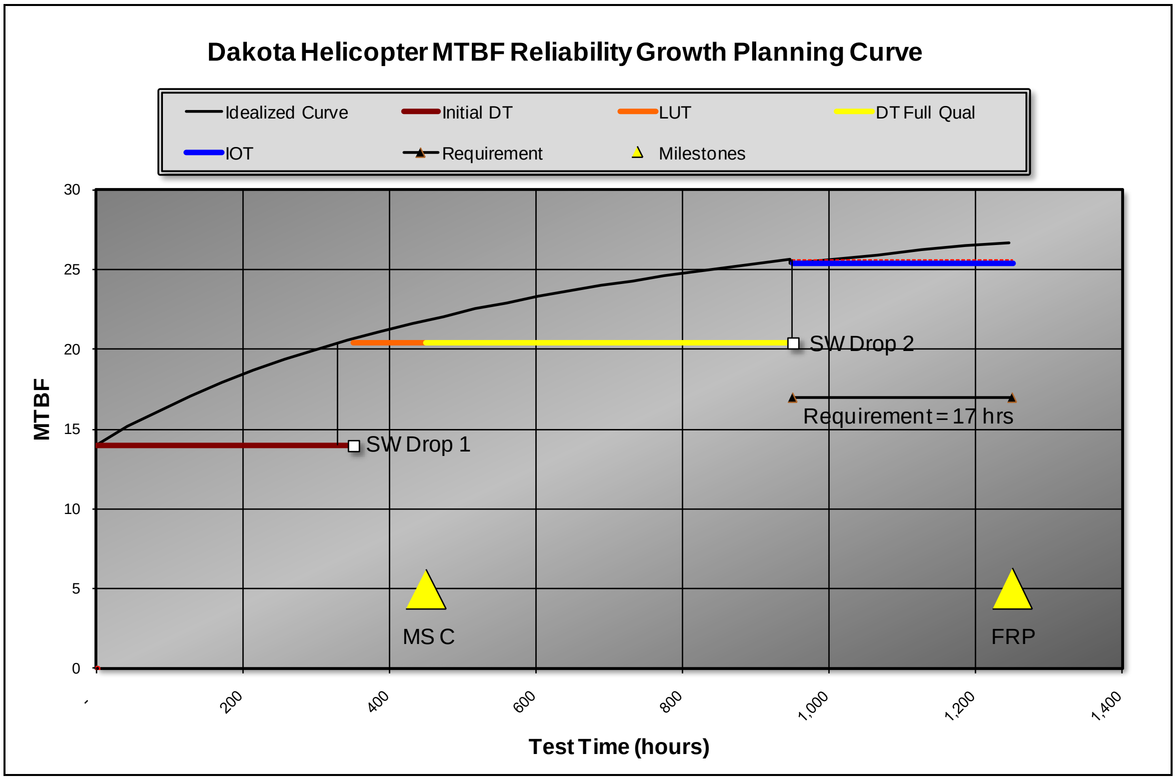


Figure 1. Reliability Growth Curve

the program will seek to demonstrate a MTBF of 20 hours during the Limited User Test (LUT). The development goals associated with this reliability growth program include corrective action or mitigation for at least 80 percent of the initial failure intensity with an average fix effectiveness factor of 70 percent.

The reliability growth plan consists of two corrective action periods for implementing corrective actions to reliability deficiencies observed during developmental test flights. There will be a major software release just prior to the LUT and another just prior to IOT&E. The majority of corrective actions discovered in developmental testing will be implemented in these software releases.

Table 1. Projected Flight Hours Supporting Reliability Growth

Test	Test Flight Hours	Cumulative Flight Hours
Initial DT	350	350
LUT	100	450
DT Full Qualification	500	950
IOT	300	1250

Design of Experiments – Guidance

General

Design of Experiments (DOE) is a statistical methodology for planning, conducting, and analyzing a test. Any program that applies DOE principles should commence doing so early in the test planning process. The program should assemble a test and evaluation working integrated product team (T&E WIPT) of subject matter experts who can identify the primary evaluation metrics (in DOE parlance: response variables) of interest that will characterize the performance of the system in the context of a mission-oriented evaluation operational. The T&E WIPT should identify environmental and operational factors that are expected to drive the performance of the system, as well as the levels of these factors (i.e., the various conditions or settings that the factors can take). A master test strategy should include the resources needed, the concept for early tests (including component tests), and the use of the results of early tests to plan further testing. One goal of the test strategy should be to ensure adequate coverage of all important factors while demonstrating the evaluation metrics (response variables) through planned testing. The testing strategy should be iterative in nature to ensure an adequate Initial Operational Test and Evaluation (IOT&E). The testing strategy should accumulate evidence that the system performs across its operational envelope before and during IOT&E. The T&E WIPT should apply DOE at each test iteration.

Elements of DOE for the TEMP

A brief overview of the design philosophy should be outlined in Section 3.2 of the TEMP. The information content may vary depending on the Milestone that the TEMP is supporting. Table 1 outlines information content that is appropriate for each milestone. Systems with legacy data will be expected to include more detail and have more robust test designs. The details of each of the test designs should be provided in a supporting appendix to the TEMP. Elements of experimental design should include the following:

- The goal of the test (experiment). See Mission-Oriented Testing Guidance.
- Quantitative mission-oriented response variables (evaluation metrics) for effectiveness, suitability, and survivability. See Mission Focused Metrics Guidance.
- Factors that affect those measures of effectiveness, suitability, and survivability. See Integrated Survivability Evaluation Guidance.

- A method for strategically varying factors across developmental, operational, and live fire testing with respect to responses of interest See Integrated Testing Guidance.
- Statistical measures of merit (power and confidence) on the relevant response variables (evaluation metrics) (i.e., those for which doing so makes sense). These statistical measures are important to understand "how much testing is enough," and can be evaluated by decision makers on a quantitative basis so they can trade off test resources for desired confidence in results.

These elements include all of the planning steps for designing an experiment, with the exception of execution order. Standard statistical designs assume the test point execution order can be randomized. This is often not the case in T&E, since many factors cannot be easily controlled or changed (e.g., weather, test range location). Therefore, designs including blocking and/or split-plot techniques should be considered. The execution of the test, including run plans/order, should be discussed in the Test Plan.

Commonly, the system under test (SUT) is a complex system with multiple missions and functionalities. The test design should reflect the complexity of the system. Often, multiple test designs will be necessary to fully characterize SUT mission performance. This might also require multiple experimental designs to capture all stages or aspects of mission execution.

Table 1: DOE Information Content for the TEMP

Milestone Supported	Information Content
A	Identify responsibilities of T&E WIPT for test design purposes The goal(s) to be addressed at each stage of testing Metrics for each goal/question Initial listing of factors Language for the overall testing strategy, including: • Screening experiments to ensure important factors are considered in operational testing • Sequential experimentation
B	Identify responsibilities of T&E WIPT for test design purposes The goal(s) to be addressed at each stage of testing Metrics for each goal/question Refined listing of factors and levels Test designs to support resourcing for limited user tests (LUT) and operational assessments (OA)

	Language for the overall testing strategy, including: • Screening experiments to ensure important factors are considered in operational testing • Sequential experimentation
C	Identify responsibilities of T&E WIPT for test design purposes The goal(s) to be addressed at each stage of testing, focusing on IOT&E Metrics for each goal/question Refined listing of factors and levels, based on prior testing and the operational mission. Details on how the factors and levels will be varied and controlled during each stage of testing Complete test designs to support resourcing for IOT&E Language for the overall testing strategy, including: • How previous knowledge is being used to inform IOT&E test planning. • Analysis plans to support power calculations

References

Guidance on the use of Design of Experiments (DOE) in Operational Test and Evaluation, DOT&E, October 19, 2010

Montgomery, D. C. (2009), *Design and Analysis of Experiments,* John Wiley and Sons

Myers, R. H., and Montgomery, D. C. (2002), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments,* John Wiley and Sons.

Design of Experiments – TEMP Body Example

3.2 Design of Experiments

Design and Analysis of Experiments will be used to develop test plans for the developmental, integrated, and operational testing of system XYZ. The T&E WIPT will identify the following components of the experimental design: (1) goals, (2) metrics, (3) factors and levels that impact the outcome of the test, (4) a strategic method for varying those factors and levels across all tests, and (5) appropriate statistical power and confidence levels for important responses for which they make sense. The T&E WIPT will use a sequential approach in test planning. The test plan outlined in this TEMP is adequate to support the OTA's evaluation plan. The evaluation plan is intended to provide a transparent, repeatable, and defensible approach to evaluation. Table 3.X provides the overall DOE strategy.

Table 3.X: Overview of DOE Strategy

		Test Phase			
		DT	MS	IT	IOT
Critical Responses		Select MOE, MOP, MOS, KPP	Select MOE, MOP, MOS, KPP	Select MOE, MOP, MOS, KPP	Select MOE, MOP, MOS, KPP
Factors	Factor Levels				
Factor 1	Categorical 2 levels	Systematically Vary (SV)	SV	SV	Record (allow to vary with operational mission)
Factor 2	Continuous	Hold Constant (HC)	HC	SV	SV
Factor 3	Continuous	SV	SV	SV	SV
Factor 4	Categorical 6 levels	SV	SV	SV	SV

Factor 5	Categorical 3 levels	HC	HC	SV	SV
Factor 6	Continuous	SV	SV	Record	Record

The factor management strategy may change after the initial test events are conducted to allow for increased information on the effect of the factors on the critical responses. See Appendix D for supporting data (factor selection, process diagrams, exact designs, and power/confidence levels).

Design of Experiments – Appendix D Example

DESIGN OF EXPERIMENTS (for a Precision Guided Weapon)

D.1 Design of Experiments (DOE) Definitions

This appendix uses terminology specific to DOE; the following definitions should be applied while reading.

- Initial Factor – A factor determined to potentially impact the performance of the precision guided weapon system in which the weapon system operates. Initial factors are pulled from the test design framework developed by the OTA or from subject matter expert inputs. Initial factors are accepted on their own, combined with other initial factors and accepted, placed in recordable status, determined to be a demo item, or eliminated from consideration for the DOE design.
- Accepted Factor – a factor accepted as a standalone from an initial factor or through the combination of multiple initial factors. Accepted factors were input into JMP software to create the DOE. Accepted factors are given levels.
- Level – the regions or levels that would be input into JMP software to create the DOE tables. Each accepted factor has a minimum of two levels.
- Recordable (Non-DOE) factor – a factor for which data are recorded during testing, but is not included in the DOE design. Factors that cannot be controlled, but might impact the performance the weapon system are placed into this category. These factors and their values will be recorded and compared against the performance of the weapon system to determine the impact they may have on the system.
- Demo Items – a factor or particular capability that will be tested against but is not incorporated into the DOE design created with JMP software. Demo items will be tested in standalone events if deemed to impact response variable, or incorporated into the DOE events when deemed to not impact response variable.
- Strike Warfare (STW) – the precision guided weapon system when used against Stationary Land Targets (SLT).
- Surface Warfare (SUW) – the precision guided weapon system when used against Moving Maritime Targets (MMT).

D.2.0 DOE (DT&IT)

The scope of DT&E and IT&E flight test planning for the precision guided weapon system against Moving Maritime Targets (MMT) was developed using DOE. A significant amount of data from previous testing of this precision guided weapon system exists, which helped to refine the test design.

The precision guided weapon system digital simulation consists of high fidelity guidance and electronics unit (GEU) and seeker models coupled with a target scene generator. The scene generator creates a perspective projection of the infrared target scene as presented to the seeker optics; the scenes are developed from empirical data and incorporate environmental effects such as time of day, sea state, humidity, and atmospheric conditions. Seeker imagery and GEU performance data captured during MMT captive carry flight testing has been used to successfully validate the all digital precision guided weapon system simulation. Following analysis of the historical data to date, the precision guided weapon system Technical Program Office, Lead Test Engineers, Systems Engineers, OTA testers, and DOE Subject Matter Experts determined the factors leading to an effect on the response variable. The response variable is the distance between seeker aimpoint and the preplanned aimpoint at the final seeker aimpoint refinement, both the captive carry (CC) and Free Flight (FF) live fire shots. For FF live fire shots the miss distance, assuming a normal distribution, will be recorded as the distance between the preplanned aimpoint and the actual impact point of the weapon for analysis purposes.

The initial set of factors selected for MMT were then ranked based on their predicted impact to the response variable and their intended use in the design. Table F-1 presents the factors.

D.2.1 Initial Factors (MMT)

Thermal contrast – Thermal contrast (temperature delta between the target and the background) allows the precision guided weapon system seeker to see the target and establish an aimpoint.

Day/Night – Sunlight will heat the target which will affect seeker performance. The absence of sunlight at night will cause lower Infra-red (IR) loading on the target and lower thermal contrast between the target and the background.

Glint – Glint, the reflection of light on the ocean surface, can affect the precision guided weapon system seeker and its ability to refine its aimpoint.

Target Speed – Target speed will affect the target location during weapon time of flight and the requirement for the data link to provide target location updates to the weapon. Target speed will also affect the delivery platform’s ability to locate and track the target and the precision guided weapon system seeker’s ability to refine its aimpoint.

Target Size – Target size will affect the delivery platforms’s ability to locate and track the target and the seeker’s ability to refine its aimpoint.

Target Aspect – Target aspect was chosen as an initial factor because it will affect the seeker's ability to refine the aimpoint as the weapon approaches the target.

Table D-1. MMT DOE for DT&E and IT&E

MMT DOE FACTORS (DT/IT)		
INITIAL FACTORS	**ACCEPTED FACTORS**	**LEVELS**
Thermal Contrast Day/Night Glint	Sun Elevation	≤ 1/2 Peak Rising - 1 > 1/2 Peak Rising - 2 > 1/2 Peak Setting - 3 ≤ 1/2 Peak Setting - 4 Night - 5
Target Speed Target Size	Target Type	Small (≤ ft) & Slow (≤ knots) Small (≤ ft) & Fast (> knots) Large (> ft) & Slow (≤ knots)
Target Aspect	Target Aspect	Head (0) Beam (90/270) Qtr (45/135/225/315) Tail (180)
TGT Maneuvering	TGT Maneuver	Evasive S Turn Non-maneuvering (constant course and speed)
RECORDABLE (NON-DOE)		
Sea State	Thermal Crossover	Humidity
DEMO ITEMS		
Multi Weapons Weapon Datalink	Datalink Source IRCM	Search Altitude WPN/Datalink RNG

Target Maneuvering – Target maneuvering will affect the precision guided weapon system's ability to refine the aimpoint as the target maneuvers within the seeker field of view. Also, target maneuvering will impact the requirement to provide updated target location and speed information to the precision guided weapon system during time of flight.

Sea State – Sea State will affect the precision guided weapon system seeker's ability to refine its aimpoint.

Thermal Crossover – Thermal Crossover was chosen as an initial factor because it will determine the temperature difference between the target and the background. This temperature difference will affect the seeker performance and ability to refine its aimpoint.

Humidity – Humidity was chosen as an initial factor because the humidity present in the atmosphere will affect the range at which the precision guided weapon system seeker will see the target and make aimpoint refinements.

Multiple weapons – Multiple weapons in flight should operate independently without influencing one another. On the chance that multiple weapons could influence one another and because the operational environment requires engagement of multiple simultaneous targets, some test events should include the near simultaneous launch of multiple weapons.

Datalink Source – The Datalink source will influence the target location error provided to the precision guided weapon system with each In-flight Target Update (IFTU). Target Location Error (TLE) will impact the seeker's ability to locate the target and refine its aimpoint. The datalink source has two levels: radar or targeting sensor.

Search Altitude – Search altitude will impact the initial image the seeker will compare against for aimpoint refinement.

Weapon Datalink – The Weapon Datalink performance is the method by which the delivery platform will provide target location information to the precision guided weapon system.

Infra-red Countermeasures (IRCM) – IRCM may affect the seeker's ability to refine its aimpoint.

Weapon to Datalink Range – Weapon datalink range will impact the target location error being provided the precision guided weapon system with each IFTU. TLE will impact the seeker's ability to locate the target and refine aimpoint.

D.2.2 Accepted Factors (MMT*)*

Sun Elevation – sun elevation is directly related to the thermal contrast and glint. Thermal contrast will be recorded during each run and will be used for analysis purposes. Sun elevation will be used for test planning and design purposes, while thermal contrast and glint will be used for analysis. Sun elevation is a five level factor. Four of the levels consider the presence of the sun (< ½ peak elevation rising, > ½ peak elevation rising, > ½ peak elevation setting, and < ½ peak elevation setting), and one level considers the absence of sun (night). Thermal crossover occurs twice daily and very quickly with a target at sea, and thermal contrast is consistent throughout the night until the sun rises in the morning. Using four levels of sun elevation in the daytime guarantees tests with high and low thermal contrast combined with multiple glint conditions.

Target Type – Target type is a combination of target size and target speed, a four-level factor combining slow (< 15 knots), fast (> 15 knots), small (< 100 feet), and large (>100 feet). The surrogate target types for DT&E and IT&E are the Mobile Ship Target (MST) (large and slow) and QST-35 Septar (QST) (small and fast). No DT target is available to support a combination of large and fast. Targets of opportunity with this combination will be explored to the greatest extent possible.

Target Aspect – Target aspect is a four-level factor: head-on (0 degrees), beam (90 or 270 degrees), quarter (45, 135, 225, or 315 degrees), and tail-on (180 degrees).

Target Maneuver – Target Maneuver is a two-level factor of yes or no. Yes is defined as an evasive S-turn; no is defined as maintaining constant course and speed.

D.2.3 Recordable Items (MMT)

Sea State – Sea state will be recorded during each run and will be limited, in the interest of manned target safety, to three levels: sea state 1, 2, or 3.

Thermal Crossover – The time and conditions (sun elevation, thermal contrast, and glint) of thermal crossover will be recorded. Predictions for the time of thermal crossover will be performed prior to each test. These predictions will be used to schedule testing during thermal crossover throughout the day in a variety of conditions (sun elevation, thermal contrast, and glint).

Humidity – Humidity will be a recorded item. DT&E and IT&E MMT flight testing will occur at different locations and times of year which will cause humidity variations throughout testing. Data from the precision guided weapon system Digital Simulation will augment captured flight test data for humidity levels not observed during testing.

D.2.4 Demo Items (MMT)

Multiple Weapons – This factor was not included in the DOE and is not expected to influence the response variable, but will be demonstrated during some of the test runs.

Datalink Source – The majority of runs will be performed using radar in the delivery platform to provide target location. A small sampling of runs will utilize the targeting sensor on the delivery platform to locate and track the target. This factor was not included in the DOE and is not expected to impact the response variable, but this capability will be demonstrated during testing during some of the DOE runs.

Weapon Search Altitude – The majority of runs will be performed using the optimum search altitude. A small sampling of runs will vary search altitude. This factor was not included in the DOE and is not expected to impact the response variable. Precision guided weapon system performance at various altitudes has been well characterized during developmental testing and in the digital simulation.

Weapon Datalink (WDL) – Every MMT run will require the WDL to perform several functions to ensure success. The WDL has already undergone extensive testing. This factor was not included in the DOE and is not expected to impact the response variable.

IRCM – IRCM may be evaluated during events to demonstrate the seeker's capability to refine the aimpoint when presented with false IR energy. The potential impact on the response variable may be evaluated and demonstrated during standalone testing, independent of the DOE described here.

Weapon/Datalink Range – The range between the weapon in flight and the datalink providing the IFTU will be recorded for every MMT run. This range will be varied but is not expected to impact the response variable.

D.2.5 DT/IT Power, Confidence, and Matrix for DOE Runs (MMT)

Using the accepted factors and assuming a normal distribution, the test design was created with JMP software for MMT using a D-optimal design for main effects and two-way interaction estimates. The matrix created includes 32 runs using 80% confidence and yielded a power of test of 93.4% to detect 0.5 sigma degradation above threshold accuracy requirements. The data will be collected during 32 captive carry runs. In addition to these 32 (20 DT&E, 12 IT&E) data runs, there will be 8 (4 DT&E, 4 IT&E) captive carry dress rehearsals and 4 (2 DT&E, 2 IT&E) free flight live fire runs where the data will be recorded during the MMT DT/IT testing.

The overall average miss distance will be compared against threshold values for the system to support the evaluation of the precision guided weapon system CPD requirements. ANOVA and regression analysis will also be performed based on the results. The analysis will provide additional evaluation understanding of overall system capabilities and limitations.

