Air Force Test and Evaluation Guide

SUMMARY OF CHANGES:

- General reformatting, language revisions, terms/titles/organizations updated
- Updated Cyber Survivability test guidance
- Amended and expanded Agile Software Test guidance
- Clarified ITT co-chair members

PURPOSE: This document contains guidance, best practices, and lessons learned relevant to Air Force test and evaluation (T&E). It is a "living" guide that reflects information compiled from many sources based on most recent collective test community experiences. Consequently, updates will occur more frequently than with official policy. Although this guide supplements and expands on HQ USAF/TE policy, is not directive nor prescriptive in nature. It mirrors AFI 99-103's broad applicability across multiple areas and complements policies and best practices encompassing a range of test communities and functions. Every effort is made to ensure this guide content is aligned with USAF and OSD guidance; however, if inconsistencies are discovered please contact AF/TEP so that these are resolved.

INTRODUCTION: Just as the warfighter plans by working back from the target or desired end state, a tester must know the requirements, CONOPS, acquisition strategy and available resources to devise an effective T&E strategy. T&E aims to mature system designs, manage risks, identify and help resolve deficiencies as early as possible, and ensure systems are operationally mission capable (i.e. effective, suitable, survivable, and safe). The Air Force T&E community must plan and execute integrated testing as an efficient continuum in collaboration with the requirements, acquisition, and user communities. A tester's role is vital throughout the program's life cycle by informing a Program Manager (PM) and end user on the projected success or failure of the system under test. A continuous flow of credible T&E data about the development and continued sustainment of combat systems is needed to keep systems and warfighters ready for current and emerging combat needs.

Table of Contents

CHAPTER 1: DOD-LEVEL TESTING ROLES AND RESPONSIBILITIES

Director, Operational Test and Evaluation (DOT&E). Given DOT&E's central role in the approval cycle of test execution, testers should familiarize themselves with how DOT&E interacts with their program. The Director is appointed by the President, confirmed by the Senate, and reports directly to SECDEF and Congress. The Director is the principle advisor to the SECDEF and USD (AT&L) on OT&E and the principal OT&E official within the DoD. Congress made the Director responsible for live fire test and evaluation (LFT&E) in 1994.

DOT&E Responsibilities. Their primary responsibilities are to give final approvals for Test and Evaluation Master Plans (TEMPs) before milestone decision reviews and OT&E plans before those tests may commence. They issue beyond low-rate initial production (B-LRIP) reports to senior DoD officials and Congress, and an Annual Report to Congress. They participate in Defense Acquisition Executive Summary (DAES) reports and reviews. Last, DOT&E's staff works closely with the Operational Test Agencies (OTAs), Integrated Test Teams (ITTs), and test teams to ensure better OT&E planning via early involvement.

DOT&E Interest. DOT&E's primary interest is to ensure OT&E and LFT&E are adequate before full rate production (FRP) or deployment, and that the T&E was properly executed according to statute and DoD policy. DOT&E makes a determination of the operational effectiveness and suitability of these systems prior to FRP. DOT&E also looks at the operational mission impacts of deploying a system. They look at the system under test from an "end-to-end" perspective with all the interoperability and supportability considerations thrown into the mix. Operational assessments of mission impacts should look at these considerations early in order to address the issues, questions, and problems that will be raised later with deployment and employment. DOT&E is very sensitive to system contractor involvement in operational tests in any way that could compromise the integrity of the T&E data or the analysis. Any involvement from a system contractor must be limited to instrumentation, data collection and data processing or be representative of wartime concept of operations (CONOPS). Early involvement of DOT&E personnel in drafting the T&E strategy, TEMP, and operational test plans saves time and trouble in the long run. DOT&E personnel must get these issues identified early so they can be more effectively dealt with.

TEMPs and test plans that were developed with early DOT&E action officer inputs stand a much better chance of smooth OSD approval. Early involvement helps prevent unknown issues from surprising and potentially derailing the program during final coordination.

DOT&E approves LFT&E strategies and management plans and alternative LFT&E strategies, if they are developed to support full-up system-level LFT waivers. With consultation from the PM, they determine the number of production or production representative test articles required for LFT&E and IOT&E of programs on the OSD T&E oversight list. Overall, DOT&E provides independent oversight, independent evaluation, and objective reporting of the results of operational test and LFT&E.

DOT&E Guidance. DOT&E is the final approval authority of TEMPs in conjunction with the DD(DTE&P) for programs on oversight. The DOT&E TEMP Guide provides guidance to defense acquisition programs for developing and documenting the program's evaluation strategy and management approach in the TEMP throughout the program's life cycle.

Deputy Director, Developmental Test, Evaluation and Prototyping (DD [DTE&P]). Under the authority, direction, and control of the USD(R&E) and in accordance with Section 838 of the NDAA for FY 2018, the DD(DTE&P) has overall responsibility for conducting the MS B and MS C DT&E sufficiency assessments for Major Defense Acquisition Programs (MDAPs) for which the USD (A&S) is the Milestone Decision Authority (MDA). Additionally, the DD(DTE&P) reports the DT&E sufficiency assessment determinations to the USD(A&S).

DT&E Responsibilities. DT&E supports program offices and the DoD T&E community early and throughout the life cycle of a program to ensure test planning and resourcing are adequate to discover system deficiencies, support system development, and evaluate system performance. DT&E works to improve DoD developmental test and evaluation practice by focusing on support to Programs and Program Offices, improving DT&E Policy & Oversight, advancing DT&E Capability and improving Service DT&E Competencies.

DT&E Guidance. DT&E provides guidance for the planning, execution, and reporting of DT&E in the DoD, as well as the integration of developmental and operational tests in coordination with the DOT&E. DT&E published a T&E Management Guide in December 2012 that defines all the required activities for DoD testing. See Test & Evaluation Management Guide.

Oversight Program List. DOT&E publishes an annual OSD DOT&E oversight list which addresses OT&E and LFT&E. According to DoDI 5000.02, USD (AT&L) has a designated special interest list and OSD (DT&E) has an engagement list. DD(DTE&P) uses the MDAP, MAIS, and AT&L designated special interest lists to identify programs for DT&E oversight. Access to the USD (AT&L) designated special interest list requires a Defense Acquisition Management Information Retrieval (DAMIR) account (DoD CAC required). "Oversight" programs typically require additional briefings and reports, supporting documentation, and increased test rigor. Title 10 lays out the requirement for DOT&E approval of operational test plans prior to conducting operational testing of all MDAPs per §2430 or if so designated by DOT&E and DD(DTE&P). DoDI 5000.02, Enclosure 5 paragraph 3 spells out this designation process by stating that all programs "on the OSD T&E Oversight List" are subject to such approval. This same paragraph also points out that approval is required before conducting operational testing whether such testing occurs prior to, or after fielding and full-rate production decisions. AFI 99-103 defines the various types of Air Force led operational testing all of which fit within OSD's understanding of operational testing and that may be subject to oversight and approval of operational test plans and reports. Each year AF/TE will recommend changes to the T&E Oversight List, and will recommend deletions once oversight serves no further purpose. All test organizations are encouraged to forward any change requests to AF/TEP for forwarding to OSD. See the Defense Acquisition Guidebook (DAG) for additional details.

What goes on the OSD T&E Oversight List? DOT&E chooses programs based on congressional or high level OSD interest. Programs acquired jointly (i.e., Joint or multi-Service), having a high dollar expenditure (i.e., ACAT I), or posing special risks are put on the list annually. The following criteria are listed in the DAG although any program can be placed on OSD T&E oversight at any time.

