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

OPERATIONAL TEST
AND EVALUATION

SEP 14 2010

MEMORANDUM FOR DEPUTY UNDER SECRETARY OF THE ARMY, TEST & EVALUATION COMMAND
DEPUTY, DEPARTMENT OF THE NAVY TEST & EVALUATION EXECUTIVE
DIRECTOR, TEST & EVALUATION HEADQUARTERS, U.S. AIR FORCE
TEST AND EVALUATION EXECUTIVE, DEFENSE INFORMATION SYSTEMS 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: Guidelines for Operational Test and Evaluation of Information and Business Systems

To support agile acquisition of Information and Business Systems, the attached guidelines may be substituted in place of the traditional operational test and evaluation approach described in DoD Instruction 5000.02 of December 8, 2008, *Operation of the Defense Acquisition System.*

It is expected a large portion of the test strategy for Information and Business Systems will utilize an integrated test approach. The degree of independent operational testing appropriate for each software increment or capability can be tailored by using the risk analysis described in the attached guidelines. The guidelines also permit delegation of test plan approval using the same criteria. These guidelines do not apply to weapons systems or strategic and tactical command and control systems.

This memorandum supersedes the Director, Operational Test and Evaluation Memorandum of June 16, 2003, *Guidelines for Conducting Operational Test and Evaluation for Software-Intensive System Increments.*

J. M. Gil

J. Michael Gilmore
Director

Attachment:
As stated

cc:
DDT&E

GUIDELINES FOR OPERATIONAL TEST AND EVALUATION OF INFORMATION AND BUSINESS SYSTEMS

1. PURPOSE AND APPLICABILITY

These guidelines provide a means to tailor operational test and evaluation (OT&E) using a risk assessment of information and business systems deliverables.[1,2] These guidelines supersede the Director, Operational Test and Evaluation (DOT&E) memo titled Conducting OT&E of Software Intensive System Increments (June 16, 2003). Throughout these guidelines, the term "Applicable Programs" connotes all information and business systems that are Major Automated Information Systems (MAIS) or are under DOT&E oversight.

In order to take advantage of rapid advancements in information technology these systems must be efficiently acquired, tested, and fielded. However, the capabilities in the deliverable must undergo "adequate" operational test and evaluation before being deployed for use within the Department of Defense. The determination of "adequate" depends on the risks to mission accomplishment involved. To that end the Operational Test Agencies (OTAs) and, for Applicable Programs, DOT&E must agree early on the level of risk and the corresponding adequate level of OT&E for all capabilities that are to be deployed.[3, 4]

2. GENERAL APPROACH

A risk analysis will be conducted by the lead OTA documenting the degree of risk and potential impact on mission accomplishment for each capability. The results of this analysis are expected to be part of the program's test and evaluation (T&E) documentation (the T&E Strategy (TES) or T&E Master Plan (TEMP), depending on program maturity) and will be used to determine the appropriate level of OT&E to assess operational effectiveness, suitability, and survivability/security.[5] For all MAIS programs as well as programs under DOT&E oversight, the results will be provided to DOT&E for approval of the tailored OT&E approach.

[1] For the purposes of these guidelines, information and business systems are computer-based information systems executing one or more application software programs used in the acquisition, storage, manipulation, management, movement, control, display, switching, interchange, transmission, or reception of data or information. The capability being developed is primarily provided through software that is hosted primarily on commercial products.

[2] For the purposes of these guidelines, the deliverable is defined by the particulars of the acquisition program as whatever will be fielded to users to provide an operational capability, e.g., a software release, upgrade, patch, install, increment, spiral, project, technology refresh, business process change, etc.

[3] Agreement required for all MAIS programs as well as those under DOT&E oversight.

[4] A risk-based approach could conceivably apply to T&E for any program whatsoever, but is unlikely to be helpful outside the realm of information and business systems due to prevalence of risks associated with the potential for loss of life.