D.3.0 Operational Test DOE Development

In order to better evaluate precision guided weapon system performance in the STW and SUW operational environments, two distinct mission-based DOEs were developed: one for engaging stationary land targets (SLT) and one for engaging MMTs. Since the STW and SUW missions and requirements for precision guided weapon system employment are so different, one combined DOE would not adequately test the system.

STW requires the delivery platform to fly to the release point and launch the precision guided weapon system with prelaunch coordinates entered into the weapon. When the weapon approaches the target, the seeker will refine the flight profile to ensure the precision guided weapon system strikes the desired impact point on a stationary target. The precision guided weapon system incorporates a new seeker design.

SUW requires the delivery platform to detect the target with either a radar or targeting sensor, fly to the release point, and launch the precision guided weapon system. The delivery platform provides IFTU support to get the precision guided weapon system as close as possible to the MMT. As the weapon approaches the MMT, the seeker takes over, refining the flight profile in the final miles to ensure the precision guided weapon system strikes at the desired impact point on a moving target. These two distinct missions are described in detail below.

D.3.1 Operational Test DOE (STW)

Using DOE, the OT team leveraged the knowledge base from previous precision guided weapon system testing in developing the streamlined STW test design. The following assumptions provided the foundation for selecting the factors and levels for the test design: the weapons procedures for employment against SLT remained unchanged from the legacy precision guided weapon system; the weapon Launch Area Region (LAR), release and separation characteristics from the launch aircraft, and warhead capabilities remained the same; the new seeker capabilities and limitations will be compared against the legacy precision guided weapon system seeker; and the same target set will be used for the comparison of seeker performance data as much as possible. Additionally, the DOE factors considered known capabilities and limitations of the legacy precision guided weapon system seeker.

The precision guided weapon system test design was created primarily for Captive Carry (CC) runs. Replication was used to increase the understanding of the effects size and variability of data for specific test runs while increasing the statistical power and confidence of the test. The breadth of the design, coupled with the ease of performing multiple CC runs in a short period of time against SLTs in STW scenarios, facilitated replication in a cost efficient matter. With targets grouped together in a target area it is possible to fly against three or four different targets during an event, but not possible to transit to a new area during the course of one flight. It was deemed effective and efficient to fly three runs against each target in the target area, allowing nine runs or greater to be performed during each flight.

Outside of the primary DOE for CC runs, a robust test against Global Positioning System (GPS) jamming and Infra-red Countermeasures (IRCM) was also developed. This test will be used to demonstrate the specific effects of GPS denial, IRCM, and camouflage on the precision guided weapon system seeker. The performance of the precision guided weapon system will be compared directly against the legacy system in this same environment.

In addition to the CC STW DOE matrix and the CC test against GPS jamming/IRCM described above, data from two Free Flights (FF)/live fire (performed in IT) will be evaluated and compared with the results from the CC runs. Each of the FF/live fire shots will have CC dress rehearsal runs performed prior to the weapon release. These CC dress rehearsal runs will occur on a flight prior to the actual FF event to run through the FF scenario and ensure pilot familiarization with the event. The data gathered during the CC dress rehearsal and the CC runs just prior to the launch will also be used to compare with previous data gathered during the CC DOE and CC test against GPS jamming.

D.3.1.1 Operational Test Response Variable (STW)

The STW designs support measurement of weapon accuracy. For the CC runs and FF/live fire shots, the response variable is the distance between seeker aimpoint and preplanned aimpoint at the final seeker aimpoint refinement. Using this miss distance as the response variable while assuming a normal distribution, the factors and levels were selected as described below. For FF/live fire shots, the distance between the preplanned aimpoint and the actual impact point of the weapon will also be recorded for analysis purposes. The assumption of a normal distribution of the data was made during the planning phase in coordination with DOE subject matter experts at both the OTA and DOT&E.

D.3.1.2 Operational Test Factors and Levels (STW)

Initially several factors were considered as being able to affect the response variable of precision guided weapon system in the STW mission. These factors were reviewed and sorted based on their anticipated effect on the response variable. Table D-2 presents the factors for STW during OT&E.

Table D-2. OT&E Factors and Levels for STW

STW DOE FACTORS (OT)		
INITIAL FACTORS	ACCEPTED FACTORS	LEVELS
Terrain	Terrain	Desert Mountain Urban Littoral
Target Orientation	Target Orientation	Horizontal Face Vertical Face
Clutter Civil Structures Snow	Contrast	High Low
Thermal Contrast	Sun Elevation	<1/2 peak AM or PM >1/2 peak AM or PM
RECORDABLE (NON-DOE)		
Thermal Crossover		Humidity
DEMO ITEMS		
IRCM	Camouflage Day/Night	GPS jamming

D.3.1.3 Operational Test Power, Confidence, and Matrix for DOE Runs (STW)

Using the factors above and assuming a normal distribution, the design was created with JMP for STW using a full factorial design for main effects and two-way interaction estimates. The matrix created includes 32 runs, which will each be replicated three times, for a total of 96 runs. The replications are a result of efficient use of flight sortie time by repeating runs rather than repeating flights. This design used 80 percent confidence level and yielded a power of test of greater than 99 percent to detect a 0.5 sigma degradation above threshold accuracy requirements. The runs are displayed in Table D-3.

The overall average miss distance will be compared against threshold values for the system to support the evaluation of the precision guided weapon system CPD requirements. ANOVA and regression analysis will be performed as well, based on the results. The analysis will provide additional understanding of overall system capabilities and limitations.

Table D-3. OT&E STW Run Matrix

OT STW Matrix Full Factorial						
High Humidity Det						
Run	Sun Elevation	Orientation	Contrast	Humidity	Terrain	Actual Target
1-3	<1/2 max	Horizontal	Low	High	Littoral	Corpus Christi Command Center Wall
4-6	<1/2 max	Horizontal	High	High	Littoral	Corpus Christi Hangar
7-9	<1/2 max	Vertical	Low	High	Littoral	Corpus Christi Small Building on Pier
10-12	<1/2 max	Vertical	High	High	Littoral	Corpus Christi Tower
13-15	<1/2 max	Horizontal	High	High	Urban	Orange Grove Roof of NE Bldg
16-18	<1/2 max	Horizontal	Low	High	Urban	Orange Grove Airfield Arresting gear building
19-21	<1/2 max	Vertical	Low	High	Urban	Orange Grove ILS Radar
22-24	<1/2 max	Vertical	High	High	Urban	Target TBD
25-27	>1/2 max	Horizontal	Low	High	Littoral	Corpus Christi Command Center Wall
28-30	>1/2 max	Horizontal	High	High	Littoral	Corpus Christi Hangar
31-33	>1/2 max	Vertical	Low	High	Littoral	Corpus Christi Small Building on Pier
34-36	>1/2 max	Vertical	High	High	Littoral	Corpus Christi Tower
37-39	>1/2 max	Vertical	High	High	Urban	Orange Grove Roof of NE Bldg
40-42	>1/2 max	Horizontal	Low	High	Urban	Orange Grove Airfield Arresting gear building
43-45	>1/2 max	Vertical	Low	High	Urban	Orange Grove ILS Radar
46-48	>1/2 max	Horizontal	High	High	Urban	Target TBD
Low Humidity						
Run	Sun Elevation	Orientation	Contrast	Humidity	Terrain	Actual Target
49-51	<1/2 max	Horizontal	High	Low	Mountain	Independence Courthouse Multi level Building
52-54	<1/2 max	Horizontal	Low	Low	Mountain	Independence Jailhouse Large building
55-57	<1/2 max	Vertical	Low	Low	Mountain	Independence Microwave Tower
58-60	<1/2 max	Vertical	High	Low	Mountain	Target TBD
61-63	<1/2 max	Horizontal	Low	Low	Desert	Trona Large Yellow Building
64-66	<1/2 max	Horizontal	High	Low	Desert	Trona Movie Theater
67-69	<1/2 max	Vertical	High	Low	Desert	Trona Post Office Wall
70-72	<1/2 max	Vertical	Low	Low	Desert	Ballarat Radar/R2508
73-75	>1/2 max	Horizontal	High	Low	Mountain	Independence Courthouse Multi level Building
76-78	>1/2 max	Horizontal	Low	Low	Mountain	Independence Jailhouse Large building
79-81	>1/2 max	Vertical	Low	Low	Mountain	Independence Microwave Tower
82-84	>1/2 max	Vertical	High	Low	Mountain	Target TBD
85-87	>1/2 max	Horizontal	Low	Low	Desert	Trona Large Yellow Building
88-90	>1/2 max	Horizontal	High	Low	Desert	Trona Movie Theater
91-93	>1/2 max	Vertical	High	Low	Desert	Trona Post Office Wall
94-96	>1/2 max	Vertical	Low	Low	Desert	Ballarat Radar/R2508

D.3.1.4 Matrix for Demo and Countermeasure Runs (STW)

The STW demonstration items (IRCM, GPS jamming, GPS availability, and camouflage) will be demonstrated during the following 30 runs, which are displayed in Table D-4.

- Twelve runs versus GPS jamming in mountainous terrain (six against co-altitude jamming)

- Twelve runs in R-2505 versus multiple countermeasures in the White Sands area
- Six runs in R-2505 versus multiple IR countermeasures.

Table D-4. OT&E STW Demo Run Matrix

Advanced Countermeasures								
Run	Sun Elevation	Orientation	Contrast	Humidity	Terrain	Actual Target	Jamming Profile	Countermeasure
1	>1/2 max	Vertical	High	Low	Mountain	GPS Jamming Parrot Peak Radar dish	25K to 20 degree	
2	>1/2 max	Vertical	High	Low	Mountain	GPS Jamming Parrot Peak Radar dish	25K to 20 degree	
3	>1/2 max	Vertical	High	Low	Mountain	GPS Jamming Parrot Peak Radar dish	25K to 20 degree	
4	<1/2 max	Vertical	High	Low	Mountain	GPS Jamming Parrot Peak Radar dish	Co altitude	
5	<1/2 max	Vertical	High	Low	Mountain	GPS Jamming Parrot Peak Radar dish	Co altitude	
6	<1/2 max	Vertical	High	Low	Mountain	GPS Jamming Parrot Peak Radar dish	Co altitude	
7	>1/2 max	Horizontal	High	Low	Mountain	GPS Jamming Parrot Peak Building roof	25K to 20 degree	
8	>1/2 max	Horizontal	High	Low	Mountain	GPS Jamming Parrot Peak Building roof	25K to 20 degree	
9	>1/2 max	Horizontal	High	Low	Mountain	GPS Jamming Parrot Peak Building roof	25K to 20 degree	
10	<1/2 max	Horizontal	High	Low	Mountain	GPS Jamming Parrot Peak Building roof	Co altitude	
11	<1/2 max	Horizontal	High	Low	Mountain	GPS Jamming Parrot Peak Building roof	Co altitude	
12	<1/2 max	Horizontal	High	Low	Mountain	GPS Jamming Parrot Peak Building roof	Co altitude	
13	<1/2 max	Vertical	Low	Low	Desert	2505 Sams Town T-Building	Point	Multiple/White Sands
14	<1/2 max	Vertical	Low	Low	Desert	2505 Sams Town T-Building	Point	Multiple/White Sands
15	<1/2 max	Vertical	Low	Low	Desert	2505 Sams Town T-Building	Point	Multiple/White Sands
16	>1/2 max	Vertical	Low	Low	Desert	2505 Sams Town T-Building	Point	Multiple/White Sands
17	>1/2 max	Vertical	Low	Low	Desert	2505 Sams Town T-Building	Point	Multiple/White Sands
18	>1/2 max	Vertical	Low	Low	Desert	2505 Sams Town T-Building	Point	Multiple/White Sands
19	<1/2 max	Horizontal	Low	Low	Desert	2505 Sams Small Building 1 Story	Point	Multiple/White Sands
20	<1/2 max	Horizontal	Low	Low	Desert	2505 Sams Small Building 1 Story	Point	Multiple/White Sands
21	<1/2 max	Horizontal	Low	Low	Desert	2505 Sams Small Building 1 Story	Point	Multiple/White Sands
22	>1/2 max	Horizontal	Low	Low	Desert	2505 Sams Small Building 1 Story	Point	Multiple/White Sands
23	>1/2 max	Horizontal	Low	Low	Desert	2505 Sams Small Building 1 Story	Point	Multiple/White Sands
24	>1/2 max	Horizontal	Low	Low	Desert	2505 Sams Small Building 1 Story	Point	Multiple/White Sands
25	<1/2 max	Vertical	Low	Low	Desert	2505 POL Coles Flat	Point	Laser CM and Flames
26	<1/2 max	Vertical	Low	Low	Desert	2505 POL Coles Flat	Point	Laser CM and Flames
27	<1/2 max	Vertical	Low	Low	Desert	2505 POL Coles Flat	Point	Laser CM and Flames
28	>1/2 max	Vertical	Low	Low	Desert	2505 POL Coles Flat	Point	Laser CM and Flames
29	>1/2 max	Vertical	Low	Low	Desert	2505 POL Coles Flat	Point	Laser CM and Flames
30	>1/2 max	Vertical	Low	Low	Desert	2505 POL Coles Flat	Point	Laser CM and Flames

D.3.2 Operational Test DOE (SUW)

Using DOE, the OT team extensively leveraged the knowledge base from previous precision guided weapon system testing in developing the streamlined SUW test design. The following assumptions provided the foundation for selecting the factors and levels for the precision guided weapon system SUW test design: the weapon Launch Area Region (LAR), release and separation characteristics from the launch aircraft, and warhead capabilities remained the same; the new seeker capabilities and limitations will be compared against the legacy precision guided weapon system seeker. Additionally, the DOE factors included limitations of the legacy precision guided weapon system seeker.

The precision guided weapon system SUW test design was created primarily for CC runs. Replication was not used due to the large number of factors to be tested against and the difficulty in performing each run.

In addition to the CC SUW DOE matrix, data from two FF/live fire shots being performed in IT and data from two FF/live fire shots being performed in OT will be evaluated and compared with the results from CC runs. Each of the FF/live fire shots will have CC runs performed prior to the weapon release. These CC dress rehearsal runs will occur on a flight prior to the actual FF event. During the event for the FF/live fire shot, the profile will be flown CC a few times to ensure everything is working properly. The data gathered during the dress rehearsal and the CC runs prior to the launch will also be compared with previous data gathered during the CC DOE matrix.

D.3.2.1 Operational Test Response Variable (SUW)

The SUW designs support measuring weapon accuracy. For the CC runs and FF/live fire shots the response variable is the distance between seeker aimpoint and preplanned aimpoint at the final seeker aimpoint refinement. Using this miss distance as the response variable while assuming a normal distribution, the factors and levels were selected as described below. For FF/ live fire shots, the distance between preplanned aimpoint and actual impact point of the weapon will also be recorded for analysis purposes. The assumption of a normal distribution of the data was made during the planning phase in coordination with DOE subject matter experts at both the OTA and DOT&E.

D.3.2.2 Operational Test Factors and Levels (SUW)

Initially several factors were considered to affect the response variable of precision guided weapon system in the SUW mission. These factors were reviewed and sorted based on their anticipated effect on the response variable. Table D-5 presents the factors for SUW during OT&E.

Table D-5. OT&E Factors and Levels for SUW

SUW DOE FACTORS (OT)		
INITIAL FACTORS	ACCEPTED FACTORS	LEVELS
Thermal Contrast Day/Night Glint	Sun Elevation	≤ 1/2 Peak Rising - 1 > 1/2 Peak Rising - 2 > 1/2 Peak Setting - 3 ≤ 1/2 Peak Setting - 4 Night - 5
Target Speed Target Size	Target Type	Small (≤100 ft) & Slow (≤ 15 knots) Small (≤100 ft) & Fast (> 15 knots) Large (>100 ft) & Slow (≤ 15 knots) Large (>100 ft) & Fast (> 15 knots)
Threat WPN Range Target Slant Range	Target Range	≤ 40 nm > 40 nm
Target Aspect	Target Aspect	Head (0) Beam (90/270) Qtr (45/135/225/315) Tail (180)
TGT Maneuvering RFCM GPS Jamming	Location Defenses	Yes
IRCM Camouflage Shipping presence	Seeker Defenses	Yes No
RECORDABLE (NON-DOE)		
Sea State	Thermal Crossover Humidity	Glint
DEMO ITEMS		
Multi-Weapons	Datalink Source	Weapon Datalink

D.3.2.3 Operational Test Power, Confidence, and Matrix for DOE Runs (SUW)

Using these factors and assuming a normal distribution, the design was created with JMP for SUW using a D-optimal design for main effects and two-way interaction estimates. The matrix created includes 64 runs using 80% confidence and yielded a power of test of greater than 93.4% to detect a 0.5 sigma degradation above threshold accuracy requirements. The runs are displayed in Table D-6.

Table D-6. OT&E SUW Run Matrix

OT SUW Matrix D-Optimal Design															
Low Humidity								High Humidity Det							
Run	Sun Elev.	Tgt Aspect	Tgt Type	Datalink Range	Humidity	Location Defenses	Seeker Defenses	Run	Sun Elev.	Tgt Aspect	Tgt Type	Datalink Range	Humidity	Location Defenses	Seeker Defenses
1	1	Tail	Large/Fast	Long	Low	No	Yes	33	1	Tail	Small/Fast	Long	High	Yes	No
2	1	Head	Large/Fast	Short	Low	No	Yes	34	1	Beam	Large/Fast	Short	High	No	No
3	1	Tail	Large/Fast	Long	Low	No	Yes	35	1	Qtr	Small/Fast	Short	High	Yes	Yes
4	1	Qtr	Large/Slow	Long	Low	No	No	36	1	Beam	Large/Slow	Long	High	Yes	No
5	1	Beam	Large/Slow	Long	Low	Yes	No	37	1	Head	Large/Slow	Long	High	Yes	Yes
6	1	Head	Large/Slow	Short	Low	Yes	No	38	1	Qtr	Small/Slow	Short	High	No	Yes
7	2	Head	Large/Fast	Long	Low	No	Yes	39	2	Beam	Small/Fast	Long	High	No	Yes
8	2	Head	Large/Fast	Long	Low	No	No	40	2	Qtr	Small/Fast	Long	High	Yes	No
9	2	Tail	Large/Fast	Long	Low	Yes	No	41	2	Tail	Small/Fast	Short	High	No	No
10	2	Head	Large/Fast	Short	Low	Yes	No	42	2	Head	Large/Slow	Long	High	Yes	No
11	2	Tail	Large/Slow	Long	Low	Yes	Yes	43	2	Qtr	Large/Slow	Short	High	Yes	Yes
12	2	Beam	Large/Slow	Short	Low	No	Yes	44	2	Tail	Large/Slow	Short	High	No	Yes
13	2	Qtr	Large/Slow	Short	Low	Yes	No	45	2	Beam	Small/Slow	Short	High	No	Yes
14	3	Qtr	Large/Fast	Long	Low	Yes	No	46	3	Qtr	Large/Fast	Long	High	No	Yes
15	3	Qtr	Large/Fast	Short	Low	Yes	Yes	47	3	Head	Small/Fast	Long	High	No	No
16	3	Tail	Large/Fast	Short	Low	Yes	No	48	3	Qtr	Large/Fast	Short	High	No	No
17	3	Tail	Small/Slow	Long	Low	No	No	49	3	Tail	Small/Fast	Short	High	Yes	Yes
18	3	Beam	Small/Slow	Long	Low	Yes	No	50	3	Tail	Small/Slow	Long	High	No	Yes
19	3	Head	Small/Slow	Long	Low	No	Yes	51	3	Beam	Large/Slow	Short	High	No	No
20	3	Head	Small/Slow	Short	Low	Yes	Yes	52	3	Beam	Small/Slow	Short	High	Yes	Yes
21	4	Beam	Small/Fast	Long	Low	Yes	Yes	53	4	Beam	Large/Fast	Long	High	Yes	Yes
22	4	Beam	Small/Fast	Short	Low	Yes	No	54	4	Qtr	Large/Fast	Long	High	No	No
23	4	Tail	Small/Fast	Short	Low	No	No	55	4	Head	Small/Fast	Short	High	Yes	Yes
24	4	Qtr	Small/Slow	Long	Low	No	Yes	56	4	Tail	Large/Slow	Long	High	Yes	No
25	4	Beam	Small/Slow	Short	Low	No	No	57	4	Head	Small/Slow	Long	High	No	Yes
26	4	Qtr	Small/Slow	Short	Low	Yes	Yes	58	4	Head	Large/Slow	Short	High	No	No
27	5	Beam	Small/Fast	Long	Low	No	Yes	59	5	Beam	Large/Fast	Long	High	No	No
28	5	Beam	Small/Fast	Short	Low	Yes	Yes	60	5	Head	Large/Fast	Long	High	Yes	Yes
29	5	Qtr	Small/Fast	Short	Low	No	No	61	5	Head	Small/Fast	Short	High	Yes	No
30	5	Qtr	Small/Slow	Long	Low	No	Yes	62	5	Qtr	Large/Slow	Long	High	Yes	No
31	5	Tail	Small/Slow	Long	Low	Yes	Yes	63	5	Tail	Small/Slow	Short	High	No	No
32	5	Head	Small/Slow	Short	Low	No	No	64	5	Tail	Small/Slow	Short	High	Yes	Yes

D.3.2.4 Additional SUW Runs

In addition to the 64 SUW test runs described above, a minimum of six CC runs will be conducted as dress rehearsal runs for the two free flight/live fire shots against MMT targets and then the two FF/live fire runs. The data will be recorded and compared to CC data. The specifics of these runs will be detailed in the Test Plan. See Table D-7.