OSD criteria for determining whether or not a program should be on formal T&E oversight includes:

a. Acquisition category level

b. Potential for becoming an acquisition program (such as an Advanced Concept Technology Demonstration project or pre-MDAP)

c. Stage of development or production

d. Whether program is subject to DAES reporting

e. Congressional and DoD interest

f. Programmatic risk (cost, schedule, performance)

g. Past history of the developmental command with other programs

h. Relationship with other systems as part of a system-of-systems

i. Technical complexity of system

If there is a question pertaining to program oversight status, query DOT&E (https://osd.deps.mil/org/dote-extranet/SitePages/Home.aspx) and/or reference the DAG. Chapter 8 in the DAG provides guidance on developing a test strategy, information on the different DoD and Service test agencies, test resources and ranges, and required documentation including detailed TEMP guidance. The DAG also describes roles and responsibilities of test personnel such as the Chief Developmental Tester (CDT), KLP, and test organizations including the Integrated Test Team (ITT) and Lead Developmental Test and Evaluation Organization (LDTO).

Interagency Cooperation. Many other agencies must be considered for ITTs and Working-level Integrated Product Teams (WIPT). It is important to understand where each office/agency is coming from and how their specific area of concern affects your test activities. Where you are in your program's life cycle can greatly change who cares about the test activities you are trying to accomplish at any given moment. Air Force Operational Test & Evaluation Center (AFOTEC) may not be involved in the early design of testing for a DT&E test plan; however, an office like Joint Interoperability Test Command (JITC) may be heavily involved depending on your system. Make sure you know all the players on the field.

CHAPTER 2:
AF T&E ORGANIZATION AND WORKFORCE

T&E Organizational Structure. While programs should strive to combine DT&E, LFT&E, and OT&E, much of it is organizationally separated. Figure 2.1 shows the decentralized nature of T&E organizations. It is our job as testers to facilitate a seamless transition from one stage of testing to the next in order to engender an integrated test continuum throughout the life cycle of the program.

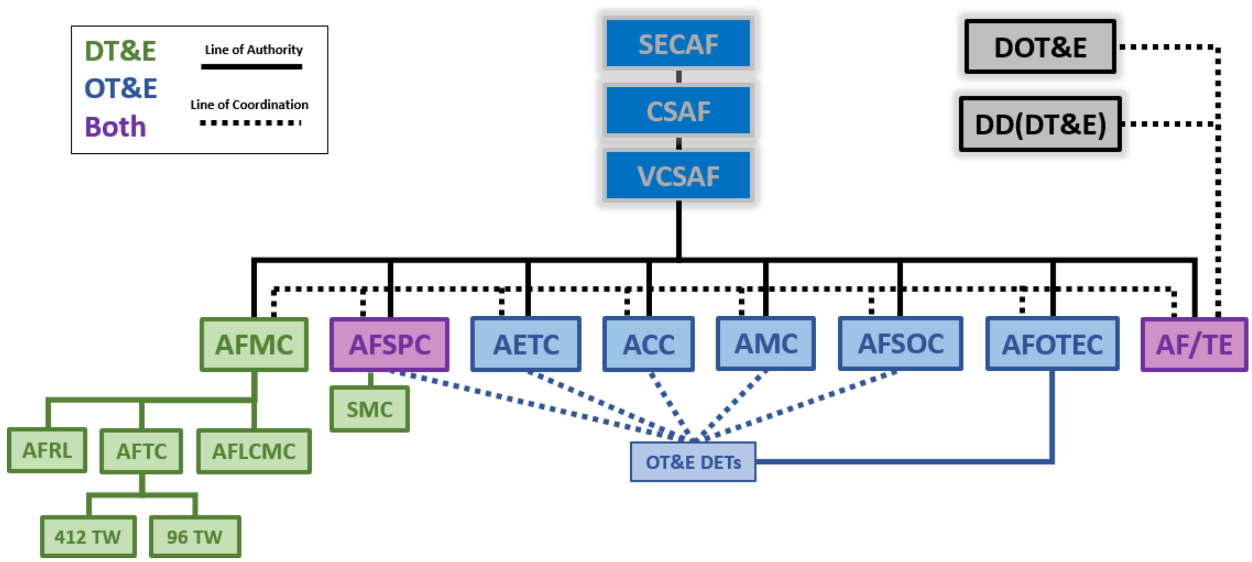


Figure 2.1 AF T&E Organization Structure

Developmental Test Units. Other than Space Systems, DT&E is primarily conducted by Air Force Materiel Command (AFMC). In 2012, AFMC re-organized to consolidate Developmental Test under the Air Force Test Center (AFTC).

Air Force Life Cycle Management Center (AFLCMC), the acquisition arm, provides the construct for the management of all weapon systems from cradle to grave.

Air Force Sustainment Center (AFSC) executes all maintenance and supply chain activities.

a. The Air Force Research Laboratory (AFRL) and Air Force Nuclear Weapons Center (AFNWC) have remained largely unaffected aside from a few minor internal efficiencies and realignments.

b. Hybrid unit Big Safari conducts DT&E and OT&E in-house (primarily for C2ISR systems).

AFTC includes the 412 Test Wing (TW) at Edwards AFB, Arnold Engineering Developmental Complex (AEDC), the 96 TW at Eglin and focuses on DT&E efforts across the enterprise. Units from AFTC make up a large portion of the LDTOs within the AF.

a. 412th Test Wing: flight and ground test of aircraft, weapons systems, software, and Modeling and Simulation (M&S) for the AF. The 412 TW is comprised of the 412 OG, 412 ENG and the Range Group.

b. 96th Test Wing: DT&E AF weapons systems, air-launched weapons, C4I, guidance systems. The 96 TW is comprised of the 96 OG, 96 TG, SEEK EAGLE, and the Range Group.

c. AEDC: 43 aerodynamic and propulsion wind tunnels, rocket and turbine engine test cells, space environmental chambers, arc heaters, ballistic ranges; comprised of the 704 TG and AEDC TG.

Operational Test Units. AFOTEC is the USAF's Operational Test Agency (OTA). It is a Direct Reporting Unit (DRU) that reports directly to the CSAF to provide leadership and procedural guidance for OT&E. AFOTEC is broken up into the following five detachments to address specific mission areas:

- Detachment 1 (Edwards AFB, CA). Lead agency for accomplishing IOT&E of the F-35 Lightning II for the Joint Operational Test Team (JOTT), Royal Air Force, Royal Australian Air Force, and the Royal Netherlands Air Force. Contact Information: (661) 275-2120.

- Detachment 2 (Eglin AFB, FL). To evaluate operational system(s) mission capability, effectiveness, and suitability for Air Force and multiservice users by conducting impartial and realistic operational evaluations and assessments. Focuses on a broad range of assets spanning munitions to command and control. Contact Information: (850) 883-1089.

- Detachment 4 (Peterson AFB, CO). Operationally tests space, missile, and missile defense capabilities in the battlespace environment for the warfighter. Contact Information: (719) 556-5850.

- Detachment 5 (Edwards AFB, CA). Ensures warfighters and logisticians have the right tools for the job, permitting them to effectively and safely accomplish their mission. Focus areas: Bomber and UAV operational testing. Contact Information: (661) 277-3666.

- Detachment 6 (Nellis AFB, NV). To plan and conduct realistic, objective, and impartial operational test and evaluation of fighter aircraft (including F-35 Block 4 IOT&E). Contact Information: (702) 652-4325.