[5] Typically, survivability testing for information and business systems will be based on information assurance. In some services and agencies, information assurance capability is addressed as security rather than survivability.

a. Risk Assessment

The risk assessment combines the distinct concepts of risk likelihood[6] and risk impact[7] of a capability failing to be operationally effective, suitable, and survivable/secure. Several methods and approaches are possible for completing the risk assessment. This document provides guidance on methods along with questions and considerations for performing the risk assessment.

b. Levels of OT&E

There are three levels of possible OT&E for Applicable Programs. Programs should always plan an integrated test and evaluation strategy to fully assess all capabilities in a given deliverable. The degree of additional independent operational testing is determined by the OTA's risk analysis. The design of testing activities at each level of OT&E must be based upon the fundamental objective of evaluating operational effectiveness, suitability, and survivability/security as expressed in the Critical Operational Issues (COIs).

Level I OT&E – An assessment primarily using data from integrated test events other than a dedicated independent operational test event, e.g., developmental tests, certification events, and independent observations of the capability being used in operationally realistic or representative conditions. Even for programs under DOT&E oversight, the assessment plan is approved by the lead Service or agency OTA.

Features of Level I OT&E are:

- The OTA influences and monitors selected test activities including recommending inclusion of test cases for examining specific operational issues, and collecting data for the evaluation.
- Contractor participation is in accordance with the nature of the test events with consideration given to fielding plans of the system.
- For acquisition and fielding decisions, the OTA must confirm that the program has plans in place that address recovery from failures and resolution of shortfalls discovered in test events.
- The OTA prepares and provides an appropriate independent evaluation or assessment to support the acquisition and fielding processes and, for Applicable, provides a copy to DOT&E.

Level I OT&E is appropriate for capabilities having low risks. Typical deliverables with low risk capabilities where Level I OT&E is anticipated are maintenance upgrades, hardware upgrades, and software patches containing only minor capabilities or enhancements.

Level II OT&E– An evaluation that includes an independent operational event, which is carried out by typical users in an operationally realistic or representative environment to assess risk-specific factors of operational effectiveness, operational suitability, and survivability/security. The evaluation primarily uses data independently collected during the independent operational event, but also includes data as appropriate from other integrated test program events. The lead Service or agency OTA approves the test plan.

[6] "Likelihood" and "probability" used interchangeably here

[7] "Impact" and "consequence" used interchangeably here

Features of Level II OT&E are:

- Typical users in their operational environment performing tasks. One or more operational sites might participate and the OTA might prescribe scripted events in addition to normal user activity.
- Contractor participation is limited to that prescribed in the program's support plan.
- For acquisition and fielding decisions, the OTA must confirm that the program has plans in place that address recovery from failures and resolution of shortfalls discovered in test events.
- The OTA prepares an appropriate independent evaluation of operational effectiveness, operational suitability, and survivability/security to support the acquisition and fielding processes and provides a copy to DOT&E.

Level II OT&E is appropriate for capabilities having a moderate level of risk with limited potential for mission disruption. The Level II OT&E is typically suitable for modest, self-contained, operational capabilities.

Level III OT&E – An evaluation of the operational effectiveness, operational suitability, and survivability/security of the operational capability using the COIs and an independent dedicated operational test. This is the highest level and most comprehensive of OT&E. DOT&E will approve the operational test plan.

Features of Level III OT&E are:

- Level III OT&E must comply with statutes and all provisions of the DoD 5000 series regulations.
- The OTA carries out test events in an operational environment.
- The OTA independently evaluates and reports on the operational effectiveness, operational suitability, and survivability/security using all available data, especially independently collected operational test data, to support the acquisition and fielding processes with a copy provided to DOT&E.
- All test data will be provided to DOT&E for independent analysis and reporting.

Level III OT&E is appropriate for Applicable Programs that have capabilities with high risks. Level III OT&E is typically appropriate for significant or new operational capabilities with high potential for mission disruption.

3. IMPLEMENTATION

a. Assess risk. The OTA, with support from the program office, user representative, and threat community, assesses and documents the risks. Risk assessments are developed using the OTA's preferred procedures. Assessments must distinguish between the likelihood of occurrence and the severity of the mission impact if the risk is realized (no matter how unlikely). The OTAs may have or develop their own risk rating scales, and no specific rating scale is required. A three point scale is used in this discussion for illustrative purposes.