Table D-7. OT&E SUW Free Flight

OT SUW Free Flight Matrix								
Run	Sun Elev.	Tgt Aspect	Tgt Type	Datalink Range	Humidity	Location Defenses	Seeker Defenses	Notes
65	2	Tail	Large/Slow	Long	Low	Yes	Yes	Dress
66	2	Tail	Large/Slow	Long	Low	Yes	Yes	Dress
67	2	Tail	Large/Slow	Long	Low	Yes	Yes	Dress
68	2	Tail	Large/Slow	Long	Low	Yes	Yes	Free Flight
69	3	Beam	Small/Fast	Short	Low	Yes	Yes	Dress
70	3	Beam	Small/Fast	Short	Low	Yes	Yes	Dress
71	3	Beam	Small/Fast	Short	Low	Yes	Yes	Dress
72	3	Beam	Small/Fast	Short	Low	Yes	Yes	Free Flight

D.3.3 Operational Test Data Analysis (STW & SUW)

The overall results of the response variable will be compared against threshold values for precision guided weapon system to support the resolution of COIs. ANOVA and regression analysis will be performed based on the results of the OT testing. This analysis will be utilized to understand system performance, the effects of the factors, and to provide tactical recommendations to the fleet operator in employment of precision guided weapon system.

Design of Experiments – Appendix D Example

DESIGN OF EXPERIMENTS (for a Milestone B ArtilleryHowitzer)

D.1 Design of Experiments (DOE) Overview

The purpose of this appendix is to provide a framework for the OTA's Design of Experiments (DOE) methodology in support of a howitzer acquisition. The OTA will plan and conduct both the LUT/OA/OA and the IOT using DOE principals. This method of assessment will provide a systematic approach to assess the effects of pre-determined factors on key performance aspects of the howitzer. The design goal is to vary key factors that affect measurable system characterizations such as timeliness and accuracy. Table D.1 below shows how the factors and factor levels will be controlled during each test event.

Table D.1: DOE Campaign Strategy

Factors	Factor Levels	Test Events	
		LUT /OA	IOT
Ammo-Lethal	Projectile 1(P1), Projectile 2(P2)	SV	SV
Ammo-Non Lethal	Smoke, Illum	Non-Lethal limited # missions	Non-Lethal limited # missions
Time	Day, Night	SV	SV
Range Band	C1 + C2, C3, C4, C5	SV	SV
Traverse	0-15, 15-45, Out of Sector	SV (0-15, 15-45), Out of Sector (limited # missions)	SV (0-15, 15-45), Out of Sector (limited # missions)
Angle	Low, High	SV	SV
Fuze	Time Delay (TD), Point Detonation(PD), Multi-option fuse (MOF)	SV	SV
MOPP	0, IV	HC-MOPP 0, MOPP IV limited # missions	HC-MOPP O, MOPP IV limited # missions
Test Elements	# of test elements	HC (1 Element)	SV (3 Elements)
IA	None, Red team	None	HC-None, Red team excursion at end of test
Notes/Definitions: *HC-Held Constant *SV – Systematically Varied *C1-MACS 1 or equivalent *C2-MACS 2 or equivalent *C3-MACS 3 or equivalent *C4-MACS 4 or equivalent			

*Ce
*High angle of fire – Above maximum range Quadrant of Elevation(>~800 mils)
*LowAngle of Fire – Below maximum range Quadrant of Elevation(<~800mils)
*IA – Information Assurance

LUT/OA/**OA:**

The objectives of the LUT/OA shall be to evaluate the howitzer interoperability, fire mission accuracy and responsiveness and automotive performance as well as mobility and reliability in support of combat operations. Table D.2 shows critical responses.

Table D.2: Critical Responses

Critical Responses	Accuracy (Miss Distance in meters, CEP)
	Timeliness (Time to Complete Mission in seconds)
	Reliability (Mean Time between Failure)

This phase of the operational testing will follow a D-optimal split-plot design of experiments approach with some of the hard to control factor systematically controlled to balance DOE and operational realism from the OMS/MP. Table D.3 lists the factors and levels for the two responses: accuracy and timeliness.

Table D.3: Factors and Levels

Factor	Levels	Control
Projectile	P1, P2	Hard, Systematic
Time	Day, Night	Hard, Systematic
Range Band	C1 + C2, C3, C4, C5	Hard, Systematic
Traverse Angle	0-15, 15-45	Hard
Angle of Fire	Low, High	Easy
Fuze Type	TD, PD, MOF	Hard

If a factor it systematically controlled it was organized in an operationally realistic manner yet based on a D-optimal design. Projectile, Time, and Range were organized so that it followed a scenario where it starts on closest range bands (C1 + C2) and then moves to the C5 range band over the first two 24-hour periods before returning to the initial bands over the next two 24-hour periods. If a factor was hard to control, these factors were randomized over whole plots (blocks of time where the time, Projectile, range band, traverse, and fuze could randomly be assigned). Angle is an easy to control so it could be randomly assigned to the individual missions or within the blocks. The DOE consists of 96 missions, but to meet the reliability requirements, 160 missions are necessary. These additional missions are distributed between special case requirements (Non-Lethal, emergency firings, MOPP IV, Out of Sections, and other long range missions to meet the MP/OMS. These additional missions will be injected into the DOE run matrix at the discretion of the Test Officer to ensure operational realism. For example, all the Out of Sector and Emergency missions will be conducted right after tactical moves. Table D.4 shows the breakout by mission.

Table D.4: Factor Breakout By Mission

	Range	Charge	P1 Missions	P2 Missions	Illum Missions	Smoke Missions	Total Missions
DOE	4 - 9 KM	1/2L	16	0	-	-	16
	9-12 KM	3H	16	0	-	-	16
	12-15 KM	4H	16	20	-	-	36
	16.4 - 20 KM	5H	-	28	-	-	28
Non-Lethal	TBD	TBD	-	-	3	3	6
emergency firings	16.4 - 20 KM	5H	-	12	-	-	12
MOPP IV	16.4 - 20 KM	5H	-	8	-	-	8
Additional Long range for RAM	16.4 - 20 KM	5H	-	26	-	-	26
Out of Sector	TBD	TBD	-	12	-	-	12
Total	-	-	48	108	3	3	160

The D-Optimal Split-Split Plot design permits the ability to estimate all main effects, all 2-way interactions with time, and the following additional interactions: range band and traverse, traverse and angle, angle and fuze, traverse and fuze, and projectile and angle. The run matrix, which it the required order that these runs must follow, is shown in table D.5 below.

Table D.5: LUT/OA D-Optimal Split-Split Plot Run Matrix

Day	**Time**	Projectile	**Range Band**	**Traverse**	**Angle**	**Fuze**
1	Day	P1	C1 + C2	0-15	High	TD
1	Day	P1	C1 + C2	0-15	Low	TD
1	Day	P1	C1 + C2	0-15	Low	TD
1	Day	P1	C1 + C2	0-15	High	TD
1	Day	P1	C1 + C2	0-15	High	PD
1	Day	P1	C1 + C2	0-15	Low	PD
1	Day	P1	C1 + C2	0-15	Low	PD
1	Day	P1	C1 + C2	0-15	High	PD
1	Day	P1	C3	30-45	Low	PD
1	Day	P1	C3	30-45	High	PD
1	Day	P1	C3	30-45	Low	PD
1	Day	P1	C3	30-45	High	PD
1	Night	P1	C3	0-15	High	TD

1	Night	P1	C3	0-15	High	TD
1	Night	P1	C3	0-15	Low	TD
1	Night	P1	C3	0-15	Low	TD
1	Night	P1	C4	30-45	High	TD
1	Night	P1	C4	30-45	High	TD
1	Night	P1	C4	30-45	Low	TD
1	Night	P1	C4	30-45	Low	TD
1	Day	P1	C4	0-15	Low	MOF
1	Day	P1	C4	0-15	Low	MOF
1	Day	P1	C4	0-15	High	MOF
1	Day	P1	C4	0-15	High	MOF
2	Day	P2	C4	30-45	High	MOF
2	Day	P2	C4	30-45	Low	MOF
2	Day	P2	C4	30-45	Low	MOF
2	Day	P2	C4	30-45	High	MOF
2	Day	P2	C4	30-45	Low	TD
2	Day	P2	C4	30-45	High	TD
2	Day	P2	C4	30-45	Low	TD
2	Day	P2	C4	30-45	High	TD
2	Night	P2	C5	30-45	Low	MOF
2	Night	P2	C5	30-45	Low	MOF
2	Night	P2	C5	30-45	Low	MOF
2	Night	P2	C5	30-45	Low	MOF
2	Night	P2	C5	30-45	Low	PD
2	Night	P2	C5	30-45	Low	PD
2	Night	P2	C5	30-45	Low	PD
2	Night	P2	C5	30-45	Low	PD
2	Night	P2	C5	0-15	Low	TD
2	Night	P2	C5	0-15	Low	TD
2	Night	P2	C5	0-15	Low	TD
2	Night	P2	C5	0-15	Low	TD
2	Night	P2	C5	30-45	Low	TD
2	Night	P2	C5	30-45	Low	TD
2	Night	P2	C5	30-45	Low	TD
2	Night	P2	C5	30-45	Low	TD
3	Day	P2	C5	0-15	Low	MOF
3	Day	P2	C5	0-15	Low	MOF
3	Day	P2	C5	0-15	Low	MOF
3	Day	P2	C5	0-15	Low	MOF
3	Day	P2	C5	30-45	Low	PD
3	Day	P2	C5	30-45	Low	PD

3	Day	P2	C5	30-45	Low	PD
3	Day	P2	C5	30-45	Low	PD
3	Day	P2	C5	0-15	Low	TD
3	Day	P2	C5	0-15	Low	TD
3	Day	P2	C5	0-15	Low	TD
3	Day	P2	C5	0-15	Low	TD
3	Day	P2	C4	0-15	High	PD
e	Day	P2	C4	0-15	High	PD
3	Day	P2	C4	0-15	Low	PD
3	Day	P2	C4	0-15	Low	PD
3	Night	P2	C4	0-15	Low	MOF
3	Night	P2	C4	0-15	High	MOF
3	Night	P2	C4	0-15	Low	MOF
3	Night	P2	C4	0-15	High	MOF
3	Night	P2	C4	0-15	Low	PD
3	Night	P2	C4	0-15	High	PD
3	Night	P2	C4	0-15	Low	PD
3	Night	P2	C4	0-15	High	PD
3	Night	P1	C4	0-15	High	PD
3	Night	P1	C4	0-15	Low	PD
3	Night	P1	C4	0-15	High	PD
3	Night	P1	C4	0-15	Low	PD
4	Day	P1	C4	30-45	Low	MOF
4	Day	P1	C4	30-45	High	MOF
4	Day	P1	C4	30-45	High	MOF
4	Day	P1	C4	30-45	Low	MOF
4	Day	P1	C3	30-45	Low	TD
4	Day	P1	C3	30-45	Low	TD
4	Day	P1	C3	30-45	High	TD
4	Day	P1	C3	30-45	High	TD
4	Night	P1	C3	30-45	High	MOF
4	Night	P1	C3	30-45	High	MOF
4	Night	P1	C3	30-45	Low	MOF
4	Night	P1	C3	30-45	Low	MOF
4	Night	P1	C1 + C2	30-45	High	PD
4	Night	P1	C1 + C2	30-45	Low	PD
4	Night	P1	C1 + C2	30-45	Low	PD
4	Night	P1	C1 + C2	30-45	High	PD
4	Night	P1	C1 + C2	0-15	Low	MOF
4	Night	P1	C1 + C2	0-15	High	MOF
4	Night	P1	C1 + C2	0-15	High	MOF

4	Night	P1	C1 + C2	0-15	Low	MOF

Power for the tests for all the factors to see how they influence the responses are listed below in Table D.6:

Table D.6: Power Effect on Factors and Responses

Effect	Variance	Power (90% Confidence, S:N=2)	Power (80% Confidence, S:N=1)
Intercept	0.228	0.994	0.789
Time	0.303	0.974	0.701
Range Band 1	0.333	0.963	0.671
Range Band 2	0.245	0.991	0.767
Range Band 3	0.180	0.999	0.855
Traverse	0.305	0.974	0.699
Angle	0.018	1.000	1.000
Fuze 1	0.208	0.997	0.816
Fuze 2	0.194	0.998	0.836
Projectile	0.390	0.937	0.624
Time*Range Band 1	0.559	0.842	0.524
Time*Range Band 2	0.273	0.984	0.733
Time*Range Band 3	0.147	1.000	0.906
Time*Traverse	0.208	0.997	0.816
Time*Angle	0.016	1.000	1.000
Time*Fuze 1	0.095	1.000	0.974
Time*Fuze 2	0.269	0.985	0.738
Time*Projectile	0.464	0.897	0.574
Range Band*Traverse 1	0.299	0.976	0.705
Range Band*Traverse 2	0.257	0.988	0.752
Range Band*Traverse 3	0.222	0.995	0.797
Traverse*Angle	0.016	1.000	1.000
Angle*Fuze 1	0.016	1.000	1.000
Angle*Fuze 2	0.014	1.000	1.000
Traverse*Fuze 1	0.145	1.000	0.908
Traverse*Fuze 2	0.182	0.999	0.852
Projectile*Angle	0.018	1.000	1.000

IOT:

Design of Experiments – Example for Software-Intensive System

(The following section would appear in the body of the TEMP for a Command and Control System at MS C. Appendix D material begins on page 3.)

3.2 TEST AND EVALUATION FRAMEWORK

The OTA must accomplish the following during IOT&E:

- Determine if thresholds in the approved capabilities documents and COIs have been satisfied
- Determine Operational Effectiveness, Survivability, and Suitability of the system under realistic operational conditions
- Assess the contribution of the system to combat operations
- Provide additional information on the system's operational capabilities and limitations.

Testing involves the physical exercising, by trial or examination, of a component, system, concept, or approach for the sole purpose of gathering data and information regarding the item under test. Evaluation seeks to ascertain the worth of, or to fix the value of, a component, system, concept, or approach. The relationship of testing to evaluation is many-to-one; that is, several tests may be required to support a single evaluation.

The OTA's evaluation plan creates a framework and methodology for evaluating the entirety of program data, obtained from assessments and IOT&E. The evaluation plan is intended provide a transparent, repeatable, and defensible approach to evaluation. The evaluation framework is captured in Table 3.1. The test team developed test concepts by employing Design of Experiments (DOE) to ensure that a rigorous methodology supports the development and analysis of test results. A designed experiment is used to determine the effect of a factor or several factors (also called independent variables) on one or more measured responses (also called dependent variables). All COI DOEs are designed with mission-oriented response variables. Each design will include an estimation of the power of the test. When gaps in the design are identified, these gaps will be listed as limitations, and a risk assessment will be provided in the appropriate Detailed Test Plan. In addition, the team will work with all appropriate parties to determine the most appropriate way to mitigate and/or manage the risks.

The OTA intends to exercise the command and control system during a training exercise (for a list of resources, see section 4.0). Although the OTA believes this event provides the necessary operational realism, some limitations to the event will affect the confidence and power of the results. The primary limitation is that the event does not allow for complete randomization of factors and levels; however, the exercise does provide some strategic variation due to the nature of the event.

The OTA has identified the response variables, factors and levels that will be exercised during each event in Table 3.1, in addition to the expected trial replications, which is outlined in Appendix A for the COIs. An optimal design has been provided (along with confidence levels and power) with the expected event replications. The identified confidence level and power are the maximums expected in a completely randomized event, due to restrictions in randomization. The major risk of not completely randomizing the design is that some factors may become confounded with uncontrollable variables. The OTA will work to avoid any obvious confounding of variables. As more information on the training exercise becomes available, design gaps will be identified and appropriately addressed in the Test Plan (to include the factors and levels).

Finally, a minimum of 228 hours of operation, spread across all of the systems employed operationally at the IOT&E, is required to evaluate RAM and Ao requirements.

Table 3.1: IOT&E Variables, Factors and Levels*

Factors	Levels	Mission
COI 1: Does the system support DASC air support and air strike requests? Response Variables: System Response Time (RT_D) and Probability of Success (P_S)		
Process Type	Pre-Planned, Immediate	Assault Support, Anti-air Warfare, Offensive Air Support, Air Reconnaissance and Electronic Warfare
Request Type	Air-strike, Air Support	
Communication Method	POTS, Chat, E-mail, Radio	
Request Flow	Low, Medium, High	
COI 2: Does the system support DASC coordination of supporting arms missions? Response Variables: System Coordination Time (CT_D) and Probability of Success (P_S)		
Communication Method	POTS, Chat, E-mail, Radio, AFATDS	Coordinate Supporting Arms
Request Flow	Low, Medium, High	
COI 3: Does the system support DASC air control missions? Response Variables: Probability of Success (P_S)		
Aircraft Type	Fixed Wing, Rotary Wing	Procedurally Control Aircraft
Sortie Flow	Low, Medium, High	
COI 4: Does the system support TAOC surveillance, direction and control missions? Response Variables: System Response Time (RT_T) and Probability of Success (P_S)		
Mission Type	Surveillance, Direction, Control	Air Direction, Supporting Arms Coordination, and Airspace Reconfiguration
Communication Method	POTS, Chat, E-mail, Radio	
Request Flow	Low, Medium, High	
COI 5: Does the system support the TAOC intercept of hostile missions? Response Variables: TAOC Intercept Time (IT_T) and Probability of Success (P_S)		
Communication Method	POTS, Chat, E-mail, Radio	Hostile Air/Missile Interception
Request Flow	Low, Medium, High	

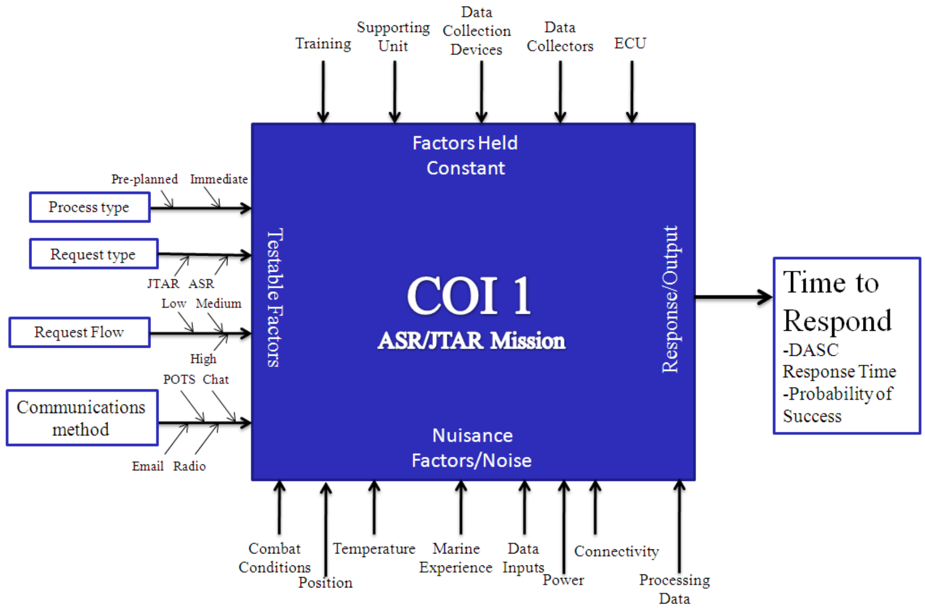
*See Appendix A for supporting data (process diagram, replication, and power/confidence level)

The Evaluation Framework in Table 3.2 captures the critical data that will be collected and recorded during DT, OA, and IOT&E.