AFOTEC is not the USAF's sole Operational Test Organization (OTO). MAJCOM test squadrons also conduct OT&E that is responsive to their respective MAJCOM priorities and mission needs. They retain strong OT&E culture at operational MAJCOM levels, offer lower cost and faster response times, and typically accomplish a greater total number of operational tests. Some examples of OTOs are:

a. 53d Wing at Eglin AFB, FL

b. 505th Command and Control Wing, Hurlburt Field, FL

c. 17th Test Squadron at Schriever AFB, CO (space test)

d. 47 Cyberspace Test Squadron at Joint Base San Antonio-Lackland, TX cyber testing a broad range of systems

e. ANG –AFR ANG, AF Reserve Test Center (AATC) at Tucson, AZ: AATC conducts operational test (OT) on behalf of each MAJCOM

Test and Evaluation Workforce Development.

Cross-Functional Authority (XFA). The CSAF appointed AF/TE as the XFA for the Test and Evaluation Enterprise on 11 Oct 16. Per AFI 36-2640, Cross Functional Authority means: "Responsible for strategic oversight and force development advocacy related to the requirements of their occupational capability…"

a. The designated XFA has valid force development requirements across various Air Force specialties.

b. The XFA can generate, implement, and manage force development strategies.

c. The XFA is NOT authorized to establish its own development teams or new career fields.

AF/TEP established an XFA working group to address management; i.e. tracking and utilization of the AF TE workforce.

Test and Evaluation Functional Authority for Acquisition Professional Development Program (APDP).

a. AF/TEP will continue to serve as the Functional Authority for APDP T&E

issues. If you do have an urgent APDP matter, please email the AF/TEP workflow usaf.pentagon.af-te.mbx.af-tep-workflow@mail.mil or call (703) 697-0252/DSN 227- 0252.

CHAPTER 3:
TYPES OF TEST

Test type overview. A program will go through multiple types of testing but not all systems will follow the same path. Chapter 3 in AFI 99-103 provides an overview of the major types of testing (Table 3.1) so this guide will only address specific areas that need further clarification. Contact AF/TEP for specific guidance to get your programs through the T&E approval process.

Type of Testing	Description	References
Advanced Technology Demonstration (ATD) (*Note 1*)	Air Force Research Laboratory-funded, MAJCOM-sponsored development efforts that demonstrate the maturity and potential of advanced technologies for enhancing military operational capabilities.	AFI 61-101, *Management of Science and Technology*
Technical Assurance Standards Testing	Evaluates offensive cyberspace operations capabilities against technical assurance standards.	DoDI O-3600.03, *Technical Assurance Standard (TAS) for Computer Network Attack (CNA) Capabilities*
Electronic Warfare Integrated Reprogramming (EWIR)	Process intended to produce and deliver software/hardware changes to electronic equipment used to provide awareness and response capability within the EM spectrum. May require changes in TTP, equipment employment guidance, aircrew training and training devices (threat simulators and emitters). Provides guidance for test / fielding of mission data (MD) changes, OFP changes, or minor hardware changes that comply with the guidance in AFI 63-131 concerning modifications.	AFI 10-703, *Electronic Warfare (EW) Integrated Reprogramming*
Emission Security (EMSEC) Assessment	Assesses against the requirement to control the compromise of classified electronic emissions.	AFSSI 7700, *Emissions Security*, AFSSI 7702, *EMSEC Countermeasures Reviews*
Foreign Comparative Testing (FCT) (*Note 1*)	FCT is an OSD-sponsored program for T&E of foreign nations' systems, equipment, and technologies to determine their potential to satisfy validated United States operational requirements.	10 U.S.C. § 2350a(g) OSD Comparative Technology Office Handbook (https://cto.acqcenter.com/)
Joint Capability Technology Demonstrations (JCTD) (*Note 1*)	Exploits maturing technologies to solve important military problems and to concurrently develop the associated CONOPS to permit the technologies to be fully exploited. Emphasis is on tech assessment and integration rather than development.	DoDI 5000.02, *Operation of the Defense Acquisition System* AFI 63-101/20-101, *Integrated Life Cycle Management*
Joint Interoperability Test and Certification	Required certification for net-readiness prior to a system being placed into operation. Must be preceded by Air Force System Interoperability Testing (AFSIT), formal service-level testing to determine the degree to which AF systems which employ tactical data links conform to appropriate DoD MIL-STDs.	CJCSI 5123.01G, *Charter of the Joint Requirements Oversight Council (JROC)* DoDI 8330.01, *Interoperability of Information Technology (IT) and National Security Systems*
Joint Test & Evaluation (JT&E) (*Note 1*)	Evaluates non-materiel capabilities and potential options for increasing joint military effectiveness. Focus is on evaluating current equipment, organizations, threats, and doctrine in realistic environments. JT&E projects are not acquisition programs.	DoDI 5010.41, *Joint Test and Evaluation (JT&E) Program* AFI 99-106, *Joint Test and Evaluation Program*
Testing of Urgent Needs *(Note 1)*	Quick reaction capability for satisfying near-term urgent warfighter needs.	DoDI 5000.02, *Operation of the Defense Acquisition System*

Unified Capabilities (UC) Certification	Certifies interoperability and information assurance for Unified Capabilities (defined as integration of voice, video, and/or data services delivered ubiquitously across a secure and highly available network infrastructure, independent of technology). AFSPC appoints the Air Force UC test organization responsible for testing technologies meeting the definition.	DoDI 8100.04, *DoD Unified Capabilities* AFMAN 33-145, *Collaboration Services and Voice Systems Management*
Note 1. Activity falls outside the traditional acquisition process; however, Air Force testers may be required to support the activity by providing T&E expertise in assessing the military utility of new technologies.		

Table 3.1 Types of Testing

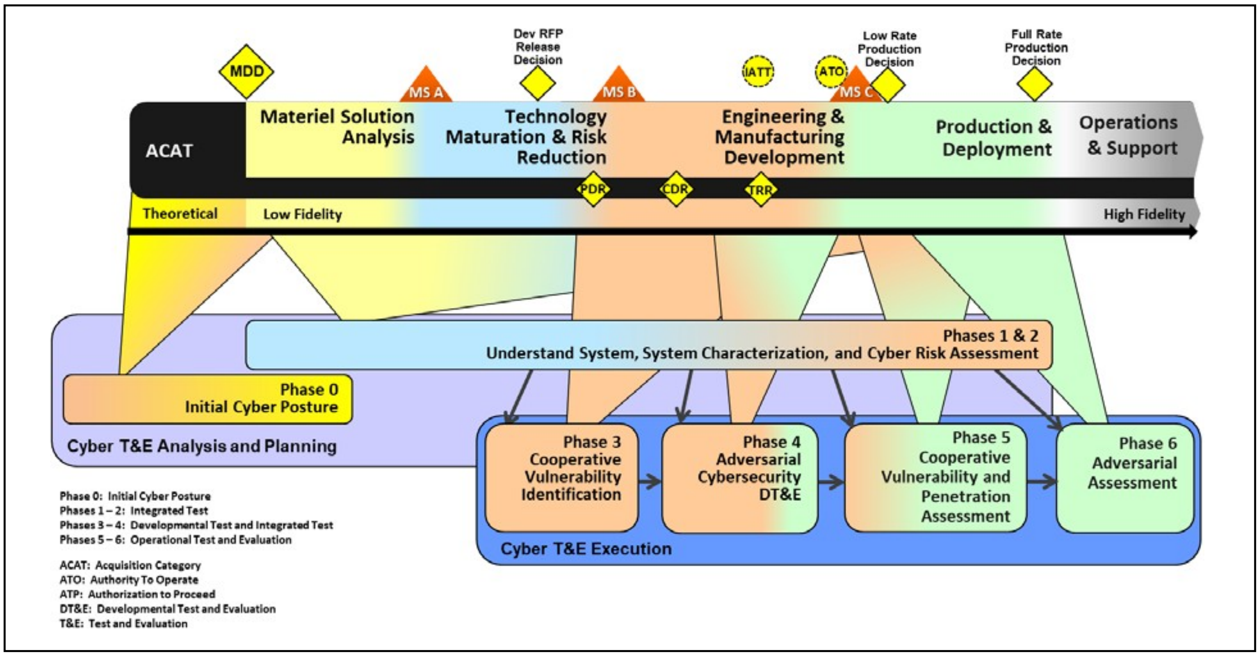
CHAPTER 4: CYBERSECURITY/CYBER SURVIVABILITY TEST