Information and business systems typically involve deliverables that are comprised of multiple operational capabilities.[8] The basic method is to assess the level of risk (likelihood and mission impact) to each of the operational capabilities comprising the deliverable. This assessment determines an appropriate level of OT&E for the operational capabilities and thus the deliverable. It is likely that some capabilities may be adequately assessed using data only from integrated testing while within the same deliverable an independent operational test may be needed to assess the most critical capabilities.

(1). Mission Risk Categories

Risk assessment begins with Risk Identification.[9] This guide addresses four risk categories that represent examples of things to consider when assessing risk. DOT&E expects the OTAs to evaluate risk categories, questions, and considerations that best reflect the deliverable and operational capabilities being assessed. In all cases, the OTA will perform the risk assessment with support of the program management office, user representatives, and threat community. The four risk categories are:

- Technology and Software Development (including software reliability)
- Integration and Deployment
- Training, Utilization, and Management
- Information Assurance

Note: The goal is to identify and assess the risks to operational mission goals that might occur once the system is deployed, not the technology maturation and manufacturing risks that might prevent system acquisition.

The risk categories include human and organizational factors that are involved in deployment, training, and business process change. The acquisition of information and business systems is not complete until they are deployed and actively in use. Operational deployment, training, and process issues are often more complex than technical issues in software-intensive projects and these human and organizational factors are often cited as the root cause of Information Technology project shortfalls.[10] Human and organizational factors must be considered in risk assessments and OT&E.

The following example questions represent topics that should be addressed when assessing the four risk categories. They should be tailored as appropriate for the particular information or business system that is being assessed. The OTAs are encouraged to add their own questions and any additional risk categories they think appropriate.

[8] For the purposes of these guidelines, an operational capability is a useful, meaningful, and supportable capability provided by a software-intensive system or information technology.

[9] See also the information assurance risk assessment methodology in NIST SP 800-30 or the Software Engineering Institute's "Taxonomy-Based Risk Identification," CMU/SEI-93-TR-6 ESC-TR-93-183.

[10] See, for example, Brooks Jr., Frederick P., *The Mythical Man-Month* (20th Anniversary Edition), Addison-Wesley, Reading, Mass., 1995.

Technology and Software Development

This risk category represents the well-known concern that software can have "bugs" and/or be developed with incorrect understanding of user needs.

- How complex is the capability (lines of code or other industry standard complexity metrics)?
- How dependent is the capability upon new technologies (hardware and software)?
- What is the commercial tempo of change in the technology areas represented in the capability and how mature are the technologies that are used?
- How many agents (government, contractors, sub-contractors) participated in the development of this capability?
- If the capability is primarily commercial off-the-shelf (COTS), non-developmental-item (NDI), or government off-the-shelf (GOTS), what is the past performance and reliability? For new technologies, what is the performance record in other applications? Are custom modifications of base software planned?
- For newly developed software, what is the developer's Capability Maturity Model rating as defined by the Software Engineering Institute? Were any Priority l or Priority 2 problems experienced with previous increments from this development team? [11]
- Does the developer employ a robust set of software management indicators?
- Is there a plan for collecting and reporting software reliability metrics, including discovery of new failure modes and closure rate?
- Do the requirements documents clearly and unambiguously state the performance requirements, testability metrics, use cases, and operational constraints for the software?
- How stable were/are the system requirements?
- What is the proportional change to system hardware and software introduced by the new capability (100 percent new; small patch to large deployed system; etc.)?
- What is the cumulative change to system hardware and software since the last full operational test?
- Does the developing contractor's test agent have sufficient experience and technical expertise to conduct a proper technical evaluation? Was thorough integration and regression testing conducted? Have clear exit criteria been identified for developmental testing of this capability?
- Are the mission, business process environment, hardware, and software hosting system clearly defined?

[11] As defined in IEEE/EIA Standard 12207.2-1997, Annex J.

Integration and Deployment

This risk category relates to signal and data environment; program interfaces to the operating system and user input, interfaces to legacy databases, messaging, communications protocols, and local configuration files; published and actual software service specifications; interoperability; real time processing issues; competency and accurate record-keeping of system administrators tasked with software installation, and other aspects of distributed computing.