Appendix D – Design of Experiment for COIs

COI 1: Does the system support the DASC ASR/JTAR missions within 8.5 minutes?
Response Variable 1: DASC Response Time (RT_D)
Response Variable 2: Probability of Success (P_S)

Process Diagram:
The process diagram illustrated below depicts all factors that could impact the system's ability to perform the DASC ASR/JTAR mission. Factors were divided among: testable factors, which will be used in the experimental design; held-constant factors, which will be set at a constant level for the full test; and noise factors which are uncontrollable and therefore cannot be directly controlled or held constant in the test. These factors will be recorded at the observed values in the data collection process.



Design:
A Full-Factorial design was determined to be the best design to fit the first-order model with interaction; therefore, 48 runs are appropriate for this mixed level design. The desired number of replicates is two, for a total sample size of 96. The power of the overall test is 1.00 when the confidence level is set to 0.80, predicted signal to noise ratio is 1, and total sample size is 96.

	Request Type →	JTAR			ASR			
	Process Type ↓	Low	Medium	High	Low	Medium	High	**← Request Flow**
Radio	Pre-planned	2	2	2	2	2	2	
	Immediate	2	2	2	2	2	2	
Chat	Pre-planned	2	2	2	2	2	2	
	Immediate	2	2	2	2	2	2	
POTS	Pre-planned	2	2	2	2	2	2	

	Request Type →	JTAR			ASR		
	Immediate	2	2	2	2	2	2
Email	Pre-planned	2	2	2	2	2	2
	Immediate	2	2	2	2	2	2

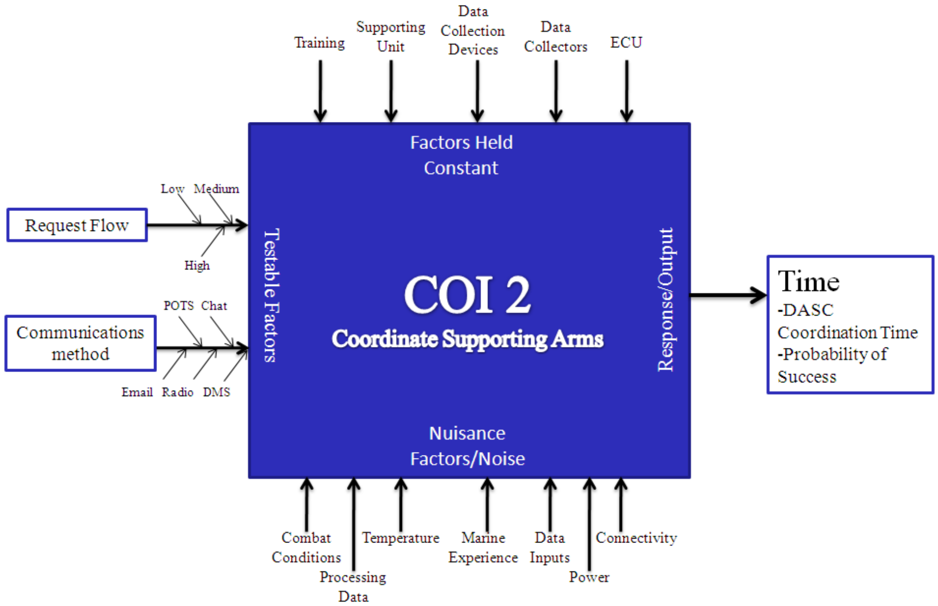
The power analysis for DASC Response Time, by term, for a signal/noise ratio of 0.75 and 1.00 is below. The IOT will provide an opportunity for a total sample size exceeding 128. No limitations are identified at this time.

Term	Power	
	0.75	1.00
Request Type	1.00	1.00
Process Type	1.00	1.00
Request Flow	1.00	1.00
Comm Method	1.00	1.00
Request Type & Process Type	1.00	1.00
Request Type & Request Flow	1.00	1.00
Process Type & Comm Method	1.00	1.00
Process Type & Request Flow	1.00	1.00
Request Flow & Comm Method	1.00	1.00

A power of 0.48 can be achieved for the probability of success if the desired effect size is 0.10 and total sample size is 96 with 2 replicates per test case.

COI 2: Does the system support the DASC Supporting Arms Coordination mission?
Response Variable 1: DASC Coordination Time (CT_D).
Response Variable 2: Probability of Success (P_S)

Process Diagram:



Design:
A Full-Factorial design was determined to be the best design to fit the first-order model with interaction; therefore, 30 runs are appropriate. The desired number of replicates is two, for a total sample size of 60. The power of the overall test is 1.00 when the confidence level is set to 0.80, predicted signal to noise ratio is 1, and total sample size is 60.

Comm. Method ↓	Low	Medium	High	← Request Flow
POTS	2	2	2	
Chat	2	2	2	
Email	2	2	2	
DMS	2	2	2	
Radio	2	2	2	

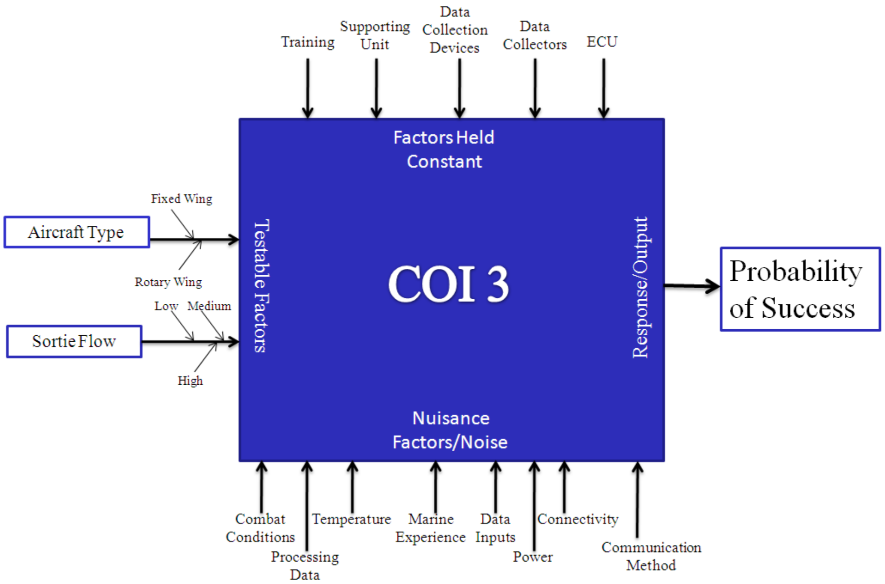
The power analysis table, by term, for signal/noise ratios of 0.75 and 1.00 is below. IOT will provide an opportunity for a total sample size exceeding 180. No limitations are identified at this time.

Term	**Power**	
	0.75	**1.00**
Request Flow	0.997	1
Comm Method	0.997	1
Request Flow & Comm Method	0.997	1

A power of 0.38 can be achieved for the probability of success if the desired effect size is 0.10 and total sample size is 60 with 2 replicates per test case.

COI 3: Does the system support the DASC Air Control mission with at least a 0.80 probability?
Response Variable 1: Probability of Success (P_S).

Process Diagram:



Design:
A Full-Factorial design was determined to be the best design to fit the first-order model with interaction; therefore, 6 runs are appropriate. A power of 0.66 can be achieved if the desired effect size for probability of success is 0.10 and total sample size is 180 with 30 replicates per test case. WTI provides the ability for 120 samples for this test. MCOTEA will increase sample size by conducting special test events.

Aircraft Type ↓	Low	Medium	High	← Request Flow
Fixed Wing	30	30	30	
Rotary Wing	30	30	30	

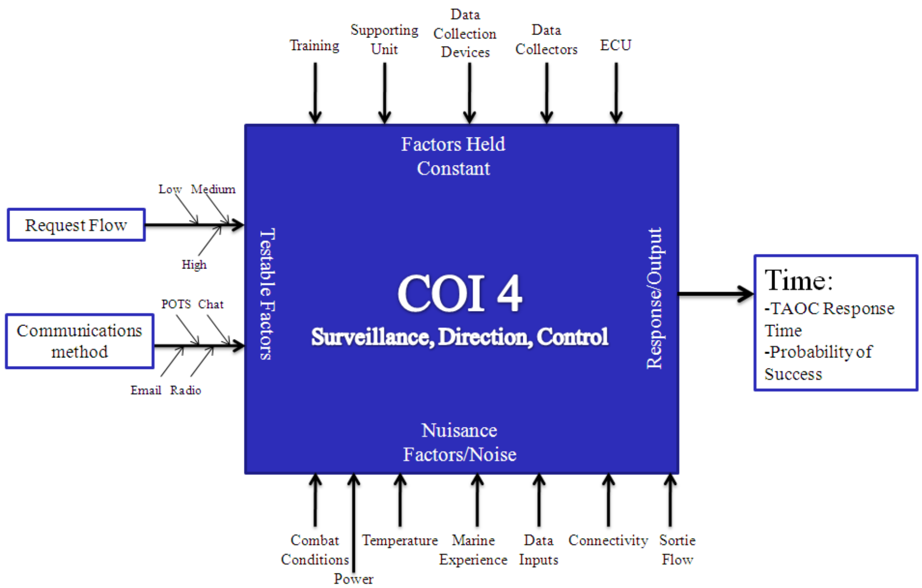
The table below shows how the effect change and number of replications affects the power of the test.

Effect Change	Replications 0	1	5	10	15	20	25	30
0.00	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
0.05	0.10	0.13	0.17	0.21	0.25	0.28	0.30	0.33
0.10	0.10	0.17	0.28	0.38	0.47	0.54	0.60	0.66
0.15	0.10	0.21	0.41	0.58	0.70	0.79	0.85	0.90

COI 4: Does the system support the TAOC Surveillance, Direction, and Control mission?

Response Variable 1: TAOC Response Time (RT_T)
Response Variable 2: Probability of Success (P_S)

Process Diagram:



Design:
A Full-Factorial design was determined to be the best design to fit the first-order model with interaction, therefore 24 runs are appropriate. The desired number of replicates is two, for a total sample size of 60. The power of the overall test is 1.00 when the confidence level is set to 0.80, predicted signal to noise ratio is 1, and total sample size is 48.

Comm. Method ↓	Low	Medium	High	**← Request Flow**
POTS	2	2	2	
Chat	2	2	2	
Email	2	2	2	
Radio	2	2	2	

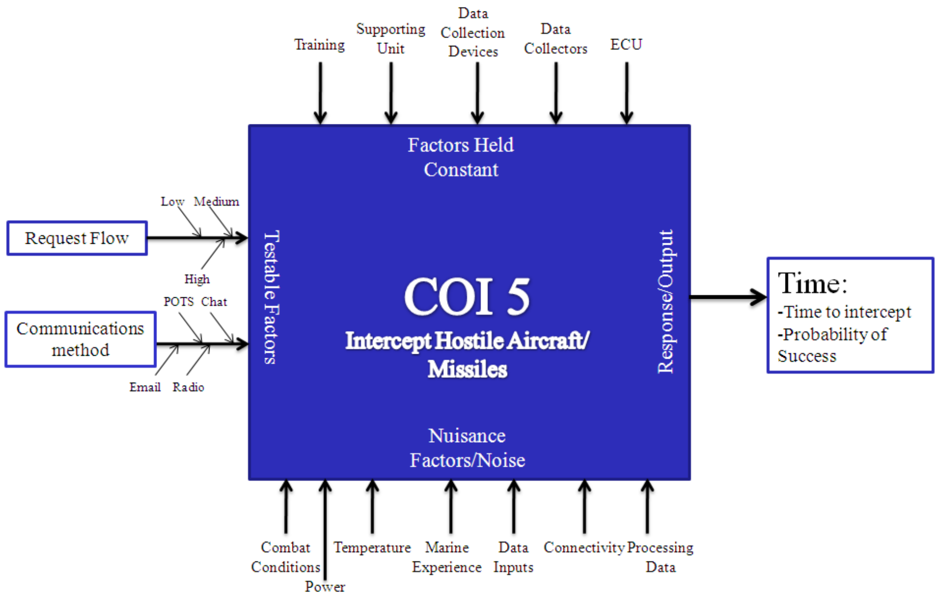
The power analysis table, by term, for a signal/noise ratio of 0.75 and 1.00 is below. The IOT will provide an opportunity for a total sample size exceeding 180. No limitations are identified at this time.

Term	**Power**	
	0.75	**1.00**
Request Flow	0.988	1
Comm Method	0.988	1
Request Flow & Comm Method	0.988	1

A power of 0.34 can be achieved for the probability of success if the desired effect change from a probability of success of 0.80 is 0.10 and total sample size is 60 (2 replications).

COI 5: Does the system support the TAOC Intercept of Hostile Aircraft/Missile mission?
Response Variable 1: Time to Intercept (TI_T)
Response Variable 2: Probability of Success (P_S)

Process Diagram:



Design:
A Full-Factorial design was determined to be the best design to fit the first-order model with interaction, therefore 24 runs are appropriate. The desired number of replicates is two, for a total sample size of 60. The power of the overall test is 1.00 when the confidence level is set to 0.80, predicted signal to noise ratio is 1, and total sample size is 48.

Comm. Method ↓	Low	Medium	High	**← Request Flow**
POTS	2	2	2	
Chat	2	2	2	
Email	2	2	2	
Radio	2	2	2	

The power analysis table by term for signal/noise ratios of 0.75 and 1.00 is below. The IOT will provide an opportunity for a total sample size exceeding 180. No limitations are identified at this time.

Term	**Power**	
	0.75	**1.00**
Request Flow	0.988	1
Comm Method	0.988	1

Term	Power	
	0.75	**1.00**
Request Flow& Comm Method	0.988	1

A power of 0.34 can be achieved for the probability of success if the desired effect size is 0.10 and total sample size is 60 with 2 replicates per test case.

The objective of the IOT shall be to evaluate the howitzer interoperability, rate of fire, fire mission accuracy, responsiveness and automotive performance as well as mobility and reliability in support of combat operations. The test results shall support a full rate production decision.

The IOT will follow the same DOE philosophy and have the same factors and levels as the LUT/OA except it will be larger. A split plot design will be created based on the same set of factors and levels. Similarly the factors will be controlled in the same manner with the missions starting out close moving to the C5 ranges and the returning to the initial range bands over the course of the three 96-hour scenarios. Due to the increased number of missions, number of rounds fired and length of the test in the IOT compared to the LUT/OA, more interactions can be estimated, to include main effects and second order interactions. IOT design will ensure a similar balance between statistical capabilities and operational coverage. Similar to the LUT/OA, the IOT will consist of a smaller subset of the total number of required missions compared to the DOE missions. The overall ratio of the DOE to the total number of missions will be the same or very similar. Thus all the non-lethal, emergency firings, out of sector missions, and additional C5 missions needed to meet the OMS/MP, which would again follow tactical moves, and additional C5 missions will be injected into the matrix at the discretion the Test Officer to ensure operational realism.

Red Team excursions will be conducted at the discretion of the IOT Test Officer. These excursions will support Information Assurance evaluation requirements in an operational environment at a system of systems level. Additional information relating to Red Team excursions can be found in paragraph 4.3.2.5 “IOT Events, Scope of Testing and Scenarios” of the TEMP.

Information Assurance – Examples

3.2 IA EVALUATION FRAMEWORK EXAMPLES

As part of the overall evaluation framework, this paragraph should include evaluation issues for information assurance. The information assurance issues should be scoped appropriately for the system under test and Table 3.1 should include appropriate measures for the information assurance issues and the nature of the system. The development, requirements, operational, and test community representatives should work together to identify appropriate issues and measures for the system.

Example Issue – How well do the system's information assurance capabilities protect the Commander's/user's required data/information?

Potential measures/metrics for this issue::

- Level of effort (e.g., time) required by the penetration team to achieve penetrations, accounting for system information made available
- Comparison of time to penetrate a system/network with the system mission duration, accounting for system information made available
- Number of attempts that failed to escalate privileges over the total number of attempts
- Adequacy of network scanning and patch management
- Adequacy of configuration management
- Effectiveness of firewall
- Effectiveness of access control list
- Impact of vulnerabilities and exploitations.

Example Issue – Will the system's information assurance detection measures support the ability of the commander/user to identify specific attacks?

Potential measures/metrics for this issue:

- Total number of events/incidents detected in the system under test (SUT)
- Time taken to analyze detected events/incidents in the SUT
- Elapsed time between when a penetration was made and when the network defenders detected the penetration in the SUT
- Number of successful detections over the total number of penetrations/exploitations

- Effectiveness of intrusion detection systems
- Adequacy of audit logging, including review and analysis.

Example Issue – Will the system facilitate the Commander's/user's ability to react to detected penetrations and exploitations?

Potential measures/metrics might include:

- Number of successful reactions over the total number of detected penetrations and exploitations attempted
- Time taken by systems/security administrators to react to each incident
- Courses of action to support system's mission operation/performance.

Example Issue – Will the system facilitate the Commander's/user's ability to restore data/information?

Potential measures/metrics might include:

- Elapsed time between when a penetration was made and when network defenders fully restored the system/network to a trusted state
- Time to restore the system's support of operations after initiating restoration plan
- Number of instances where data/information were successfully restored over the total number of instances where data/information needed to be restored
- Assessment of continuity of operations.

3.3 DEVELOPMENTAL TEST OBJECTIVES

Example

Information Assurance: As part of System Developmental Test, information assurance testing will be conducted to ensure compliance with DIACAP and DoDIIS certification programs. Consistent with the program's overall approach to operational testing of information assurance, the operational test agency will be furnished with completed test reports from each event for review and support of the system's Information Assurance COI.

Mission-Oriented Evaluation - Examples

3.3.1 Mission-Oriented Approach

Evaluation of the XYZ Anti-Submarine Warfare (ASW) system will be completed in realistic at-sea scenarios using a production-representative system. This testing will assess whether the system meets the performance thresholds in the CPD but will primarily focus on the operational effectiveness of the system. The test ship will be tasked to conduct ASW as well as intelligence, surveillance, and reconnaissance (ISR) tactical missions. The ASW test platform will be directed to clear an area with a suspected hostile submarine; the test ship will search for, detect, report, and initiate engagement of hostile submarines up to, but not including launch of live ordnance. The test ship will also be tasked to conduct an ISR mission in a high-density surface contact environment. In both cases, the tasking will provide an element of surprise or uncertainty for the test ship; the test platform commander will be able to respond to the tactical situation as perceived when employing the XYZ system. Successful accomplishment of testing events will support an evaluation of system operational effectiveness, operational suitability, and a recommendation on fleet release of the system.

Mission-Oriented Evaluation – Guidance

While the test and evaluation strategy should provide opportunities to determine whether a system meets documented requirements, the ultimate purpose of the test and evaluation strategy is to demonstrate the operational effectiveness, suitability, and survivability of the system in an operational environment. Operational effectiveness is defined as the overall ability of the system to support successful mission accomplishment, when used by representative operators in the intended environment. This definition takes into account the interplay of the system under test and interrelated or supporting systems. In many cases, the system performance specifications in the requirements document will assist in the assessment of mission accomplishment, but the overall evaluation will not be limited to these specifications.

Often, system requirements are best demonstrated in a controlled developmental test that might exclude or control important elements of the expected operational environment. Still, there are some developmental test events that can be conducted with a mission focus using representative users in the intended operational environment. To assist in early identification of system problems that might only be manifest in operational environments, developmental test planners should incorporate elements of the operational environment (representative users, weather, threat systems, end-to-end missions, weapons, secure communications gear, user maintainers, etc.) into developmental testing whenever possible.

References

Reporting of Operational Test and Evaluation Results, DOT&E, January 6, 2010

Examples

Paragraph 3.3.1 Examples

VV&A

Under DoDI 5000.61, M&S must undergo verification, validation, and accreditation (VV&A). M&S that has previously undergone VV&A for other purposes still must undergo VV&A for the new intended use. However, previous VV&A generally simplifies the process because the previous efforts have been documented and the new VV&A effort typically can focus on the changes.