Cyber test evaluates and characterizes systems and sub-systems operating in the cyberspace domain, and the access pathways of such systems. Cyber survivability (which includes cybersecurity and cyber resiliency) is a key factor in a system under test (SUT) operational effectiveness, operational suitability, and lethality and may drive a need for updated Tactics, Techniques, and Procedures (TTP). Evaluations must include the elements of susceptibility, vulnerability, and recoverability of SUTs when exposed to cyber or kinetic threat weapons. The Critical Operational Issue (COIs) include any issues that need to be addressed in the OT&E evaluation of the system's cyber survivability. The CDT works with select SMEs to create a Decision Support Key (e.g., the Developmental Evaluation Framework Table included in Chapter 3 of the TEMP) that describes the key program decisions, the cybersecurity and cyber resiliency information that is required to inform those decisions, and the CT&E events that will be accomplished to gain the needed information.

Cyber Test Phases. As currently outlined in the DoD Cybersecurity Test and Evaluation Guidebook, cyber testing is comprised of six distinct Cyber Test and Evaluation phases:

- Phase 1: Understand the System
- Phase 2: System Characterization and Risk Assessment
- Phase 3: Cooperative Vulnerability Identification
- Phase 4: Adversarial Cybersecurity Developmental Test and Evaluation
- Phase 5: Cooperative Vulnerability
- Phase 6: Penetration Assessment; and Adversarial Assessment

A preceding Phase 0, Initial Cyber Posture, is being added by the Air Force to incorporate cyber survivability considerations into the Analysis of Alternatives (AoA) process. Each phase builds upon previous activities, providing invaluable information a program office can use to improve the cybersecurity and cyber resiliency of its system. Early cybersecurity test involvement maximizes these opportunities. Figure 4.1 depicts the seven Cyber Test and Evaluation Phases as they align with a standard Acquisition Category (ACAT) program timeline. Gray diamonds depict anticipated Mission-based Risk Assessment Process for Cyber (MRAP-C) activities. Actual alignment of the cybersecurity T&E phases is tailorable to accommodate the unique structure and requirements of each program but regardless of where a program enters after AoA (e.g., Milestone A, Milestone B), the cyber test activities must start at Phase 1. Programs that initiate cyber test activities prior to the AoA decision will begin at Phase 0.



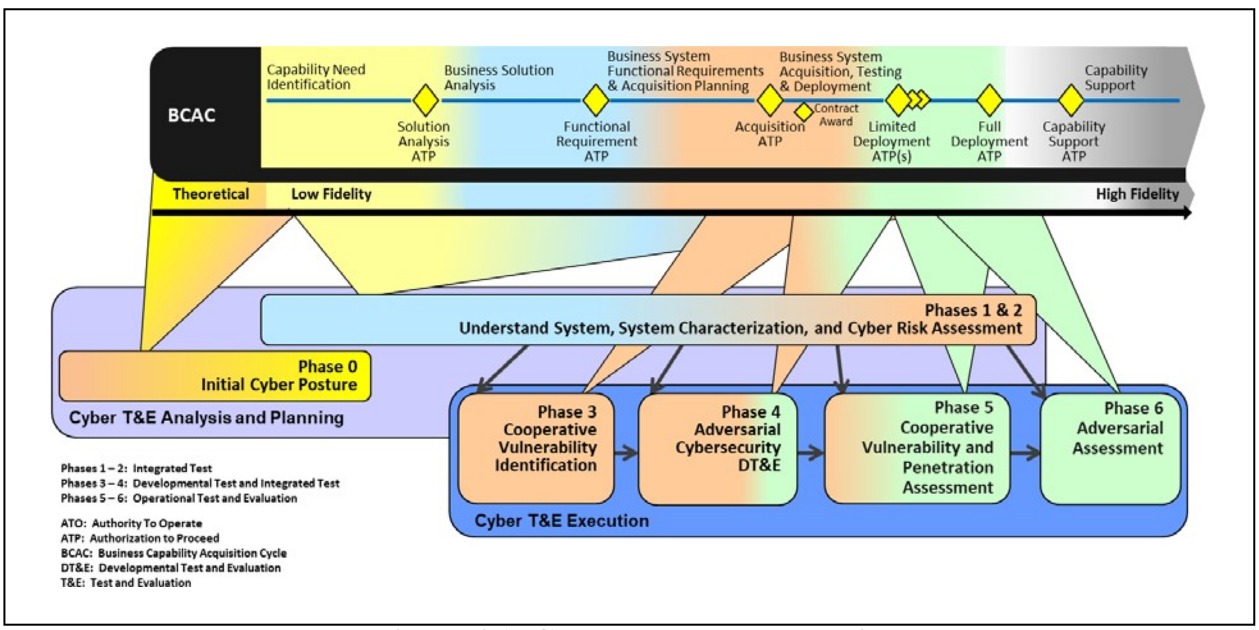


Figure 4.1: Cyber Test and Evaluation Phases

Mission Based Cyber Risk Assessments. Mission Based Cyber Risk Assessments (MBCRA) are a key component to developing a cyber survivable system. Executed by an integrated team of personnel from the various stakeholders and functional areas, MBCRAs are performed iteratively throughout the acquisition lifecycle to inform:

- Milestone A/B/C
- Authorizations to Proceed
- Requirements Identification
- Request for Proposal Generation
- Preliminary Design Review
- Critical Design Review
- Test Strategy, Design, and Execution
- Engineering Changes
- Deployment Decisions
- Test Resources
- Authorization to Operate
- Other engineering, programmatic, risk, and test decisions

The MBCRA process helps identify, assess, and prioritize cyber-risks based on potential impacts to the system's capability to perform its operational mission(s). The MBCRA process begins with a thorough review of available system requirements, system documentation (including CONOPS/CONEMP, architecture documents, design documents, previous test results), and current threat information. Detailed technical, mission, and intelligence analyses are then used to understand and characterize the system, map dependencies, identify potential attack vectors and attack paths, and identify cyber risks. The technical analysis focuses on understanding the system's architecture, interconnections, data flows, and external interfaces. The mission analysis focuses on the potential impact an exploitation of each vulnerability could have on the systems' capability to perform its mission. Finally, the intelligence analysis addresses the likelihood that an adversary has the means and desire to exploit a given vulnerability. Cyber risk is derived by combining the potential mission impact and the exploitation likelihood. The MBCRA results are then used to inform requirements, engineering/design, programmatic, risk, and test decisions. The final iteration, performed after Phase 6, is used to support the program in sustainment and during follow-on upgrades.

Cyber test characterizes and evaluates systems and sub-systems operating in the cyberspace domain, and the access pathways of such systems. Cyberspace is defined as a domain characterized by the use of electronics and the electromagnetic spectrum to store, modify, and exchange data via networked systems and associated physical infrastructures. The primary objective of cyber test is to evaluate the system's cyber survivability by testing the system's cybersecurity and cyber resiliency. Cyber test should be integrated throughout system development and testing by both the contractor and government test agencies. Early cyber testing should be incrementally accomplished in the lab and/or developmental test environment(s), eventually progressing to operationally representative cyberspace environments. Developmental Test and Evaluation and Operational Test and Evaluation plans must be developed considering system architecture and all attack surfaces (interfacing and embedded systems, services, and data exchanges that may expose the system to potential cyber threats) through all applicable domains. Refer to the *DoD Cybersecurity Test and Evaluation Guidebook* for additional guidance.