- Does the capability require integration with another new or immature system? Are interfaces with existing systems fully documented and under configuration control?
- How complex are the external system interface changes (hardware, software, data, signals, and messaging) required for the capability?
- What is the status of the Certification and Accreditation package?
- Are mature Interface Control Documents available and approved?
- Must the capability interact with fielded, legacy databases?
- Must the capability interact with fielded, legacy configuration files?
- Must the capability interoperate with other systems? Are any of those systems also in development?
- How complex are the user interactions? How mature is the interface that captures and conditions the user input?
- How numerous and complex are the data inputs? Are there real time signals? What is the bandwidth of the data processing demands? How many different input or data types? How mature is the interface that captures and conditions the input data?
- Does the capability implement a change in executive software (operating system or database management system)?
- Does the new capability introduce changes that place in jeopardy or modify the system data structures?
- Does the capability introduce any new standards or protocols?
- Does the integration of the entire system (e.g., hardware, software, communications, facilities, management, operations, sustainment, personnel) present unusual challenges?
- Has the receiving organization developed plans for continuity of operations during the installation of the new increment?

Training, Utilization, and Management

This risk category relates to user training and organizational buy-in; tactics and procedures for sustainment; and usability issues.

- Is the user committed to the successful implementation of the new capability? Is the receiving organization committed to the successful implementation of the new capability?
- Have mission needs been adequately described and user requirements clearly identified? Is the capability traceable to requirements? Do the requirements address operational needs rather than specifying a technical solution?
- How extensively have prototypes been used to evaluate acceptance by typical users? Is the user interface intuitive or does it require extensive explanation?
- Does the system include the necessary system administration capabilities?
- Have operational and user support procedures been developed and readied for implementation? Have user representatives developed appropriate concepts of operations, policies, procedures, training, support, and contingency plans for a full operational deployment?
- Do the operators possess the skill levels required to use the increment's capabilities effectively? Has an adequate training plan been developed or implemented to include reorientation and sustainment training?
- Has a point of contact been established to represent the views of users?
- Is the receiving organization prepared for the changes in business processes associated with the new capability?
- Have the human factors and man-machine interface impact on the system performance been adequately considered?
- Does the capability or system present any safety hazards to the operators or operational environment? What data and physical effectors does it control and what damage can they do?
- Does the capability present lists or other data (menus, file names, comments, tags, etc.) that are created by users and will accumulate over time? Is there any effective limitation on that data accumulation or process for data aggregation?
- Are there simultaneous, asynchronous database access threads? Can row or transaction locking result in wait-cycles for users? Has the maximum number of simultaneous user requests in a given time window been determined?

Information Assurance

This risk category relates specifically to the survivability/security assessment.

- Are there foreign sources of component chips? Is there a risk of counterfeit parts?
- Are the network interfaces and information exchanges defined, including all relevant data sources?
- Does the new capability affect system security via new or altered external interfaces?

- Is administrative access granted to software in order to enable installation? Are new high-access accounts required for the newly installed capabilities?
- Do security protocols of new deliverable map cleanly to existing protocols?
- What is mission assurance capability (MAC) and confidentiality level (CL) for the deliverable?
- Who is the Designated Approval Authority to connect to the network?
- Have the DoD 8500 information assurance controls been assessed?
- Who will do the scanning for the vulnerability assessment?
- Who will do the vulnerability assessment and penetration testing if necessary?
- Who will identify mitigation techniques for vulnerabilities found?
- Are there tools available to detect penetrations?
- If vulnerabilities are detected, are the react and respond procedures identified?

(2). Likelihood of Risk Occurrence

Once risks have been identified, the risk assessment will distinguish the likelihood that a risk will occur vice the consequence if the risk does occur. The OTAs may use any point system they commonly use for rating risk likelihood. Table 1 below is an example of a three point scale. Every risk identified in step (1) for a given capability should be rated for its likelihood. When in doubt, the likelihood should be rated high.

Table 1. Likelihood of Occurrence at IOT&E/Follow-on Test and Evaluation

		Estimate of likelihood of issue occurrence during OT&E given the program's demonstrated maturity rate to date:
Level	**Descriptor**	
1	Negligible	One can reasonably assume no occurrence.
2	Possible	Issue is possible but not likely. It cannot be ruled out.
3	Likely	Issue has a significant chance of occurrence. It would not be surprising.