Verification determines whether the M&S accurately represents the developer's specifications. The M&S is expected to add two numbers; does it add two numbers? Validation determines whether the model is an accurate representation of the real world. The M&S is expected to add two numbers; does it provide the correct sum? Accreditation is the official certification that the M&S and its associated data are acceptable for an intended use.

For accreditation, the intended use is important because M&S that is useful in one area may not be useful in another because of known limitations. The accreditation will explicitly state the intended use, such as: "The Big Weapon Model will be used to estimate the miss distance between the weapon and the target in support of developmental test DT-II." It also should acknowledge any significant limitations: "The Big Weapon Model does not include threat countermeasures, and consequently all scenarios are simulated in a clear environment."

The scope of the VV&A effort is a function of how each M&S will be used. On one extreme are simple models used early in a program (e.g., a spreadsheet model used to estimate system performance) that demand little effort for VV&A, and frequently rely on pre-existing documentation and prior VV&A efforts. At the other extreme are high-fidelity models an evaluator might use to assess a Key Performance Parameter or to resolve a Critical Operational Issue (e.g., a hardware-in-the-loop missile model used to estimate performance against countermeasures); these must undergo a rigorous VV&A effort. In general, the more important the M&S is to the final evaluation, the more rigorous the VV&A must be.

References

DoDI 5000.59

DoDI 5000.61

Defense Acquisition Guidebook

DoD Instruction 5000.02

M&S for Test Planning, Pretest Predictions, and Evaluation - Guidance

The Modeling and Simulation (M&S) sections should address how M&S will be employed in the overall test strategy and how the M&S will be verified, validated and accredited (VV&A). Specifically, the TEMP should list any M&S expected to be used, the intended use, any data requirements, the test objectives to be addressed and/or how test scenarios will be supplemented with M&S, the planned VV&A effort, and who will conduct the VV&A effort (ref. DoDI 5000.61). The TEMP should list any specific test events required for VV&A of the M&S. The resources for the specific test events will be included in Part IV.

M&S is used throughout developmental, operational and live fire testing. M&S can support evaluation of requirements, trade studies, test planning, pre-test predictions, evaluation of system performance, and other uses. It includes a broad set of tools including spreadsheet models, high-fidelity digital models, and hardware and computer in the loop facilities.

M&S for Test Planning and Prediction

M&S is frequently used in test planning and pre-test predictions. If M&S will be used as a basis for decisions regarding test scope or test conditions, describe how it will be used. Which tests will have pre-test predictions? Which pre-test predictions will be based on M&S?

M&S for Evaluation

Another common intended use is that M&S can support the evaluation of system performance. The TEMP should describe how M&S will be used as a basis for evaluating any important system metrics and/or make significant contributions to the assessment of any Critical Operational Issues. Will M&S be used to supplement, interpolate, or extrapolate results from live testing? Briefly summarize the degree of reliance on M&S. Which evaluation issues will be addressed through M&S? Explain why M&S is being used (e.g., safety restrictions preclude testing).

Test Limitations – Guidance

Guidance

Ideally, the test and evaluation strategy would have no limitations that could degrade or prevent resolution of the critical operational issues (COIs) or formulation of conclusions concerning system effectiveness, suitability, or survivability. In those instances when test limitations cannot be avoided, the TEMP should enumerate them. For each limitation, the TEMP should explain the problem(s) in enough detail to describe specifically how the limitation will affect the evaluation and the conclusions that can be drawn from the test.

A program might have test limitations that affect DT, LFT&E, and/or OT. Each limitation should be addressed in TEMP sections 3.3.3 DT Test Limitations, 3.4.3 LFT&E Test Limitations, or 3.6.3 OT Test Limitations, as appropriate.

Rarely should a TEMP that anticipates a critical limitation for planned test events be submitted to DOT&E for approval. The TEMP should explain plans, if any, to mitigate limitations.

Definition

Generally, test limitations are constraints that cause differences between the test environment and the expected operational environment (combat or peacetime, as appropriate), which in turn could cause the test results to differ from the results in the expected operational environment. A test might also have limitations if it is impossible to establish ground truth or evaluate results with certainty. The test might be limited in scope because there are inadequate resources to test in all of the relevant operational environments, e.g., extreme cold or hot weather. Other limitations might include altered procedures because of safety concerns, constrained test infrastructure, lack of threat surrogates, inadequate target realism, or the immaturity of the system or any subsystems.

References

Defense Acquisition Guidebook, sections 3.4.3, 3.6.3, 9.6.1, and 9.6.2

Test Limitations – DT Example

3.3.3 Test Limitations

Aerial targets will not fully represent the full spectrum of threat anti-ship cruise missiles (ASCM) in terms of speed, altitude profile, maneuverability, radar cross section, size and shape, infra-red (IR) signature, countermeasures, counter-countermeasures, radar emissions, and survivability (in the event of warhead-configured Sea Sharks). In those areas where the target fidelity differs substantively from the most prevalent ASCM threat, the Sea Shark and its supporting NCS may not be stressed to a comparable extent as they would be by the actual threat, thereby bringing into question the relevance of the operational test results when using the lower fidelity target. The areas in question are the target speed and the target altitude profile.

Planned mitigation efforts include:

- NCS and Sea Shark modeling and simulation will explore Sea Shark missile performance and in-flight support against all expected threat/target speed/altitude profiles. This will be followed by validation of the M&S simulation with developmental test results and pre-shot predictions for operational testing.
- Development and procurement of an upgraded threat target that can match the speed/altitude profile of the most challenging threats.

Background for Maritime Air Defense Example

This example is for the hypothetical Sea Shark missile (ship-launched, anti-air, semi-active radar homing missile, supported by the hypothetical Neptune Combat System (NCS)). Critical operational issues (COIs) for Sea Shark and its supporting combat systems include:

- Area Air Defense Capability (Can Sea Shark, supported by the NCS, provide air defense for other ships within the Aircraft Carrier Strike Group?)
- Own Ship Air Defense Capability (Can Sea Shark, supported by the NCS, provide own ship defense against air threats while also conducting Area Defense?)
- Availability (Can Sea Shark, after a representative shipboard storage time in the vertical launch cell, provide the required launch availability?)
- Reliability (Can Sea Shark, after a representative shipboard storage time in the vertical launch cell, provide the required in-flight reliability?)

M&S for DT and OT - Examples

Aircraft OT&E Example

The F-100 fighter aircraft will use the Aerial Combat Simulation (ACS) to support evaluations of F-100 operational effectiveness in air-to-air missions. The ACS will provide data in support of the following metrics: Air-to-Air Kill Ratio, Blue-on-Blue Kills, and Blue-on-White Kills. Other secondary metrics also will be evaluated.

The ACS consists of four actual F-100 cockpits installed in visual scene domes and ten other manned interactive cockpit stations. The ACS includes high fidelity models of the F-100's cockpit and sensor suite and integrated threat models developed by MSIC, NASIC, and ONI. Scenarios will be focused around two simultaneous Major Contingency Operations threats. The ACS is intended to model a dense surface-to-air and air-to-air threat and electronic signal environment, which is impractical to create on an open-air range (OAR).

The ACS will support operational test design, test team and pilot training, and test preparation and rehearsal. In addition, ACS will be used to mitigate test limitations and to support the evaluation of F-100 effectiveness under conditions not possible on an OAR. OAR limitations that ACS can address include constraints due to flight security concerns, the lack of realistic threat assets (types and/or numbers), and limited battle space.

AFOTEC will perform Verification, Validation, and Accreditation (VV&A) of the ACS, which will include the use of F-100 DT validation data, Intelligence agency support of validated threat models, and operational test data collected on the OAR against available threats or surrogates. A model-test-model approach will be used. If intelligence shortfalls limit the ability of AFOTEC to accredit an ACS component, AFOTEC will consider the operational context of the shortfall to assess the likely outcome and impact to the evaluation. ACS limitations will be included in the F-100 IOT&E test plan. AFOTEC has defined the ACS requirements to support the F-100 IOT&E via the Integrated Test Team (ITT).

Funding and resources for ACS validation, ACS operation and AFOTEC test activities in the ACS for FY-10 through FY-15 are detailed in Part IV.

not included in the HWIL facility will be simulated by the IFS model. The second AMSTAR facility to be used will be the Production Test Bay, still under development, and will incorporate every hardware and software component of tactical missiles.

Production Hardware in the Loop

The Production Test Bay will be used primarily as a safe, non-destructive production acceptance test capability with the objective of cost savings from performing less destructive testing of production missiles. The Production Test Bay will use IFS models to stimulate the missiles under test. Both the Performance Test Bay and Production Test Bay are a combined development effort of the AMRDEC and the Redstone Technical Test Center (RTTC), a subordinate command of the Army Test and Evaluation Command (ATEC) that was the primary financial sponsor during development. The Production HWIL will support AUR testing in a non-destructive environment prior to GFT. The Production HWIL will be on line prior to the end of SDD and utilization will continue during the production phase of the program. The Production HWIL will use IFS drivers to stimulate the tactical hardware and will use equivalent scene generators to those developed for the Performance HWIL. VV&A of the Production HWIL will be completed prior to FRP.

Simulation Based Performance Assessment

The simulation based performance assessment (PA) will address the BAMM key performance parameters; probability of hit, probability of kill, and probability of incapacitation. While the flight test program will demonstrate a limited number of scenarios, the simulation will be used to assess the performance for a broad range of scenarios under a broad range of conditions. This approach will not only assess performance for the broad range of scenarios but also BAMM performance robustness to various conditions within those scenarios. The PA will use the IFS all digital capability, with subsets being conducted using the IFS in the STS and the performance HWIL. Various levels of preliminary assessments will be conducted throughout SDD. The results of these initial assessments will be provided to the prime contractor to support design and algorithm enhancements. The milestone C PA, which will calculate the probability of hit and probability of kill against the BAMM-specified targets, will occur during the latter portion of SDD, after the system design is solidified and after the simulation has been validated against flight tests. The PA will consist of a large number of simulation executions for the different launch platforms, all modes of operation, stationary and moving targets, and target aspect. The BAMM Simulation IPT will

Missile DT and OT Example

Modeling and Simulation (M&S) is an integral part of Bama Missile (BAMM) T&E. Below is a discussion of the BAMM simulation and associated tools.

Integrated Flight Simulation (IFS)

The BAMM IFS is a complete, closed-loop simulation of the BAMM system and is considered the authoritative representation of the BAMM for simulation purposes. The BAMM IFS contains five main models: (1) environment model, (2) seeker model, (3) tactical software including the missile tracker, (4) six degrees of freedom (6-DOF), and (5) launcher model. The five main models contained in the BAMM IFS are independent of any contractor's technical solution and any simulation architecture. The BAMM IFS is a contract deliverable to the Government by the prime contractor and will be hosted by the government at the Army's Aviation and Missile Research, Development and Engineering Center at Redstone Arsenal and the Navy's Naval Air Warfare Center Weapons Division at China Lake. Independent Verification and Validation will be conducted by the government under the auspices of the BAMM Simulation Working Group.

Software Test Station (STS)

The BAMM STS contains tactical processor boards which replace the equivalent models contained in the IFS, along with the tactical software. The other models of the IFS remain the same. The STS is used to perform further checkout of missile tracker algorithms and tactical software, but its primary function is to perform the Formal Qualification Testing (FQT) of the tactical software prior to loading on tactical hardware for guided flight testing.

Performance Hardware in the Loop

Throughout the SDD acquisition phase, the prime contractor will be required to provide to the Government missile hardware and support to allow the government simulation team to complete development of the Advanced Multispectral Simulation, Test, Acceptance Resource (AMSTAR), consisting of two hardware-in-the-loop (HWIL) facilities located at Redstone Arsenal.

The first AMSTAR facility to be used will be the Performance Test Bay, which will be used by the government and prime contractor as a risk reduction tool for missile seekers by performing system and subsystem tests, and performing pre-flight test predictions and post-flight test reconstructions and analysis. Those missile components develop the exact structure of the PA. The PA will be conducted for benign atmospheric conditions, selected countermeasures, APS/DAS, obscurants, and different weather conditions. The magnitude and structure of the countermeasures, APS/DAS, obscurant, and weather matrices will also be defined during the SDD contract.

The PA will include a Monte Carlo analysis of the missile seeker parameters, 6-DOF variables, different geographic locations, and different target locations within a geographic location. Target conditions will include moving and stationary, solar loaded, and non-solar loaded. Geographical locations will include temperate, arid, and cold weather areas.

Verification, Validation, and Accreditation

The most important activities to be performed in M&S on BAMM are Verification, Validation, and Accreditation (VV&A). As such, the VV&A strategy will be aggressive and rigorous for the prime contractor as well as for the Government. The BAMM System Simulation Working Group (SWG) will be the overseeing organization for VV&A. A VV&A subgroup will be formed within the SWG and will be required to report regularly to the SWG and will document their efforts to the T&E Integrated Product Team (IPT). The VV&A subgroup will contain members from the JAMS PO, the prime contractor, AMRDEC and NAWC subject matter experts (SMEs), ATEC, OPTEVFOR, and other interested organizations.

SMEs from the Army, Navy, and the prime contractor will be used in the model verification effort. To assist the SMEs in their effort, the Common Simulation Evaluator (CSE) will be used and tailored for the particular model being verified. This provides a method of quantifying and documenting the models. The compilation of the CSEs for the models will constitute a major portion of the verification documentation contained in the BAMM System Verification Report. This report will be augmented by the prime contractor's contractually required deliverable "IFS Model and System Level V&V Report," which will include test data from various tests conducted. The initial delivery of the prime contractor's report is due at the Preliminary Design Review. The next required update will be at the Critical Design Review with additional updates as required.

Validation of the IFS will be a multi-faceted approach. Validation will be accomplished based upon component level tests as well as vendor test data. The test data will be compared to the applicable IFS model. The validation of the component model will be made by the SMEs, presented to the VV&A subgroup of the SWG, and presented to the T&E IPT

The accreditation of the IFS for the BAMM System will be a joint accreditation by the Army and the Navy evaluation and development communities. The accreditation approach will be for the VV&A subgroup to develop the IFS Accreditation Plan, then present the plan through the SWG to the T&E IPT for concurrence. The VV&A subgroup will also develop the Accreditation Support Package and the Accreditation Report. It is currently intended for the IFS accreditation methodologies to be tailored from existing Army and Navy accreditation methodologies.

The IFS system level validation will be based upon a Model-Test-Model approach. The prime contractor, as well as the Government, will perform pre-flight predictions using the IFS of the scenario to be used in an upcoming flight test. The scenario will include the test range to be used, range from missile at trigger pull to the target, target aspect angle relative to the missile at trigger pull, and target motion at trigger pull. During the flight tests, telemetry data will be collected on the missile, either with the mini-telemetry section that is a part of the missile or with the warhead replacement telemetry that will only be on pre-determined missiles. Other data to be gathered include range and target metrology data, and the infrared target signature measurements that will be collected pre-flight test and post-flight test as allowed by range control/safety. The data gathered for the flight test is then used in the post-flight reconstruction in the IFS. Key missile parameters are analyzed for the flight test and for IFS Monte-Carlo runs. The comparison of the flight test results and the IFS results will show the validity of the IFS. The VV&A subgroup will oversee this effort and present results to the SWG and the T&E IPT as required.

IOT&E Scenarios

The actual test scenarios executed during the IOT will be prepared to maximize the operational realism of the test. These scenarios will be generated using the AH-64D and AH-1Z Concept of Operations (CONOPS) and TTPs and be centered on the execution of their assigned missions.

AH-1Z scenarios will include Close Air Support (CAS), Deep Air Support (DAS), armed and visual reconnaissance, Forward Air Control Airborne (FACA), escort, and interdiction/ emergency defense of the expeditionary strike group. Forward Arming and Refueling Point (FARP) and CBRN operations will be conducted as needed in support of these scenarios.

AH-64D scenarios will include both short and maximum range engagements normally associated with Close Combat with Ground Forces, Interdiction Attack, and

Vertical Maneuver missions. A/C acquisition sources matched with BAMM multiple seeker-mode capabilities will be used to test BAMM integrated seeker-mode performance based on established TTPs. The engagements will include moving and stationary targets and targets within MOUT-type environments. FARP and CBRN operations will be conducted as needed in support of these scenarios. Six AH-64D A/C will be required to support operational testing, four with FCR and two without the FCR. Engagements will be fired using the desert type terrain at China Lake/YPG.

As a minimum, the target list will include Tanks, Air Defense Artillery (ADA) weapons, MOUT targets, Armored Vehicles, maritime targets, and both stationary and moving targets. The test will be conducted in the natural environment of the operational test range. The test officer will collect measurements of temperature, pressure, humidity, precipitation, clouds, winds, blowing sand, or other conditions that may influence system performance. BAMM capabilities and limitations in various SAL/EO/IR/RF CM environments will be assessed to determine effects on operational performance and possible BAMM tactics and improvements. Acquisition denial and tracking interference susceptibility testing will be conducted in both captive-carry and live-fire missions/scenarios against known battlefield obscurants, such as APS/DAS, host platform expendable CM, support jamming operations, and any additional CM determined to affect operations of the BAMM as specified in the STAR and Threat TSP.

Data will be captured on target acquisition performance, engagement/download timelines, missile diagnostic checks, human factors feedback, onboard A/C video, and other measures. ***To the degree possible, engagements/missions will be flown in simulation prior to the test to verify that each meets test performance requirements in terms of launch conditions, flight profiles, and target conditions.***

Collected data will include measurements of missile-hit performance, target acquisition and transfer performance, engagement timelines, flight profiles, reliability, and other measures. Questionnaire information will also be collected from pilots on A/C/missile interface performance and from support personnel on support issues. Data on suitability and survivability will be collected where possible during the test.

Integrated Survivability Evaluation – Guidance

Summary

The Developmental Test and Evaluation (DT&E), OT&E, and Live Fire Test and Evaluation (LFT&E) strategies should be integrated so that the full spectrum of system survivability is assessed in a consistent manner. For some systems, it might be appropriate for Critical Operational Issues (COIs) to address system and/or personnel survivability. Personnel survivability (force protection) must be addressed for systems under LFT&E oversight and should be integrated into the overall system evaluation of survivability.

Best Practices

The evaluation of survivability for many combat systems can be subdivided into assessment of susceptibility (probability of hit), vulnerability (probability of kill given a hit), and force protection (measures or features to protect occupants), as shown in Figure 1.

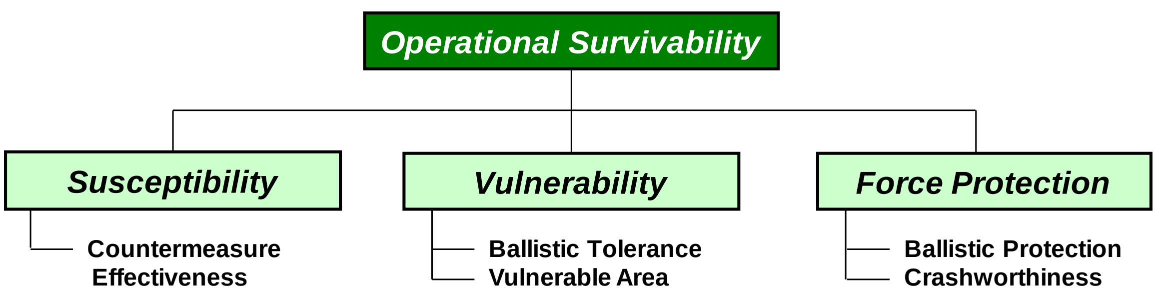


Figure 1. Operational Survivability

An integrated survivability test strategy might envision several operational scenarios or mission threads that guide the design of developmental testing of countermeasure systems, live fire testing of ballistic tolerance, vulnerable area analyses, and force protection assessments. Operational testing might, for example, generate the most likely threat engagement scenarios (shot lines) that are subsequently investigated in LFT&E against system components. The vulnerability assessment might provide the probabilities of kill given a hit for use in real-time casualty assessment instrumentation during operational testing. Other LFT&E insights available from DT&E and OT&E testing of susceptibility might include information on signatures, employment of countermeasures, and tactics used for evasion of threat weapons.