Cybersecurity Testing is the testing of the systems' and sub-systems' ability to protect or defend against a cyber-attack. Cybersecurity testing focuses on identifying and eliminating or mitigating system cyber vulnerabilities. Cybersecurity testing should be conducted by contractors and government testers throughout the development process and is scoped to assess the system's or sub-system's cyber boundary.

Cybersecurity testing utilizes the MBCRA artifacts, Security Assessment Plan, Program Protection Plan, Life Cycle Sustainment Plan, Information Support Plan, and Risk Management Framework artifacts to help assess risk to mission assurance. It assesses the presence and effectiveness of cyber protections and provides the data necessary to the Authorizing Official to render a determination of risk to DoD operations and assets, individuals, other organizations, and the Nation from the operation and use of the system. Cybersecurity testing does not assess the system's or sub-systems' cyber resiliency (i.e., capability to perform its functions/missions when subjected to cyber-attacks).

Cyber resiliency testing evaluates a system's ability to meet operational requirements while subjected to cyber attacks. A cyber attack is defined as an attack, via cyberspace, designed to infiltrate, disrupt, disable, deceive, destroy, or maliciously control a target within cyberspace or a physical system. Cyber resiliency testing focuses on preventing, mitigating, and recovering from a successful cyber attack, and determining mission system and mission effectiveness. Cyber resiliency testing should include the information identified in the System Survivability Key Performance Parameters/Cyber Survivability Endorsement Implementation Guide, including addressing the ten Cyber Survivability Attributes.

The Integrated Test Team and test organizations must plan for appropriate cyber test to assess system vulnerabilities and mission impact. If the Integrated Test Team or test organization cannot comply with cyber test requirements, the Integrated Test Team or test organization must document the limitations and rationale in the Test and Evaluation Master Plan and test plans.

Automated testing is critical to designing and fielding a secure system but it does NOT replace hands-on testing. Automated testing is an exceptional way to conduct baseline cybersecurity testing (checking for the presence and effectiveness of basic controls and protections) but it does not adequately emulate creative pen testers nor test system resiliency. Both automated and hands-on testing are required.

AF Cybersecurity Test Policy. The latest update of AFI 99-103 defines cyber testing as a combination of cybersecurity testing and operational resiliency testing. Much of present TEMP content is focused on cybersecurity testing and certification. Testing of operational resilience is largely encapsulated by “detect, react to, and restore from” cyber attacks or planning for mission continuation in the event restoration is impossible or impractical.

Cybersecurity in TEMPs. The Test and Evaluation Master Plan should explain what cyber testing will be accomplished during each cyber test phase throughout the acquisition lifecycle. This includes describing the planned cybersecurity and cyber resiliency testing and the associated scope, and should follow the process covered by the Joint Cyber Survivability Endorsement Implementation Guide (CSEIG). Explain program capabilities and limitations with regard to test. Create a classified annex if needed. While the entire realm of “cyber” continues to evolve, make sure to address the following elements as part of a comprehensive test strategy during your planning and for your TEMPs:

- Define the cybersecurity strategy for the system under test.

  a. What elements are being incorporated into the system’s design to ensure resiliency during a cyber attack?

  b. What is your plan for managing supply chain threats, permanent and air gap connections, software development and updates, etc.?

- Based on the program strategy, identify the elements that require testing (verify and validate, or V&V).

  a. How will the program V&V that the system under test is survivable in a cyber attack (or against classes of attacks)? This strategy must include formal DT (i.e., measureable, repeatable, etc.) and OT.

  b. Include a plan for how the system will be characterized, what will be tested, when in the lifecycle, the resources required (i.e., where, using what, and whom), and the organizations that will perform the testing. Limitations/risks should also be identified alongside mitigation plans.

- Identify the cyber hygiene strategy for the weapon system.

  a. How will cyber hygiene be verified/tested throughout the weapon system lifecycle?

  b. Include the program plan for identifying vulnerabilities, incorporating the risk management framework (RMF), performing penetration testing, and achieving an authority to operate (ATO) certification. This plan should identify the resources required (ranges, organizations, etc.)

- Cybersecurity Test Integration. Cybersecurity testing has become an integrated part of the test like every other subsystem (propulsion, avionics, weapons, e.g.). Cybersecurity test should be integrated throughout DT and OT, and executed in operationally representative cyberspace environments. All DT and OT documents and plans must be developed the system architecture and all potential attack surfaces (interfacing and embedded systems, services, and data exchanges that may expose the system to potential cyber threats) throughout all applicable domains. See Part III for more information regarding cybersecurity testing within an ASD framework.

Cyber Survivability Endorsement (CSE). The Joint Staff developed the CSE process to improve cybersecurity requirements within development and test documents. The CSEIG helps sponsors articulate cyber survivability requirements with the appropriate detail based on the system under test's software dependency, adversary threat tier, and impact of system compromise. For more information on the CSEIG, visit the Risk Management Framework Knowledge Service and request access using the following link: https://rmfks.osd.mil/login.htm.

CHAPTER 5:
AGILE SOFTWARE TESTING

Overview. Agile Software Development (ASD) is an iterative methodology performed in a highly collaborative manner. In contrast to Waterfall (distinct stages of software development tied to rigid requirements, with each stage finishing before the next one can begin), ASD is characterized by early and continuous cross-functional/stakeholder involvement and is responsive to requirements that may change in priority throughout the system's development. ASD produces successive usable software iterations that gradually build and continuously evolve through successive development cycles. Maintaining development/test velocity and cadence allows ASD to regularly inform specific risk-based decisions while driving down overall risk during the software's development and sustainment (Figure 5.1). For more information regarding general ASD concepts reference https://www.agilealliance.org/agile101/.