(3). Operational Impact of Risk Occurrence

The assessment of ***operational*** impact, which is the operational consequence from the risk occurring, in the context of OT&E activity is somewhat different from the assessment of impact in a standard risk assessment.[12] First, operational impacts relate only to performance impacts, not impacts on cost and schedule. Second, some risks can have performance impacts that do not greatly affect the mission, and therefore have low operational impact. For example, a risk could cause a complete loss of a specific operational capability but still have low mission impact because of reasons such as the redundancy within the system of interest or from the presence of another system that is able to support completing the mission. So, operational

[12] Such as defined in the "Risk Management Guide for DoD Acquisition" (http://www.dau.mil/pubs/gdbks/docs/RMG%206Ed%20Aug06.pdf).

impact involves an understanding of performance impacts of risks plus an assessment of the operational/mission relevance of those impacts.

The operational impact question is: If this risk occurs and has potential performance impact on the capability, will that performance impact undermine mission goals?

In order to determine operational impact, the risk assessment must start with a description of performance impacts. Software capabilities can fail (hang or crash a system), but they can also store or transmit incorrect data, emit confusing messages and graphics, add new avenues of unauthorized access, slow system response, hinder training, and so on. The risk assessment must provide the specifics of how a realized risk would unfold.

The risk categories in step (1) above are organized by similar performance outcomes:

- Technology and Software Development Risks: The performance impacts of these risks will tend to be computer crashes, data errors, and excess resource usage on desktops, servers, and local networks.

- Integration and Deployment: The performance impacts of these risks tend to be problems in data transmission and inter-operability. Deployment risks also include one-way data or computer transformations that cannot be rolled back. Irreversible changes must be tested at OT&E level II or III.

- Training, Utilization, and Management: The performance impacts of these risks tend to be problems in user acceptance, excessive training requirements, software administration, and data maintenance.

- Information Assurance: These are risks related to security exposure, intrusion detection, and ability to respond.

After assessing how the anticipated risks to a capability might unfold as performance impacts, the OTA must determine how the impacts on performance of capabilities will translate into operational impacts of those capabilities on the mission goals. Mission goals are expressed in the measures of operational effectiveness, operational suitability, and survivability/security, and the COIs stated in the TEMP. The OTAs should use their preferred risk assessment approaches to provide an assessment of operational impact for each risk/operational capability combination. Table 2 is an example of a three point scale for operational impacts.

Table 2. Operational/Mission Impact Classification

Mission Impact Level	Descriptor	Issue Definition
1	Minimal	Annoying system characteristic or nuisance that does not degrade operational/mission performance or suitability. Little to no impact on resolving COIs.
2	Moderate	Issue degrades operational/mission performance or suitability, and no acceptable operator compensation/workarounds exists. Issue prevents operational/mission performance, but can be overcome with operator compensation/workaround. System COIs are moderately dependent upon increment performance.
3	Severe or Catastrophic	Issue that prevents operational/mission performance cannot meet mission objectives or suitability threshold, no workarounds available. The element is required for mission success, and its malfunction could cause significant damage to the installed system, to other interconnected systems, or to personnel. System COIs are critically dependent upon increment performance.

b. Determination of level of operational test required: On completion of the risk analysis effort, the levels of OT&E for each risk can be determined from the matrix below. The level of OT&E for each capability is the maximum of the OT&E levels determined for each of the risks to the capability.

Likelihood of Risk Occurrence

3	II	III	III
2	I	II	III
1	I	I	II
	1	2	3

Operational/Mission Impact of Risk

Legend (see paragraph 2b for explanation):
I = Level I OT&E
II = Level II OT&E
III = Level III OT&E

c. Obtain approvals. Once the risk assessment is complete, if an Applicable Program, the OTA will provide DOT&E with the risk assessment (likelihoods of occurrence and mission impacts) and the corresponding proposed level of OT&E for approval.

4. **POINT OF CONTACT.** If additional clarification is needed, contact DOT&E's Deputy Director for Net-Centric and Space Systems.