References

10 USC 2366

Defense Acquisition Guidebook

Force Protection and Personnel Casualties - Guidance

Summary

Force Protection attributes are those that contribute to protection of personnel. In particular, they are closely linked to the issue of personnel survivability. For programs on oversight for survivability Live Fire Test and Evaluation (LFT&E), the critical LFT&E issues must include personnel survivability. In general, personnel survivability should be addressed through dedicated measures of evaluation, such as "expected casualties." The ability of personnel to survive should be addressed even in cases where the platform cannot survive.

Key Performance Parameters (KPPs) for force protection and survivability are required for any manned system that is expected to be deployed in an asymmetric threat environment. Although force protection is a primary issue for programs on LFT&E oversight, force protection as an evaluation issue is not limited to such programs.

All Department of Defense (DoD) hard body armor acquisition programs under DOT&E oversight will execute, at a minimum, a DOT&E-approved protocol for testing that results in a decision to qualify a design for full-rate production (i.e., First Article Testing).

References

10 USC 2366

Policy for Updating Capabilities Documents to Incorporate Force Protection and Survivability Key Performance Parameters, The Joint Staff, 13 June 2005

Defense Acquisition Guidebook

Standardization of Hard Body Armor Testing, DOT&E, 27 April 2010

Test Planning Documents – Guidance

Summary

For all operational tests, live fire tests, and all other tests that support DOT&E evaluations, the TEMP should include a matrix that identifies which test planning documents will be submitted for DOT&E approval and which will be submitted for information and review only. The lead OTA shall brief the DOT&E on T&E concepts for the Operational Test Plan as early as possible but no fewer than 180 days prior to start of any such testing. The lead OTA shall deliver the Operational Test Plan to DOT&E for approval no fewer than 60 days before the start. Use of developmental test data for an operational assessment or evaluation should be coordinated with the lead OTA and DOT&E prior to the start of testing, and, when feasible, shall receive prior approval. The DOT&E shall require approval of LFT&E strategies, LFT&E plans, and survivability test plans for covered systems.

References

Defense Acquisition Guidebook

Title 10 USC 2399

Timeliness of Operational Test and Evaluation (OT&E) Plans, DOT&E, 24 June 2011

Example Document Approval Matrix

Test Plan Approval Matrix - Example

Table 19 – Document Review and Approval Matrix

Test	Document	Delivery Date	DT&E	DOT&E
LFT&E				
	Armor Coupon Detailed Test Plan	30 days before test		X
	BH&T OTA TP	60 days before test		XX
	BH&T Detailed Test Plan	30 days before test		X
	Controlled Damage Experiment Detailed Test Plan	30 days before test		XX
	FUSL Pre-Test Predictions	15 days before test		X
	FUSL OTA TP	60 days before test		XX
	FUSL Detailed Test Plan	30 days before test		XX
	M&S Accreditation Report including V&V Report(s)	Before start of FUSL test		X
	M&S Comparison Report	90 days after final FUSL test event		X
Developmental Testing				
	Component Qualification Test Plans	60 days before each test	X	X
	Weapons Performance Test Plan	60 days before test	X	X
	Sensor Performance Test Plan	60 days before test	X	X
Operational Testing				
	Operational Assessment Test Plan	60 days before test	X	XX
	IOT&E Test Plan	60 days before test	X	XX
	FOT&E Test Plan	60 days before test	X	XX

X – Denotes Review
XX – Denotes Review and Approval

BH&T	Ballistic Hull and Turret
OTA TP	Operational Test Agency Test Plan
FUSL	Full-up system-level
M&S	Modeling and Simulation
V&V	Verification and Validation

M&S for LFT&E - Examples

Ship LFT&E Example

M&S for Test Planning and Prediction. For the tests of surrogate ships, the Internal Blast (INBLAST) model and the Blast Damage Assessment Model (BDAM) will be used for pretest predictions of blast pressure loading and ship structural response to the loading, and SVM will be used for fragment penetration predictions. The Consolidated Model of Fire Growth and Smoke Transport (CFAST) will be used for fire growth curve development in the post-shot analyses for the CG 19 testing, and for the pretest fire spread predictions and post-test data analyses for the ex-*Maui* test. The Advanced Survivability Assessment Program (ASAP) will be used for primary damage pretest predictions and post-test analyses for the DD 930 test. ASAP, BDAM, CFAST, and the Fire and Smoke Simulator (FSSIM) will used for the ex-*Larson* Autonomic Fire Suppression System (AFSS) Weapons Effects Test (WET).

Reliance on M&S for Evaluation. M&S is a primary method of executing the alternative LFT&E program. The Shock Trial, TSST, component shock testing, surrogate testing, combat incidents, and peacetime accidents supplement the M&S and serve in part to validate the modeling that is performed. Realistic tests of surrogates will address the most significant areas of uncertainty, e.g., fire spread and the ability to extrapolate shock trial results to realistic encounter conditions for proximity underwater bursts. One of the primary objectives of both the Advanced Threat Weapons Effects tests is to obtain data that could be used to improve or validate damage algorithms used in ship vulnerability models.

Susceptibility analyses will be performed to determine likely hit points for the threats to be assessed in the Final Vulnerability Assessment Report. The M&S tools that will be used to generate hit points included CRUISE_MISSILES, Total Mine Simulation System (TMSS), and the Technology Requirements Model (TRM).

A full ship DYSMAS finite element model is being used to predict the structural damage and equipment shock environments with greater fidelity. Deactivation diagrams for the prediction of secondary damage will replace the Integrated Recovery Module (IRM). Since deactivation diagrams do not enable the generation of recoverability time lines, recoverability will be addressed through other means.

The program office VV&A process relies heavily on data from legacy models, and will use test data to assist in the validation of new model functionality. ASAP was accredited with limitations for the Initial Vulnerability Assessment Report. The Program Manager is funding a project to improve the fidelity of blast projections in the ASAP model.

Aircraft LFT&E Example

M&S for Test Planning and Prediction. Susceptibility and vulnerability issues will be examined with modeling and simulation. M&S will be used to scope the ballistic series of tests and the specific tests within each series. Pre-test predictions are being made for all tests, with the intent of using test results to identify M&S improvements.

A Modular UNIX-based Vulnerability Estimation Suite (MUVES-S2) vulnerability assessment model will be employed to support the overall aircraft vulnerability assessment. It will be used to select shotlines for testing and to generate pre-shot predictions.

Reliance on M&S for Evaluation. System-level survivability will be assessed using the aircraft signatures and known threat weapon system accuracies to evaluate the susceptibility and the vulnerability analysis results. Aircraft signatures will be measured in flight testing and used in models to predict countermeasure effectiveness. Infrared signatures will be used in Hardware-in-the-Loop (HITL) simulations to determine realistic impact locations on the aircraft for man-portable air defense system (MANPADS) threats and to evaluate the ability of aircraft survivability equipment to detect and counter MANPADS threats. The vulnerability analysis will use a 26-view average to determine vulnerable area and probability of kill given a hit for fragments and non-bursting projectiles.

A hierarchy of M&S will be used to analyze aircraft survivability and effectiveness. Engineering-level analyses will be used to assess vulnerability aspects such as structural response to hydrodynamic ram, fire and explosion, and vulnerable area. Higher level M&S will be used to assess one-on-one encounters, mission effectiveness, and force effectiveness. The models include:

- FPM – Fire Prediction Model
- ARAM – Advanced Ram Model
- FASTGEN – target description and Fast Shotline Generator model

• COVART – Computation of Vulnerable Area Tool model

• SHAZAM – missile warhead endgame model

• ESAMS – Enhanced Surface-to-Air Missile Simulation

• Brawler – air-to-air combat model

• JIMM – Joint Interim Mission Model

• Thunder – Force effectiveness model.

Since model improvements are always being made, model versions are not listed.

IOT&E Entrance Criteria – Guidance

Guidance

The purpose of IOT&E Entrance Criteria is to ensure that the system under test is ready to commence IOT&E and the required resources are in place to support the test. The intent of this requirement is to ensure that systems do not enter IOT&E before they are sufficiently mature. Premature commencement of IOT&E can waste scarce resources if IOT&E is suspended or terminated early because of technical problems that should have been resolved prior to the start of IOT&E. Commencement without all required resources can result in an inadequate test.

Best Practices

A determination that IOT&E is ready to proceed should be based on the following criteria:

- The system has demonstrated acceptable hardware and software performance during mission-focused DT conducted in operationally realistic environments with the hardware and software to be used in IOT&E.
- IOT&E test articles are production representative (as determined using DOT&E criteria).
- Threat surrogates and targets have been validated and approved by the DOT&E.
- All critical issues identified in the Assessment of Operational Test Readiness (AOTR) have been resolved.
- The required test ranges are ready to support all planned events as described in the IOT&E plan, including environmental, safety, and occupational health requirements.
- All required certifications and accreditations are in place, and DASD (DT&E) and DOT&E have been provided all data, including a description of the level of operational realism under which testing was conducted.
- Adequate reliability data are available (or planned) to enable prediction with statistical rigor of reliability growth and expected IOT&E reliability results.
- The staffing of the system is consistent with Concept of Operations and training has been completed consistent with that planned for intended users.
- Pre-IOT&E M&S predictions are based on verified, validated, and accredited modeling and simulation.

- DOT&E has approved plans to use DT data to support the evaluation. The required data have been provided to the OTA and DOT&E.
- The logistics system and maintenance manuals intended for use with the fielded system are in place for IOT&E.
- If operational force support is required for IOT&E, there is a documented agreement between the operating forces and the Component Acquisition Executive (CAE) describing respective roles and responsibilities during the test.
- DOT&E has approved the IOT&E plan.

References

Defense Acquisition Guidebook

DoDI 5000.02

Test Limitations – LFT&E Example

3.4.3 Test Limitations

LFT&E will not confirm or demonstrate through ballistic testing the actual vulnerability of the wiring or avionics subsystems of the Dakota aircraft. LFT&E and combat data have shown that ballistic damage to wiring or avionics can result in loss of mission critical systems such as: EO/IR sights/displays, communications, and weapons systems. In mitigation, the effects of avionics and wiring failures will be tested through fault insertion in the Avionics Integration Laboratory. Those results will then be incorporated into the system-wide M&S vulnerability assessment.

IOT&E Entrance Criteria – Examples

3.5 Certification for IOT&E The Component Acquisition Executive (CAE) will evaluate and determine system readiness for Initial Operational Test and Evaluation (IOT&E). Prior to the CAE's determination of readiness for IOT&E, an independent Assessment of Operational Test Readiness will be conducted by OUSD (AT&L). It shall consider the risks associated with the system's ability to meet operational suitability and effectiveness goals and will be based on capabilities demonstrated during DT&E and OAs, as well as on the criteria described in this TEMP. The final report for DT will provide insight into the system's readiness for IOT&E.

3.5.1 DT&E Information Required Adequate test data will be collected during DT-IIG and DT-IIH to allow the Program Manager to assess and report the system's capabilities against the stated COIs using the MOE/MOS listed in this TEMP prior to IOT&E.

3.5.2 IOT&E Entry Criteria

- All Milestone C exit criteria have been met.
- Department of the Navy Criteria for Certification listed in Secretary of the Navy Instruction 5000.02 of December 8, 2008 have been satisfied and the system is certified for test.
- All deficiencies identified in previous testing have been resolved.
- All required targets have been accredited and the test range has been adequately surveyed.
- Production representative test articles are available to conduct IOT&E.
- Red Team for information assurance penetration testing has been identified and is funded for testing.
- OTRR is completed and DOT&E concurs with proceeding to test.

End-to-End Testing – Guidance

Guidance

End-to-end testing is the logical means to conduct a mission-based evaluation. End-to-end testing is easiest thought of as testing a mission thread. Mission threads result from a careful analysis of a unit's mission using the system and can be derived from the Joint Mission Essential Task List or from the Component-specific Mission Essential Task List. The threads should make operational sense and evaluate the intended operational mission from beginning to end. The end-to-end evaluation of each mission thread should rely on testing that includes the entire thread in a single operational event. For example, a rocket or missile end-to-end test would include acquiring the target, passing the target information to a launch platform, firing the rocket or missile, hitting the target, and achieving the intended level of damage.

End-to-end testing is not just interoperability testing, which is to say that it is simply not enough to verify that critical information can pass throughout the mission thread. The end-to-end evaluation must assess the quality of the information and whether it results in a successful mission. For example, the evaluation of a munition should address the ability of targeting systems to provide accurate and timely targeting suitable for the munition and its intended target. The evaluation of a sensor platform should address the ability to provide the data products to the end user in order to complete the mission successfully. The evaluation of a ship or aircraft should include the performance of all onboard and other supporting systems required for mission completion.

If it is not possible (due to cost or safety issues) to include all aspects of a mission in a single operational end-to-end test, separate portions of the mission threads can be included in multiple test events. Each of these events should include some overlap, so that the start of test B includes the end of Test A. Conditions affecting mission performance should be duplicated in overlapping events as much as possible. Each test of the thread parts should be operationally representative and all should represent similar operational environments and threats. If separate test events are used, the TEMP should explain why it is not possible to conduct the end-to-end mission in a single event; this is a test limitation, and the TEMP should discuss how this limitation is likely to affect the evaluation, and how the limitation will be mitigated.

For munitions, the end-to-end test can become a critical part of the LFT&E strategy. In an end-to-end test, the target aimpoint is selected operationally. Including this data increases the operational realism of the LFT&E. To be used as part of the LFT&E, full-up munitions must be used, targets must be realistic, and a damage assessment must be completed.

Systems often rely on other systems to complete missions. For these system-of-systems, the test and evaluation should address the impact of all systems to the mission, not just the system under test. It is possible that the system under test meets its requirements, yet cannot accomplish its mission due to the performance of another system.

For system-of-systems, end-to-end testing will involve systems other than the system under test. This can complicate test coordination when the additional systems are under the control of another program office. In these cases, DOT&E may require:

- That the availability of the critical system become an entrance criteria
- TEMP coordination signatures of the project office(s) responsible for the system(s)
- A capstone TEMP as described in DoD Instruction 5000.02.

References

Reporting of Operational Test and Evaluation Results, DOT&E, January 6, 2010

Guidance on the use of Design of Experiments (DOE) in Operational Test and Evaluation, DOT&E, October 10, 2010

Examples

End to End Testing – Examples

CARGO AIRCRAFT EXAMPLE

3.6 Operational Evaluation Approach. Operational testing of the C-100 cargo aircraft will employ the mission profiles as required by the CPD and described below. The missions will demonstrate delivery of time-sensitive/mission-critical supply items and/or personnel over operational/tactical distances to forward-deployed forces in remote and austere locations. Approximately 50 missions will demonstrate all variations of the mission profiles. Missions will include short notice logistical re-supply, casualty evacuation, troop movement, and aerial sustainment. The C-100 will operate to and from smaller, unimproved tactical landing strips and improved airfields up to the maximum cargo gross weight. The C-100 will be off-loaded to tactical rotary-wing aircraft and ground vehicles to demonstrate transloadability at Forward Operating Bases (FOBs) located near supported tactical units. The ability to rapidly reconfigure the C-100 will be evaluated. To evaluate adverse weather capability, the C-100 will conduct missions during day, night, night vision goggles (NVG), Visual Meteorological Conditions (VMC), and Instrument Meteorological Conditions (IMC).

The first three mission profiles will be flown under day/night/NVG conditions to improved and unimproved runways, carrying various load configurations (463L pallets, troops, and vehicles), and will require 20 missions and approximately 64.0 flight hours.

Mission profiles 4 and 5 will include aircraft reconfiguration for aeromedical evacuation. Missions will be flown under day/night/NVG conditions to improved runways carrying various load configurations (463L pallets, troops, vehicles, and litter patients), and will require 16 missions and approximately 48.0 flight hours.

Mission profiles 6 and 7 will demonstrate single and multiple airdrops (four static line airlifts with door bundles and static line paratroop drops, and four military freefall airlifts). Airdrop missions will be flown under day/night/NVG conditions and will require eight missions and approximately 30 flight hours to demonstrate.

Mission profile 8 will demonstrate aerial sustainment under day/night/NVG conditions to improved runways, and will require approximately five missions and 34 flight hours.

Mission profile 9 will demonstrate self-deployment under day/night, visual flight rules/instrument flight rules (VFR/IFR), and will require one mission and approximately 40 flight hours.

ARMY MUNITION EXAMPLE

3.6 Operational Evaluation Approach. The guided missile will be evaluated end-to-end. It is not possible to conduct the end-to-end mission in a single event due to availability of the unit, availability of real-time imagery of the test area, and delays between firing missions caused by the need to collect target data. Instead, the evaluation will be based on two operational events. The ground IOT&E will test the ability of a fire support unit to plan, target, and execute guided missile missions. The flight IOT&E will test the unit's ability to fire guided missiles and examine the missile's effects on actual threat targets. During the ground phase, an operational unit will target and execute guided missile missions while executing other missions at an operational pace. Using satellite imagery of the actual test targets, the unit will mensurate the image using fielded equipment to estimate the target's location. Using fielded command and control equipment, the unit will determine the number of missiles and aimpoints. The mission information will be sent through the command and control chain to the launcher, which will dry-fire the missile. The flight phase will execute the missions generated during the ground phase. The test officer will digitally send a fire mission with aimpoints and number of missiles (determined in the ground IOT&E) to a battery command post. The battery will forward the fire missions to the launcher, which will move to a launch point and, after a brief safety delay, fire the missiles. The flight phase targets are threat-representative targets with threat-approved countermeasures. The Army Research Laboratory will conduct a damage assessment for each mission. The assessments are a critical component of the LFT&E strategy.

Details of the ground IOT&E, flight IOT&E, and LFT&E would be provided in other sections of the TEMP.

Information Assurance – Examples

3.6.1 Operational Test Objectives

Example 1

Computer network operations threat testing will be conducted by a certified and accredited Threat Computer Network Operations Team. Threat representation and portrayal will be consistent with the approved threat description documentation. Events may include but are not limited to insider/outsider computer network exploitation, captive/overrun attempts, network penetration, data compromise, denial of service attacks, network flooding, spoofing, data corruption, radio frequency/directed energy weapons, physical destruction, direction finding, jamming, hacking, malicious code, and unauthorized users.

Blue Team and Red Team vulnerability assessments will be conducted by the Service Information Warfare Center. Blue Teams will conduct scans to identify vulnerabilities and assist in assessing tactics, techniques, procedures and training. Red Teaming will consist of insider/outsider computer network exploitation, including captive/overrun attempts at password cracking, network penetration, access to router configuration files, data compromise, denial of service attacks, network flooding, spoofing, and potentially some type of benign data corruption. Continuity of operations, including alternate site transfer, will be exercised during the operational test.

Example 2

Information Assurance. An evaluation of the system's operational vulnerabilities and information assurance protect, detect, react, and restore capabilities will be conducted. Information assurance evaluation will include an observation of fleet operators performing a posture transition. An operational Information Assurance Vulnerability Evaluation will include the technical and non-technical assessment of information assurance implementation measures to discover vulnerabilities. A Protection, Detection, Reaction, and Restoration Evaluation will use penetration and exploitation techniques to measure the exploitation of discovered vulnerabilities and the performance of IA capabilities under operational conditions. The following test events will be required to complete this evaluation:

- Penetration testing of the premise routers via the network for a 7-day period or for two 4-day periods. Unclassified and Secret testing will orginate remotely from NIOC. Top Secret testing will originate from the [organization deleted]. The test platform will be required to maintain continuous network connectivity via radio frequency or pier connection. This event must be completed prior to testing of the Periods Processing LAN.[1]

- Penetration testing of the Periods Processing LAN in Secret Posture via the network for an 8-day period will originate remotely from NIOC. The 8 days of testing do not include any NAVSEA or TYCOM required pre-test work or post-test certification.
- Penetration testing of system representative Enclave Guard in an accredited laboratory.
- Penetration testing of a system ESM with the ESM Enclave Guard in an accredited laboratory.

[1] The "Periods Processing LAN" is the name of one of the subsystems of the larger platform.

Production-Representative Test Articles - Guidance

Summary

Consistent with the goal of "flying before buying" major systems for the Department of Defense, operational testing in support of Full-Rate Production decisions must be conducted with production systems or production-representative test articles. Whenever practicable, production systems are be furnished from low-rate initial production (IOT&E) quantities. Through the TEMP, DOT&E can approve the use of production-representative test articles in lieu of production test articles. In evaluating whether systems are production-representative, DOT&E will consider whether the test articles were assembled using the parts, tools, and manufacturing processes intended for use in full-rate production. The system should also use the intended production versions of software. In addition, the logistics system and maintenance manuals intended for use on the fielded system should be in place. DOT&E must be provided detailed information describing any process differences in order to independently evaluate whether the differences are acceptable.