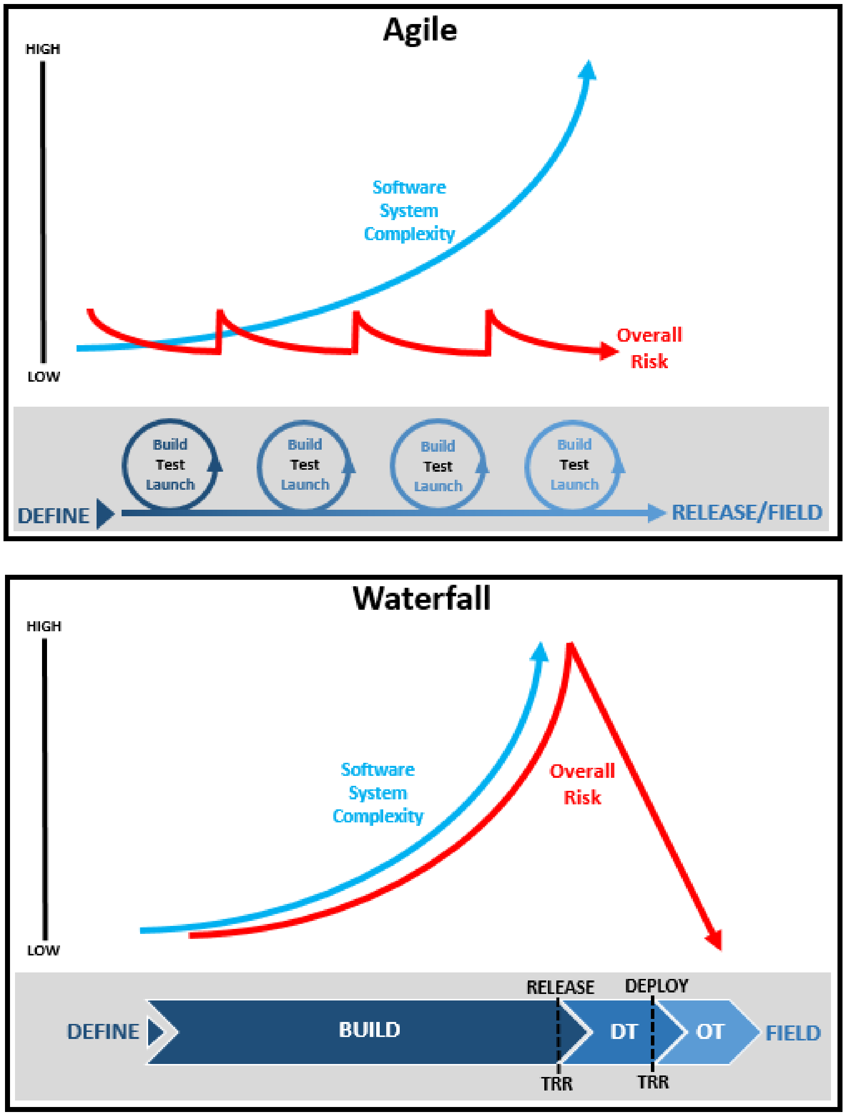


Figure 5.1: Agile Versus Waterfall Risk Profiles

Two constructs within ASD philosophy have emerged as the dominant USAF acquisition approaches:

DevOps: An extension of an Agile methodology where development, test, and operations share responsibility for delivery of functioning services or products (Figure 5.2). Common components in a DevOps ecosystem include configuration management and infrastructure-as-code, virtual or cloud infrastructure, data-rich feedback loops, and process automation that drives continuous integration, packaging, auditing, and monitoring. Reference: DevOps for Federal Acquisition, MITRE, 2015.

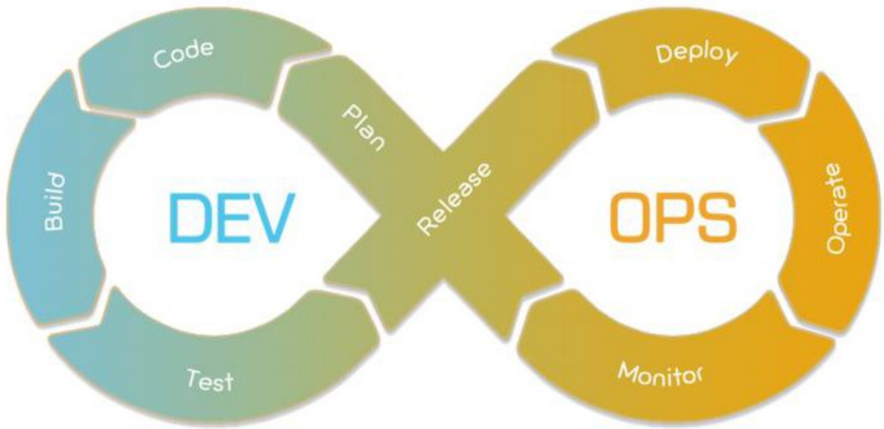


Figure 5.2: DevOps Process

DevSecOps (DSO): DSO wraps cyber survivability into every step of development to safely deliver produced software. DSO strives to automate core cyber survivability tasks by embedding key security policies such as code analysis, compliance monitoring, threat investigation and vulnerabilities assessments into typical DevOps workflows (Figure 5.3). DSO is currently the most desirable software development approach.

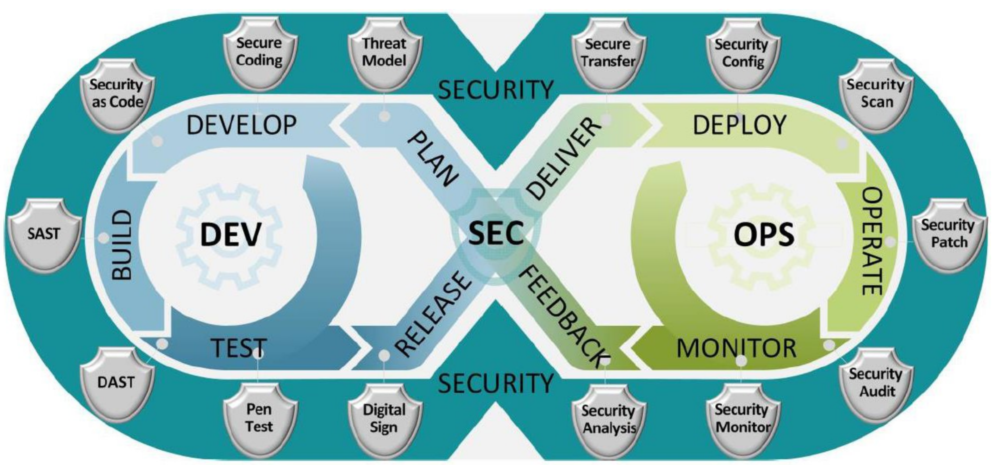


Figure 5.3: DevSecOps Process

Common terms.

Backlog: A list of the new features, changes to existing features, bug fixes, infrastructure changes or other activities that a team may deliver in order to achieve a specific outcome. Product backlog items take a variety of formats, with user stories being the most common. *Reference: Agilealliance.org*

CI/CD Pipeline: The set of tools and the associated process workflows to achieve continuous integration (CI) and continuous delivery (CD) with build, test, security, and release delivery activities, which are steered by a CI/CD orchestrator and automated as much as practice allows. *Reference: DoD Enterprise DevSecOps Reference Design Version 1.0, 2019.*

Deployment: All of the activities that make a software system available for operational use. Whereas a release is transferred from one organization to another (such as a developer to a different organization's test/staging environment), deployed software is accepted by the lead command for use in the actual operational environment. *Reference: Roger S.Pressman Software engineering: a practitioner's approach (eighth edition).*

Done: The cross-functional software development team agrees on a list of criteria which must be met before a user story is considered "done". *Reference: Agilealliance.org*

Integrated product team: A multidisciplinary group of people who are collectively responsible for delivering a defined product or process. *Reference: DoD Integrated Product and Process Development Handbook.*

Iteration: A set period of time during which specific work has to be completed and made ready for review. Also commonly referenced as sprint. *Reference: Agilealliance.org*

Minimum Viable Capability Release (MVCR): The first version of the software that contains sufficient capability to be fielded to an operational user. The MVCR focus is on delivery of a minimum set of features that are relevant to the system under development and represent a deployable release to mission operations. *Reference: OUSD A&S Interim Policy for Operation of the Software Acquisition Pathway 3 January 2020.*

Minimum Viable Product (MVP): Provides users with working software to demonstrate initial capabilities, to accelerate learning, and to shape needs/requirements, designs, and future iterations. *Reference: OUSD A&S Interim Policy for Operation of the Software Acquisition Pathway 3 January 2020.*

Release: A distinct, tested, deployable software element of a militarily useful capability to the government. A release is an increment (version) of the system/software that is transferred from one organization to another. *Reference: Software Engineering Institute Technical Memorandum, 9 June 2015.*

Software Factory: A collection of multiple pipelines, which are equipped with a set of tools, process workflows, scripts, and environments, to produce a set of software deployable artifacts with minimal human intervention. A software factory automates the activities in the develop, build, test, release, and deliver phases. *Reference: DoD Enterprise DevSecOps Reference Design Version 1.0, 2019.*

Test Manager: For non-MDAP or MAIS programs, a Test Manager (TM) can fulfill the functions of a CDT. *Reference: AFI 99-103, Capabilities-Based Test and Evaluation.*

User story: Information or feedback which is expected to yield, once implemented, a contribution to the value of the overall product, irrespective of the order of implementation. Multiple user stories may make up a feature, which is a broader element of functionality. An epic is comprised of numerous user stories/features that cannot be delivered as defined within a single iteration. *Reference: Agilealliance.org*

Agile Software Test Strategy.