References

Use of Production-Representative Test Articles for Initial Operational Test and Evaluation (IOT&E), DOT&E, October 18, 2010

Defense Acquisition Guidebook, Paragraph 2.3

DODI 5000.02

Examples

Paragraph 3.6.1 Examples

Paragraph 4.1.1 Example

Production-Representative Test Articles – Examples

Example 1

3.6.1.2 Configuration Description. The IOT configuration will be a Dakota helicopter company with five LRIP Dakota aircraft and all authorized equipment, pilots, and maintenance personnel and support equipment.

Example 2

3.6.1.2 Configuration Description. The IOT configuration will be 15 production-representative Gemini missiles with complete capability as required by the CPD. The missiles are production systems with the exception of "white wires" in the guidance module used to fix a problem discovered late in developmental testing. In production, this "white wire" will be replaced by firmware circuitry. These missiles have been assembled at the production facility. Maintenance and support equipment is production representative.

Note the exceptions to the use of production test articles

Threat Representation – Examples

3.6.1 Operational Test Objectives

[The following example addresses only threat representation in the context of overall OT objectives. Typically, other OT objectives will also be described in this paragraph.]

The IOT&E for the Sea Shark missile will require the development of a new threat surrogate that matches the anticipated threat in altitude and speed. The Program Manager will fund the development of 10 surrogate threat systems and the associated verification/validation studies. Operational Test Activity will accredit the surrogates for use in IOT&E. In addition to developing a high fidelity threat surrogate for IOT&E, the Navy will develop the capability to launch multiple simultaneous threat surrogates to support the first FOT&E.

Test Limitations – Examples

Background for Maritime Air Defense Example

This example is for the hypothetical Sea Shark missile (ship-launched, anti-air, semi-active radar homing missile, supported by the hypothetical Neptune Combat System (NCS)). Critical operational issues (COIs) for Sea Shark and its supporting combat systems include:

- Area Air Defense Capability – Can Sea Shark, supported by the NCS, provide air defense for other ships within the Aircraft Carrier Strike Group?
- Own Ship Air Defense Capability – Can Sea Shark, supported by the NCS, provide own ship defense against air threats while also conducting Area Defense?
- Availability – Can Sea Shark, after a representative shipboard storage time in the vertical launch cell, provide the required launch availability?
- Reliability – Can Sea Shark, after a representative shipboard storage time in the vertical launch cell, provide the required in-flight reliability?

3.6.3 Test Limitations

Quantities of Sea Shark missiles will be limited, possibly precluding re-engagement of surviving simulated threats. In some scenarios, threats/surrogates might survive initial engagement, thus requiring deployment of a second Sea Shark. The test plan does not provide enough Sea Shark missiles to support a second launch. This is a departure from operational realism. At most, the test unit will conduct a simulated Sea Shark missile launch against surviving surrogates.

Planned mitigation includes:

- Once the M&S is validated with the initial IOT&E results, conduct simulation using the available Office of Naval Intelligence digital models for the threats and simulated Sea Shark missile re-engagement of surviving threats. This would provide an early prediction of how Sea Shark and the NCS could respond against surviving Anti-Ship Cruise Missiles (ASCM) threats.
- Follow-on OT&E (FOT&E) will be scheduled at the earliest opportunity when production Sea Sharks are available to support OT addressing re-engagement of simulated threats that survive initial engagement.

Current test range target launch and control capability will limit the number of simultaneous targets in flight and thus, the size of simulated ASCM raids. Sea Shark is required to defend against multiple simultaneous threats, but the test range is unable to launch and track multiple simultaneous threat systems.

Mitigation efforts include the following:

- Once the Sea Shark and NCS M&S capability is validated by initial IOT&E results, simulated engagements will be conducted against threat large ASCM raids to predict results for interim fleet tactics development.
- The Navy will upgrade the test range facilities to support multiple simultaneous engagements prior to the first FOT&E.

Missiles will not have representative shipboard magazine storage times by the time of operational testing. Missiles must be fielded and in representative storage magazines for one year before steady-state availability and reliability levels will be known.

Mitigation efforts include the following:

- The M&S reliability growth curve will estimate system reliability after fielding. The growth curve will be adjusted as needed based on results of IOT&E and accelerated life testing of guidance, fuze, and propulsion components.
- Availability and reliability of Sea Shark missiles with representative magazine storage times will be evaluated during the first FOT&E.

Adequate Test Resources – Guidance

Guidance

The program manager, in coordination with all T&E stakeholders, must identify and plan for all T&E resources needed to adequately support DT&E, OT&E, and LFT&E. The TEMP must describe the T&E program in sufficient detail for DOT&E to determine whether the resource estimates in the TEMP are reasonable and sufficient. TEMP updates must include updated T&E resource estimates, since the required resources may change as the understanding of the program matures. (Reference, DoDI 5000.02)

Requirements at specific milestones include the following DOT&E interest items:

- **At Milestone A**: Address the detailed test program resource requirements for the Technology Demonstration phase and the initial estimated lifecycle T&E program resources.
- **At Milestone B**: Update estimated T&E resource requirements (such as test articles, instrumentation, targets, threat simulators, modeling and simulation, distributed test networks, testbeds, range requirements, test support, etc.) for conducting all activities in the TEMP.
- **At Milestone C**: Include updated resource estimates for IOT&E, which shall be derived from defensible statistical measures of merit (power and confidence) associated with the coverage of the factors.
- **Post Milestone C**: The TEMP update shall provide for resources to support Follow-on Test and Evaluation activities.

Best Practices

Effectively planning for adequate OT&E and LFT&E resources requires *early agreement among DOT&E, the OTA(s), and the Service(s) on the scope of testing*. For its determination of whether adequate resources are planned and documented in the TEMP for each phase of OT, DOT&E will be particularly interested in the size of the test unit and threat force, the number of test articles, other operational force test support (personnel and equipment) (including provisions for baseline systems where appropriate to the evaluation strategy), test location and duration, OT-related modeling and simulation, ammunition, munitions, targets, and OT-related instrumentation (particularly instrumentation that requires separate developmental efforts). See example test resource table.

To possibly reduce required OT resources, training events, Service and Joint Forces Command experiments, and command post exercises should be considered as possible venues either to replace dedicated OT events (provided the test objectives can be met in such venues) or to augment operational testing. In terms of resources and funding for these events, DOT&E will expect the TEMP to be clear about who will fund/resource the training events, including augmented events and data collection requirements. The program would likely be required to contribute funding and resources to accomplish OT&E specific objectives.

For some systems, associated but separate programs provide OT&E funding or other resources. For example, the costs for testing a ship's air warfare mission area might be divided among the ship class being tested and one or more installed air defense systems that are themselves distinct programs. DOT&E and DASD(DT&E), in consultation with T&E and system stakeholders, may identify closely associated programs that require coordinated T&E activities. These programs must document a strategy for coordinated and combined T&E activities for their System of Systems in a Capstone TEMP, and must include adequate resources and where appropriate a defined distribution of test costs and resources among the related programs. Although coordinating the defined distribution can be complicated, there must be agreement among the TEMPs of the participating programs in terms of funding and resources. Regardless of whether funding and test resources are addressed in a Capstone TEMP, DOT&E will still expect that T&E funding and other test resources will appear in each programs' TEMP with the Service-established allocations described therein. See example funding table.

Finally, the TEMP's funding estimates in part 4 are essential for budgeting purposes. As a general rule, though, DOT&E expects that the testing described in the approved TEMP will be conducted, even if actual test costs later exceed the TEMP's cost estimates.

Adequate Test Resources – Example

Operational Test Events						
Test Event	**Date (Qtr/FY)**	**Test Articles**	**Test Sites**	**Funding" ($000)**	**Threat Representation Test Targets/Ammo**	**Operating Forces (OPFOR) (Personnel and Vehicles)**
Single Vehicle Directional Stability DT/OT	1Q/09	1 MCVP (EMD vehicle)	CamPen	Provided in Part IV	None	17 Marines with approach march load
Multi-Vehicle Directional Stability DT/OT	2Q/09	2 MCVP (EMD vehicles)	CamPen	Provided in Part IV	None	2 Reinforced Rifle Squad
Land Gunnery DT/OT	3Q/09-4Q/09	2 MCVP (EMD vehicles)	29P	Provided in Part IV	600 MK268 APFSDS-T; 600 MK264 MPLD-T/MK266 HEI-T LINK; 600 MK239 TP-T; 4000 7.62mm; 20 2.5D & 3D targets (BMP, BMD, BTR, BRDM)	None
Hot Weather DT/OT	4Q/09	2 MCVP (EMD vehicles)	29P	Provided in Part IV	2500 MK239 TP-T; 7200 7.62mm; 20 2.5D & 3D threat targets (BMP, BMD, BTR, BRDM, T72); 5 2.5D friendly targets (LAV, Bradley)	2 Reinforced Rifle Squad
MS C OA	2Q/11	3 MCVP & 1 MCVC (EMD vehicles)	CamPen, 29P	Provided in Part IV	600 MK268 APFSDS-T; 600 MK264 MPLD-T/MK266 HEI-T LINK; 4200 MK239 TP-T; 15000 7.62MM; 20 2.5D & 3D threat targets (BMP, BMD, BTR, BRDM, T72): 5 2.5D friendly targets (LAV, Bradley)	1 Reinforced Rifle Platoon, 1 Battalion Staff, 1 AAV Section w/crews, 1 M1A1 Section w/crews, 2 LAV Sections w/crews (1 section designated as OpFor), MAGTF Afloat Node, 1 Amphibious Ship (LPD), 1 LCAC, 1 81mm Mortar Section, 1 60mm Mortar Section Engineer Squad w/designated attachments, 1 Inf Co FST, FoF OpFor (2-4 LAV Sections and 1-2 Platoons of dismount infantry)
PABM DT/OT	1Q/12	2 MCVP (EMD vehicles)	29P	Provided in Part IV	700 rds MK239 TP-T; 2100 rds PABM; 4000 rds 7.62MM; 20 2.5D & 3D threat targets (BMP, BMD, BTR, BRDM, T72); 2 BTRs; 1 BRDM; 60 3D ballistic plywood mannequin	None
Regimental COC DT/OT	3Q/12-4Q/12	1 MCVP & 1 MCVC (EMD vehicles)	29P	Provided in Part IV	None	1 Regimental Staff
HW (Hot Wx) OA	3Q/12-4Q/12	3 MCVP & 1 MCVC (EMD vehicles)	29P	Provided in Part IV	2500 MK239 TP-T; 7200 7.62mm; 20 2.5D & 3D threat targets (BMP, BMD, BTR, BRDM, T72); 5 2.5D friendly targets (LAV, Bradley)	1 Reinforced Rifle Platoon, 1 Regimental Staff with COC, 1 Battalion Staff, 1 AAV Section w/crews, 1 M1A1 Section w/crews, 2 LAV Sections w/crews (1 section designated as OpFox), 20 threat representative targets (BMP, BMD, BTR, BRDM)
CW (Cold Wx) OA	2Q/13	3 MCVP & 1 MCVC (EMD vehicles)	CRTC, Valdez AK	Provided in Part IV	1000 Mk239 TP-T 3000 7.62mm	1 Infantry Platoon (reinf), 1 Bn Staff (Composition TBD), 20 Data Collectors, 1 DC Chief, 8 Control Cell, live fire and maneuver ranges, 1 Amphibious Ship (LPD)
IOT&E	4Q/14-2Q/15	12 MCVP 2 MCVC (LRIP Vehicles)	CLNC, CamPen, 29P	Provided in Part IV	7800 rds 30mm (AP and HE); 7000 rds 7.62mm; 5000 rds 40mm; 2500 rds 50cal Threat Rep EW & targets 8000 rds 30mm; 4000 rds 7.62mm; 5000 rds 30mm; 2500 rds 7.62mm; 5000 rds 40mm; 2500 rds 50cal 100 2.5D & 3D threat targets (BMP, BMD, BTR, BRDM, T72); 2 BTRs; 1 BRDM; 60 3D ballistic plywood mannequin	14 AAVP7A1, 1 reinforced rifle company(-), 1 AAVC7A1, Bn/Reg HQ staff, 4 M1A1 tanks, 10 LAVs (6 LAV-25, 2 LAV-AT, 1 LAV-L, 1 LAV-C) 8 Javeline Msl Sys, Mortar/Arty FDCs, 8 weapons vehicles (4 Mk19, 4 M2, 50 cal), GSR, 2 AH-1Ws, 1 UH-1N w/C&C or Airborne Relay, 2 AV-8s (20 flight hours), 2 F-18s, live fire test range, USN – 10 steaming days LSD/LPD (Flag configured), 2 LCACs, 2 RACs; Exercise control group personnel at MAGCC 29 Palms and CamPen CSSG Maint. Detachment; 1 CAX BLT exercise and 1 RLT size exercise

Adequate Test Resources – Funding Example

T&E Funding RDT&E (Then-yr $M)	**FY03**	**FY04**	**FY05**	**FY06**	**FY07**	**FY08**	**FY09**	**FY10**	**FY11**	**FY12**	**FY13**	**FY14**	**Total**
Requirements	2.9	10.5	5.62	5.5	9.923	5.91	2.348	4.354	8.0	3.13	8.463	1.8	68.448
DT&E/OT&E TOTAL								0.44	0.45				0.89
--- Testbed Develop, Integrate and Execute								0.44	0.45				0.89
DT&E Total		0.1	0.9	0.7	0.2	0.4	0.4		3.7				6.4
LFT&E (Total)	2.8	10.4	4.2	4.2	8.8	4.4	1.3	2.9	2.5	1.9	0.7	0.4	44.5
- TSST								1.9	1.0	0.4			3.3
- SAR/VAR	1.1	1.7	1.8	2.2	2.8	2.4	1.3	1.0	1.5	1.5	0.7	0.4	18.4
- Surrogate Tests	1.7	8.7	2.3	2.0	6.0	2.0							22.7
- FSST Comparative Study			0.1										0.1
OT&E (Total)	0.1		0.52	0.60	0.923	1.11	0.648	1.014	1.250	1.230	7.763	1.40	16.558
- Information Assurance Analysis, Reporting									0.100				
- Early Involvement (IT&E, EOA, OA)	0.1		0.52	0.60	0.923	1.11	0.648	0.808	0.614				5.323
- OPEVAL Planning, Analysis, Reporting									0.424	1.012	1.146		2.582
- OPEVAL – Air Warfare (AAW)/Ship's Self Defense Testing								0.206	0.212	0.218*	3.203*	0.230	4.069
- OPEVAL – Amphibious Mission											3.414		3.414
- FOT&E												1.170	1.170
RDT&E Funding Budgeted	2.9	10.5	5.62	5.5	9.923	5.91	2.348	4.354	8.0	3.13	8.463	1.80	68.448

All funding provided by PMA999, unless otherwise indicated

*$0.2M funded by PMA888 in FY12 and $1.2M funded by PMA888 in FY13

Production-Representative Test Articles – Example

4.1.1 Test Articles. The test articles and testing sequence for the Dakota program are defined in Table 19, Test Article Matrix. See Chapter 3 for additional details on each test event in this table.

Test Article	Test Event	Quantity	Start Date	Source
Prototype aircraft	DT	2	FY07	Contract
Prototype aircraft with ASE	LUT	2	FY10	Contract
Spare Parts for flight testing	All	As Needed	FY07	Contract
LRIP aircraft	IOT&E	5	FY12	Contract
LFT&E Components	LFT&E	See LFT&E Strategy	FY11	USG/Contract

Table 19 - Test Article Matrix

Note confirmation in resources section of the TEMP that LRIP test articles are planned.

Instrumentation - Guidance

Summary

In the conduct of operational testing, instrumentation is vital to identify with clarity what happens during test events. However, instrumentation data alone is generally not sufficient to explain why events unfold as they do and thus requires other sources of information, including interviews with operators and commanders. In general, instrumentation data is helpful in characterizing the environment and assessing Measures of Performance, but makes up only a portion of the data needed to assess Measures of Effectiveness.

When preparing a TEMP, specify in detail what instrumentation will be used to collect data on the system under test, and precisely what the instrumentation data will be used for in the evaluation. Factors and levels that are crucial to the evaluation should be identified in the Design of Experiments methodology. When possible, both DT and OT events should use common instrumentation to facilitate interpretation of the instrumentation outputs. The instrumented data should be collected carefully during the event to ensure that harvesting does not interrupt the operational context.

In addition to specifying the system performance instrumentation, the TEMP should delineate the real-time casualty assessment (RTCA) instrumentation to be used in OT events. This should include the description of the RTCA systems to be used and their quantities in both the Red and Blue forces.

Best Practices

An example of instrumentation used in support of operational testing is the Instrumented Field Data Collector[1] (IFDC) used in the Force XXI Battle Command Brigade and Below (FBCB2) and Early Infantry Brigade Combat Team (E-IBCT) assessments. The instrumentation system was physically attached to the test vehicles to capture and record all of the electronic message traffic that passed through the FBCB2, and was crucial to understanding the volume of message traffic flow between combat units, and the degree of situational awareness subordinate units had as a result of the presence of the digitization equipment. However, the presence of the IFDC was not sufficient to disclose everything necessary about the FBCB2 during the OT. Other

[1] IFDCs monitored digital message traffic and provided data on message completion rates.

sources of information, such as interviews with unit leaders and system operators, were also needed to assess the impact of improved situational awareness during operations.

Time/position/velocity/acceleration sensors are commonly used in developmental and operational testing.

References

Reporting of Operational Test and Evaluation (OT&E) Results, DOT&E January 6, 2010

Threat Representation – Examples

1.3.3.2 Special Test Requirements Example

As explained in section 1.3, Anti-Ship Cruise Missiles (ASCMs) are the primary threat to Naval Surface Ships. Critical attributes of ASCMs include speed, altitude profile, maneuverability, radar cross section, size and shape, infra-red (IR) signature, passive homing capability, countermeasures, and radar emissions. In planning for IOT&E, the ship-launched Sea Shark missile must intercept several ASCM threats, including the most prevalent ASCM, which has a cruise speed of 1.5 Mach and, upon achieving radar lock on its ship target, accelerates to 1.8 Mach and maintains that speed until ship impact. The threat also has the ability to descend from a 50-foot cruise altitude to 25 feet.

The available aerial threat surrogate has a relatively constant speed of 1.6 Mach and can be flown no lower than 40 feet. Accordingly, the adequacy of the IOT&E for the Sea Shark missile will hinge on the development of a new threat surrogate that more closely matches the anticipated threat in altitude and speed. The evaluation will also leverage missile flight test results from developmental testing to validate an end-to-end simulation model of threat and Sea Shark engagements. In addition to developing a high fidelity threat surrogate for IOT&E, the Navy will develop the capability to launch multiple simultaneous threat surrogates to support the first FOT&E.

Threat Representation – Examples

4.1.4 Threat Systems for Testing

Threat Nomenclature	Test				Source
	DT	LUT	IOT	FOT&E	
BRDM	1	3	6	6	PM ITTS/TSMO/ YPG
BMP (any variant)	2	2	4	4	PM ITTS/TMO/YPG
BTR (any)	1	2	3	3	PM ITTS/TMO/YPG
Red Tank (T72 or later model)	4	5	5	5	PM ITTS/TMO/YPG
T-80 Surrogate	1				
Red Truck (2.5T variant)	1	2	4	4	PM ITTS/TMO/YPG
ZSU-23		1	2	2	PM ITTS/TMO/TSMO
ZPU-23-4	2	1	2	2	PM ITTS/TSMO/YPG
2S1		2	3	3	PM ITTS/TMO/YPG
C3 (van)		--	1	1	PM ITTS/TMO
Blue Tank (M1)	2	3	5	5	PM ITTS/TMO/YPG
Militarized Civ Vehicles (Mix truck/SUVs/sedans)		6	10	10	YPG
HMMWV	2	2	6	6	
Blue Truck (LMTV)	2	2	4	4	FORSCOM/YPG
IFV (M2/3)	1	2	5	5	FORSCOM
M-60	2				PM ITTS/TMO/YPG
M113	2	1	4	4	FORSCOM

M&S for Test Planning, Pretest Predictions, and Evaluation - Guidance

The Modeling and Simulation (M&S) sections should address how M&S will be employed in the overall test strategy and how the M&S will be verified, validated and accredited (VV&A). Specifically, the TEMP should list any M&S expected to be used, the intended use, any data requirements, the test objectives to be addressed and/or how test scenarios will be supplemented with M&S, the planned VV&A effort, and who will conduct the VV&A effort (ref. DoDI 5000.61). The TEMP should list any specific test events required for VV&A of the M&S. The resources for the specific test events will be included in Part IV.

M&S is used throughout developmental, operational and live fire testing. M&S can support evaluation of requirements, trade studies, test planning, pre-test predictions, evaluation of system performance, and other uses. It includes a broad set of tools including spreadsheet models, high-fidelity digital models, and hardware and computer in the loop facilities.

M&S for Test Planning and Prediction

M&S is frequently used in test planning and pre-test predictions. If M&S will be used as a basis for decisions regarding test scope or test conditions, describe how it will be used. Which tests will have pre-test predictions? Which pre-test predictions will be based on M&S?

M&S for Evaluation

Another common intended use is that M&S can support the evaluation of system performance. The TEMP should describe how M&S will be used as a basis for evaluating any important system metrics and/or make significant contributions to the assessment of any Critical Operational Issues. Will M&S be used to supplement, interpolate, or extrapolate results from live testing? Briefly summarize the degree of reliance on M&S. Which evaluation issues will be addressed through M&S? Explain why M&S is being used (e.g., safety restrictions preclude testing).

sources of information, such as interviews with unit leaders and system operators, were also needed to assess the impact of improved situational awareness during operations.

Time/position/velocity/acceleration sensors are commonly used in developmental and operational testing.

References

Reporting of Operational Test and Evaluation (OT&E) Results, DOT&E January 6, 2010

3.2 Reliability Growth

The software reliability growth effort will start at the beginning of the software design effort in each of the nodes and/or components. Code design reviews will be held for each code module to ensure conformance with the particular contractors' standards and to identify and correct obvious errors. Beginning at the start of the Code and Unit Test (CUT) activity, quality metrics will be collected at all subcontractors for each of their coding efforts. For the Engineering, Manufacturing, and Development (EMD) phase of the program, collection and analysis will continue through all levels of code development, from CUT through Software Integration, Subsystem (node level) Integration, and System Integration.

3.2.1 Discrepancy Report (DR) Status

Each DR written against contractor-developed software will be prioritized into five levels as defined by the IEEE 12207 specification. Each DR will be initially assigned a level by the subcontractor developing that particular software. The prime integrator and the Government Program Office will perform an independent analysis and redefine levels accordingly. Graphs similar to Figures 1 and 2 will be maintained showing the number of open, closed, and resolved (fixed but not tested) statistics over time, by priority level.

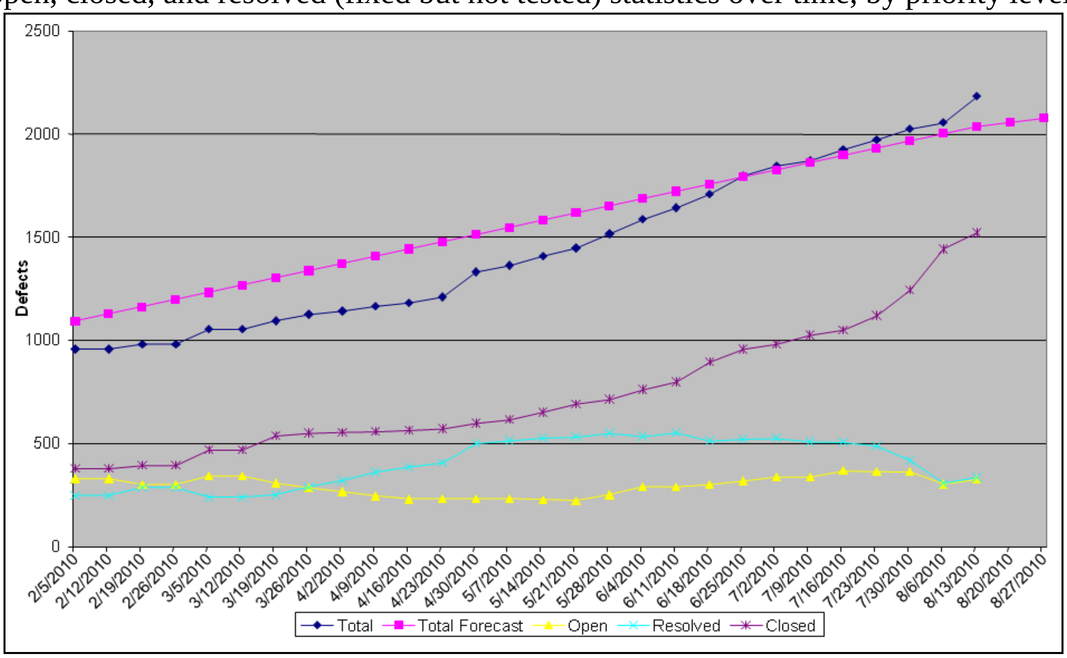


Figure 1. Example DR volume tracking (all priorities)

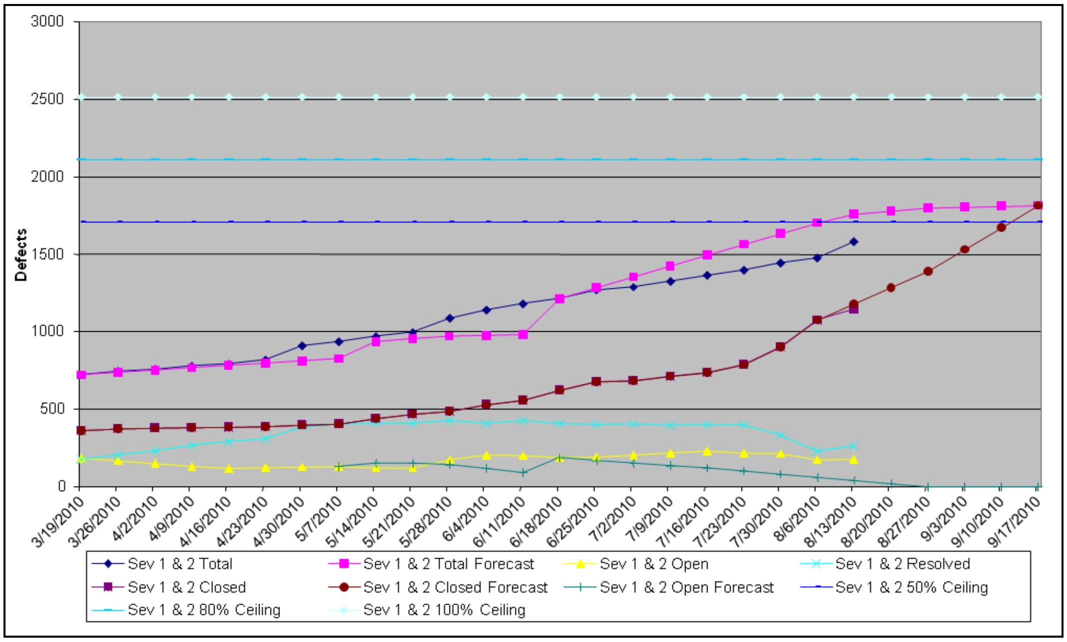


Figure 2. Example DR volume tracking (Priorities 1 and 2)

3.2.2 DR Aging

DRs at each priority level will be tracked to show how many of each level were open for a particular timeframe by priority. The timeframes will be separated into 30-day increments, up to a column for >120 days. The values in parentheses reflect the status from the previous reporting period. Example data are shown in Table 1.

Table 1: Sample DR Aging Metric

	Assigned and Submitted Defects – Days Open				
Severity	**0-30**	**31-60**	**61-90**	**91-120**	**>120**
1	9(18)	12(3)	1(1)	2(3)	4(2)
2	92(99)	41(28)	13(11)	5(8)	19(18)
3	48(45)	6(4)	8(11)	3(0)	16(18)
4	16(15)	3(3)	3(3)	1(1)	4(4)
5	0(1)	5(4)	0(1)	3(4)	4(3)
Total	165(178)	67(42)	25(27)	14(16)	47(45)

3.2.3 Commercial Off The Shelf (COTS) DRs

The ageing statistic described above will be maintained for issues found with commercially purchased equipment, such as routers, servers, etc.

3.2.4 Software Management Strategy

Every DR will be analyzed to determine the effect of the failure. Using this information, a determination will be made as to the severity of the problem (Priority, as defined by the IEEE 12207 specification). All failures that rate a Priority 1 or 2 will be fixed prior to entering the next phase of testing. These data will be collected and curves will be maintained throughout development and OT&E.

Realistic Operational Test Conditions - Guidance

General Guidance

Operational testing in support of Full-Rate production decisions shall be conducted under realistic operational conditions.

The Operational Test Agencies shall design the test and provide detailed tactics, techniques, and procedures to the participating forces to ensure realistic operational conditions. Other considerations for realistic operational conditions include typical operators and maintainers, a mission-oriented evaluation, the use of production representative test articles, adequate threat representation, end-to-end testing and baseline evaluation when appropriate, information assurance testing, and selection of mission-focused metrics in the design of experiments (DOE) analysis.

For each operational test, the TEMP will describe the resources, personnel, site selection, tactical considerations, and other factors intended to ensure appropriately realistic operational conditions. Specific resources and test articles will be described in of the TEMP.

References

Title 10, U.S. Code, Section 139

Test and Evaluation Policy Revisions**,** DOT&E, December 22, 2007

Realistic Operational Test Conditions - Examples

Example TEMP entry for generic sonar system:

3.6.1 Operational Test Objectives. OT will be conducted using an event driven and operationally realistic end-to-end scenario. Data gathered during previously completed IT and DT events will be considered in the evaluation. OT will be conducted using test events designed to assess all required capabilities of the sonar system and the ship's crew in operation of the system. The scenario will require the system to provide Undersea Warfare surveillance support to a Naval Strike Group. Within this scenario, the Blue Force test ship will sortie from port, conduct active, passive, and coordinated USW with friendly forces, and return to the port. USW operations will be conducted in deep, open ocean waters and Littorals against SSK and SSN threats executing validated threat tactics. Test sites will include representative levels of neutral shipping to provide realistic levels of interfering contacts. Threat forces will be tasked to aggressively pursue and attack the Naval Strike Group, and may preemptively engage the Blue Force test ship if possible.

Threat Representation – Examples

1.3.3.2 Special Test Requirements Example

As explained in section 1.3, Anti-Ship Cruise Missiles (ASCMs) are the primary threat to Naval Surface Ships. Critical attributes of ASCMs include speed, altitude profile, maneuverability, radar cross section, size and shape, infra-red (IR) signature, passive homing capability, countermeasures, and radar emissions. In planning for IOT&E, the ship-launched Sea Shark missile must intercept several ASCM threats, including the most prevalent ASCM, which has a cruise speed of 1.5 Mach and, upon achieving radar lock on its ship target, accelerates to 1.8 Mach and maintains that speed until ship impact. The threat also has the ability to descend from a 50-foot cruise altitude to 25 feet.

The available aerial threat surrogate has a relatively constant speed of 1.6 Mach and can be flown no lower than 40 feet. Accordingly, the adequacy of the IOT&E for the Sea Shark missile will hinge on the development of a new threat surrogate that more closely matches the anticipated threat in altitude and speed. The evaluation will also leverage missile flight test results from developmental testing to validate an end-to-end simulation model of threat and Sea Shark engagements. In addition to developing a high fidelity threat surrogate for IOT&E, the Navy will develop the capability to launch multiple simultaneous threat surrogates to support the first FOT&E.

Threat Representation – Guidance

Guidance

This is a sample of DOT&E guidance addressing the topic "Threat Representation."

The policy guidance contains additional links to the source policy documents if you wish to further investigate or understand the policy.

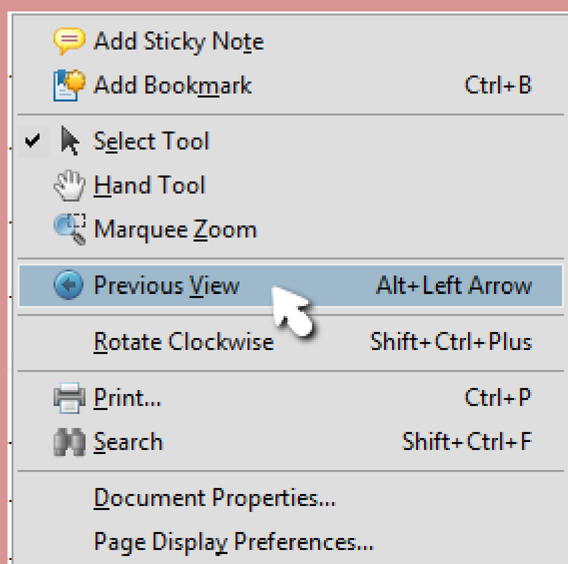
In … assist in evaluation … e goal for threat pres… UT), based on Defens… sessments. Particular … t are most relevant to … targets for demonstra…

To return to the TEMP Guide, press Alt + Left Arrow

Or

Right-click your mouse and select

"Previous View" in the dropdown menu



Th… in testing by includi…

- … he system
- … threat or
- … command … ed, attack … lity of the
- … imitations … gate those
- … ps, attack … t systems

Ide… IA threat assessments may lead to an early conclusion (that should be flagged as early as Milestone A TEMPs) that a credible, threat-representative surrogate does not exist and may require development to achieve an adequate IOT&E.

Thorough Service-sponsored technical and operational comparisons (validation) must be made between the threat and candidate surrogates. Validation culminates in a report documenting validation results. DOT&E monitors the validation and approves the Service-validated reports. Furthermore, careful consideration and documented

Threat Representation – Examples

This is a sample of an example paragraph 1.3.1 in a TEMP addressing the topic "Threat Representation."

The systems in these examples are notional, but based on real programs.

The examples are meant to convey the approximate level of detail and breadth that should be in most TEMPs. The content should be specific to your program and not a direct copy from this example.

1.3.1 Syst

Th … ntelligence Division, … ntelligence Agency-va … l of 2010. The follow…

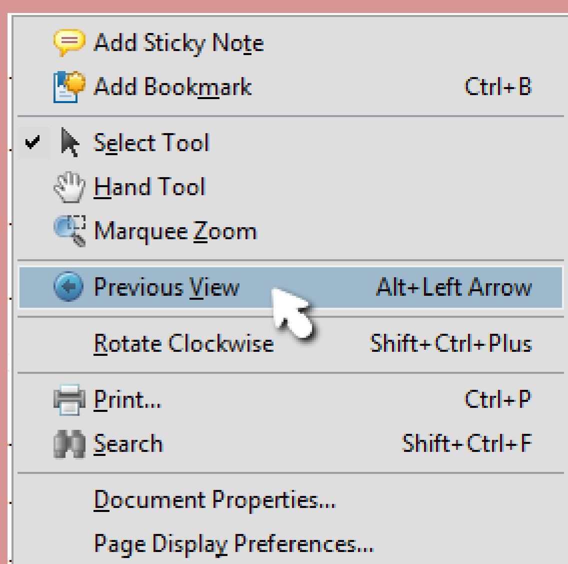
Mo … by fixed- and rotary … face-to-air missiles, a … electronic warfare systems, ground-based and airborne reconnaissance, surveillance, target acquisition … um, and a sophisticat … ern major weapon sy … Most of the regional … d various counterme … of enemy sensors an … ntries are exploring … vices that could evol … Most of the regional p … and some will acqui … me of the more tech … o conduct informatio … y modern warfightin … vary from country to …

To return to the TEMP Guide, press Alt + Left Arrow

Or

Right-click your mouse and select

"Previous View" in the dropdown menu



In … v-intensity conflict, a … ll employ tactics an … Generally, asymmetri … ated urban terrain, for concealment as well as political advantage, exploiting the indigenous environment and its inhabitants for surprise, escape routes, and shielding while negating a conventionally armed adversary's strength in numbers, equipment, and firepower. Asymmetric combatants will be armed with infantry small arms, rocket propelled grenades, light artillery and antiaircraft machineguns, man-portable antitank and surface-

Reliability Growth –Example

Figure 3.1. Top-Level Evaluation Framework Matrix

Key Requirements and T&E Measures				**Test Methodologies/Key Resources (M&S, SIL, MF, ISTF, HITL, OAR)**	**Decision Supported**
Key Reqs	**COIs**	**Key MOEs/ MOSs**	**CTPs & Threshold**		
KPP #2	COI #2. Is the Dakota suitable for…	Reliability & Maintainability	MTBSA ≥ 20 flight hrs MTBSA ≥ 26 flight hrs	Component level stress testing LUT point estimate Demonstrate at IOT with 80% confidence	PDR MS-C FRP
			MTBEMA ≥ 2.3 flight hrs MTBEMA ≥ 2.6 flight hrs	LUT point estimate IOT with 80% confidence	MS-C FRP

OFFICE OF THE SECRETARY OF DEFENSE
1700 DEFENSE PENTAGON
WASHINGTON, DC 20301-1700

OPERATIONAL TEST
AND EVALUATION

NOV 0 4 2010

MEMORANDUM FOR DEPUTY ASSISTANT SECRETARY OF DEFENSE FOR INFORMATION AND IDENTITY ASSURANCE
DEPUTY UNDER SECRETARY OF THE ARMY, TEST AND EVALUATION COMMAND
DEPUTY, DEPARTMENT OF NAVY TEST AND EVALUATION EXECUTIVE
DIRECTOR, TEST AND EVALUATION, HEADQUARTERS, U.S. AIR FORCE
TEST AND EVALUATION EXECUTIVE, DEFENSE INFORMATION SYSTEM AGENCY
TEST AND EVALUATION EXECUTIVE, NATIONAL SECURITY AGENCY
COMMANDER, ARMY TEST AND EVALUATION COMMAND
COMMANDER, OPERATIONAL TEST AND EVALUATION FORCE
COMMANDER, AIR FORCE OPERATIONAL TEST AND EVALUATION CENTER
DIRECTOR, MARINE CORPS OPERATIONAL TEST AND EVALUATION ACTIVITY
COMMANDER, JOINT INTEROPERABILITY TEST COMMAND

SUBJECT: Clarification of Procedures for Operational Test and Evaluation of Information Assurance in Acquisition Programs

DOT&E memorandum of January 21, 2009, *Procedures for Operational Test and Evaluation of Information Assurance in Acquisition Programs*, provides guidelines for verifying the effectiveness of information assurance (IA) and computer network defense (CND) measures in acquisition programs during operational test and evaluation events. I expect these guidelines to be adhered to for programs that exchange information and are under DOT&E oversight. Operational Test Agencies should consider incorporating that policy into their procedures for similar programs that are not under oversight. It has come to my attention, however, that some programs have misinterpreted that memorandum to imply IA and CND testing must be done during the Initial Operational Test and Evaluation (IOT&E) period. This memorandum is to clarify that is not the case.

To be effective, IA and CND measures need to be tested as early as possible once hosted on the operational network. I encourage this testing be planned and executed in an early integrated test venue, if possible. The results of the testing should be shared to all stakeholders as soon as they are available. If a penetration test was conducted within one year, even if by another entity, the results of that test can be used in lieu of another test as long as sufficient information is available to understand the threat emulated and verify the adequacy of the test. Information from this test as well as that obtained from Developmental Test and Evaluation and Information Assurance Certification efforts must still be assessed and included in the Operational Test Agency's IOT&E report.

To support proper planning, the Test and Evaluation Master Plan (TEMP) should clearly identify what events will support the IA and CND assessment and address the requirements in DOT&E's memorandum of January 21, 2009. In particular, the TEMP must specifically identify the Red Team requirements for the penetration test and the required threat to be emulated. If the system is connected to a network with other systems with higher Mission Assurance Category (MAC) and Confidentiality Level (CL) than the system under test, I expect the threat emulation to be planned to the higher MAC and CL. If the threat level cannot be emulated on the operational network, the TEMP should identify what closed test network will be used and the verification and validation effort needed to ensure that network provides an operationally realistic environment.

As with all guidelines, I will consider deviations from prescribed practices on a case by case basis.

J. M. Gilmore

J. Michael Gilmore
Director

cc:
ASD (NII)
DDT&E