Agile Software Test (AST) is highly integrated into the Agile software development and release process. Instead of grading capability at the end of the development phase, AST informs early, often, and combines with user feedback to inform stakeholders while helping evolve requirements that drive future software iteration development. Figure 5.4 provides a simple view of how test integrates into a typical ASD iteration. Test must foster rapid delivery of quality software by being adaptive, responsive, and providing sufficient safeguards to manage risk while ensuring the program achieves intended capability.

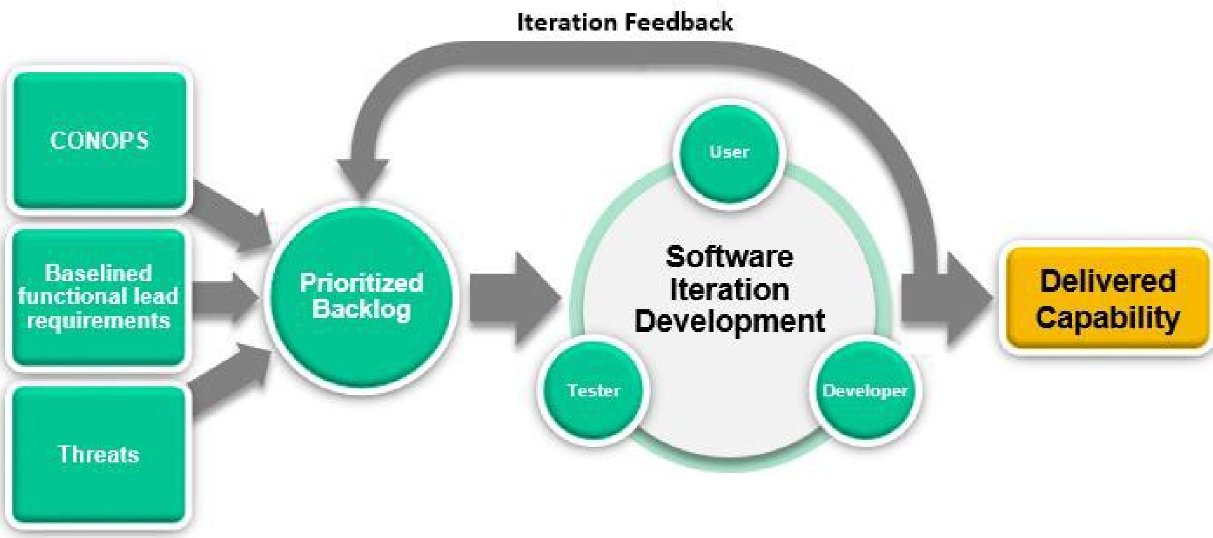


Figure 5.4: Iteration Process With Integrated Tester

An effective AST strategy hinges on effectively integrating across cross-functional integrated product teams (IPT) while maintaining objectivity, testing to traceable requirements, ensuring "done" definitions include passing specific acceptance and test criteria, scaling alongside software growth and smartly leveraging automated tools to efficiently maintain test rigor.

The strategy must take the capabilities (e.g. tool sets) and any limitations (e.g. ability to reside in specific security enclaves) of the host platform in account. An established platform/software factory provides a stable foundation for iterating and delivering code. If legacy software is being transitioned to an Agile methodology, ensure test is adequately engineered to discover any inherent code deficiencies before these migrate into subsequent software iteration cycles.

A fixed cadence/schedule enables more effective integration of independent test resources, particularly with respect to high-demand, low-density assets such as cyber red teams and operational users. Schedules also determine how projects and teams will be composed to support near- and longer-term releases. A fixed, short (typically one to two week) cadence is key to keeping overall risk from accruing and should also support software deployment on user demand. Test must therefore be present in software release planning meetings to help determine the best development cadence.

The test strategy must leverage automated testing to the maximum practical extent. Automated test tools must check for vulnerabilities in code and controls, operational functionality, integration, and perform chaos and regression testing. Properly certified and designed automated test tools enable increased test rigor for the duration of the software development effort. Automation does not completely replace the need for test involving typical users. Furthermore, automated tools must evolve to account for system changes as development progresses, interfaces/dependencies are added and overall software complexity increases. This may be done by regularly folding in manual or exploratory test into automated test code.

The test strategy must address test scalability to keep pace with software growth. Embedding human testers with every IPT will quickly limit how much test can scale with growing product. An effective strategy applies a systems engineering approach to identify the highest risk areas (such as the systems integration level) to gain maximum utility and efficiency from the human test force.

While integrated and tailored test is a key aspect of USAF ASD programs, the test strategy must ensure that embedded testers remain organizationally independent from developers and the program office. The Test Strategy should define the degree of test independence required, the test type, and level of rigor required based on the system's developmental stage. The risk assessment will drive the level of tester independence and determine the test approach.

Programs with lower risk may allow for increased levels of internally-executed test, while higher levels of risk (e.g. interacting with mission/safety critical external systems or databases) should drive increased external testing. PEOs must still designate a Lead Developmental Test and Evaluation Organization (LDTO) or obtain a waiver IAW AFI 99- 103.

AST Planning and Execution.

An Agile-compatible test strategy involves rigorous planning before the contracts are written. AST planning is informed by available requirements and resources, and enabled by informed contracting. Therefore, PMs and TMs must drive test participation left beginning with the contracting stage. Testers must be active participants in program contract writing to establish, at a minimum, test tool selection and initial test resource commitments from the contractor and government (this includes testing of software factory/pipeline security attributes). If agile contracting methods involve frequent contract revisits, test must be a standing member within this concept. Figure 5.5 illustrates the coordination channels between program management, testers, and software developers in order to select the best test tools.

Effective test planning depends on tight collaboration within the cross-functional IPT. IPTs are responsible for delivering certain components of the overall software system. The team composed of, at minimum, the lead command representative or functional lead, system users, software developers and testers. The integrated tester must contribute to and understand the team's definition of "done" for every user story within that team's assigned backlog. IPTs must ensure the user story-defined "done" is compatible with the overlying Feature and Epic that trace back to the acquisition strategy. A key role for the embedded tester is ensuring test is part of the "done" user story.

The tester also actively informs automated test tool tasks, and ensuring collaborative environments are adequate for rapid reporting. As the number of IPTs grow, the TM and PM must allow for the tester(s) to float wherever they are needed most.

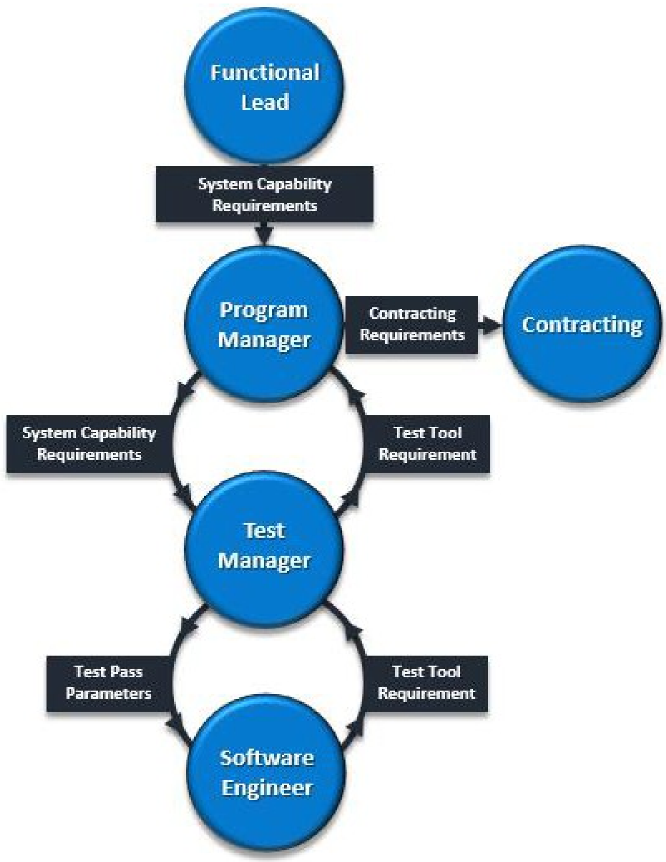


Figure 5.5: Effective Test Tool Selection/Contracting Process

Requirements should be tiered to inform Epics, Features and User Stories. In other words, overarching lead command/user requirements must be deconstructed into smaller, more specific elements that allow themselves to be sorted into backlogs that feed software iteration development cycles. Informed by an acquisition strategy, these requirements may be re- prioritized based on evolving mission needs during periodic reassessments that must include lead command representation. Testers must determine level of test rigor within the release and its iterative cycles based on assessed mission impact/safety/lethality factors. This determination must be informed by the lead command's willingness to accept risk. An example, if a user story or feature is deemed low risk by the lead command/functional lead, the feature can be fielded while sparing resources and saving time. However, if the feature is deemed to be medium or high risk, additional government test resources will be brought to bear to ensure the feature is adequately characterized and evaluated prior to fielding. Any manual testing done during this effort should be transferred to automated software factory pipelines to eliminate future redundant testing. Figure 5.6 provides a basic overview of this concept.

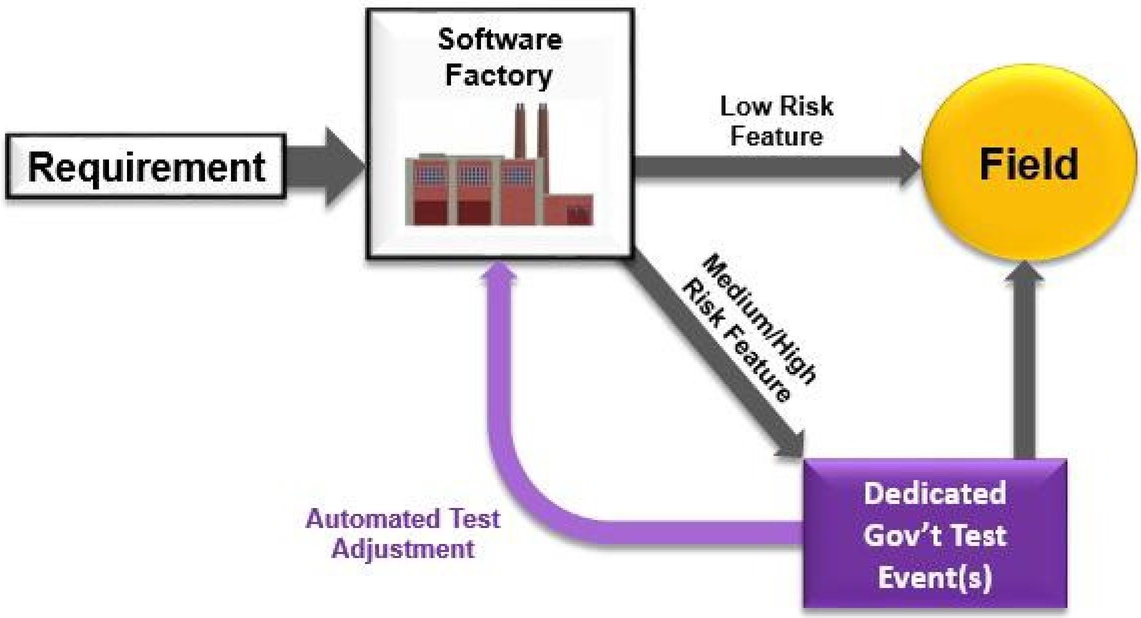


Figure 5.6: Low Versus Medium/High Risk Feature Fielding Decision

AST Approach Selection. Risk (programmatic, safety, lethality, etc.) analysis conducted throughout the program will drive test approach selection. Test approaches are tailorable based on program needs, flexible based on software system progress, adaptive to iterations produced by developmental cycles, and anticipate external test resources alignments. By accomplishing the following steps, TMs and PMs will posture the IPTs to maintain development velocity while maintaining appropriate test rigor.

Identify critical areas to test which are not captured in software developer's test-driven design process. Although proactive test planning should minimize this possibility, ASD pacing and fluidity could result in unguided, undocumented exploratory test. The fluid nature of ASD highlights a need to continuously capture the test cases and outcomes, compare those against current and subsequent releases, and continually re-evaluate developmental risk and priority for test objectives.

Determine DT/OT weight of effort to best match resources according to the assessed capability/safety/lethality/impact of the functional code being developed. It should be noted that there is a fundamental limit to when OT cyber data can be generated: The system needs to be in an operational configuration with operationally realistic inputs and outputs, and operational users. If the system is not in this state, there is typically very limited data to be generated that is relevant for answering operational questions.

Determine DT/OT integration and balance to best match resources according to the assessed impact of code being developed per release cycle. Cyber survivability testing experts must ensure the software factory security posture and CI/CD pipeline incorporates all required security features and stays updated; otherwise, factor in when spot checks (to establish or continue “certification" of an automated security scanning pipeline) should occur, and verify the results are being addressed along with a periodic penetration testing.

Determine level of contractor testing and their resources to conduct test. An Operational Test Agency (OTA) or organization (OTO) will determine the operational relevance of contractor-furnished environments to collect valid test data. Furthermore, determine whether developers can effectively test if government data/environment restrictions are in place.

Determine test frequency, particularly when defining periodic independent testing and risk thresholds (edges of the envelope) that require independent “looks.” It is important to balance frequency (and its resulting impact on resources) with the ability to inform development and satisfy user demand.

Determine the requirements for environments and/or platforms to be operationally representative, or at least as closely representative as possible. This facilitates more comprehensive automated testing and allows OT to be integrated right from the start. Cloud environments or multiple software integration labs (SILs) prevent progress bottlenecks as one can be used for development while the other can be leveraged for training. Collocated SILs are best, but if not possible then research connectivity issues to be resolved prior to software development commencement. Furthermore, if external ranges are involved, determine their capabilities and limitations as well (which may include latency limitations). Keep tester access in mind while planning supporting ranges or other test environments.

Building the AST Road Map. The ITT must help the PM build an adaptive road map that captures what, when, and how to test throughout the program’s software development.

Start with the bigger picture. Deconstruct overall CONOPS and program requirements and prioritize resulting work across the software release plan as much as possible. Determine what constitutes an MVP and MVCR (or program-analogous terms), and what testing must be accomplished in order to reach these stages of software maturity.

Test resources and limitations should be identified and accounted for to the maximum extent.

Using this strategic-level data, the ITT co-chairs (working with the PM, see Attachment 5) should further refine the software release plan and schedule. Test involvement during release planning provides a solid foundation for synchronizing test that supports the release, providing a better framework for lining up applicable test resources.

Using the release plan road map, IPT determine such things as software iteration timing, team collaboration details, and feedback mechanisms linking sequential software development iterations that build up to a release. The ITT watches for any test resource conflicts or synchronization issues between multiple software release trains and coordinates with the PM/functional lead to resolve these issues or reformulate the plan.

Testers coordinate within their IPTs to optimize test during each software development iteration to maintain release velocity but also characterizes software and informs decision makers appropriately. Figure 5.7 illustrates the recommended flow of events within a software development iteration (this can be applied from release to iteration levels).

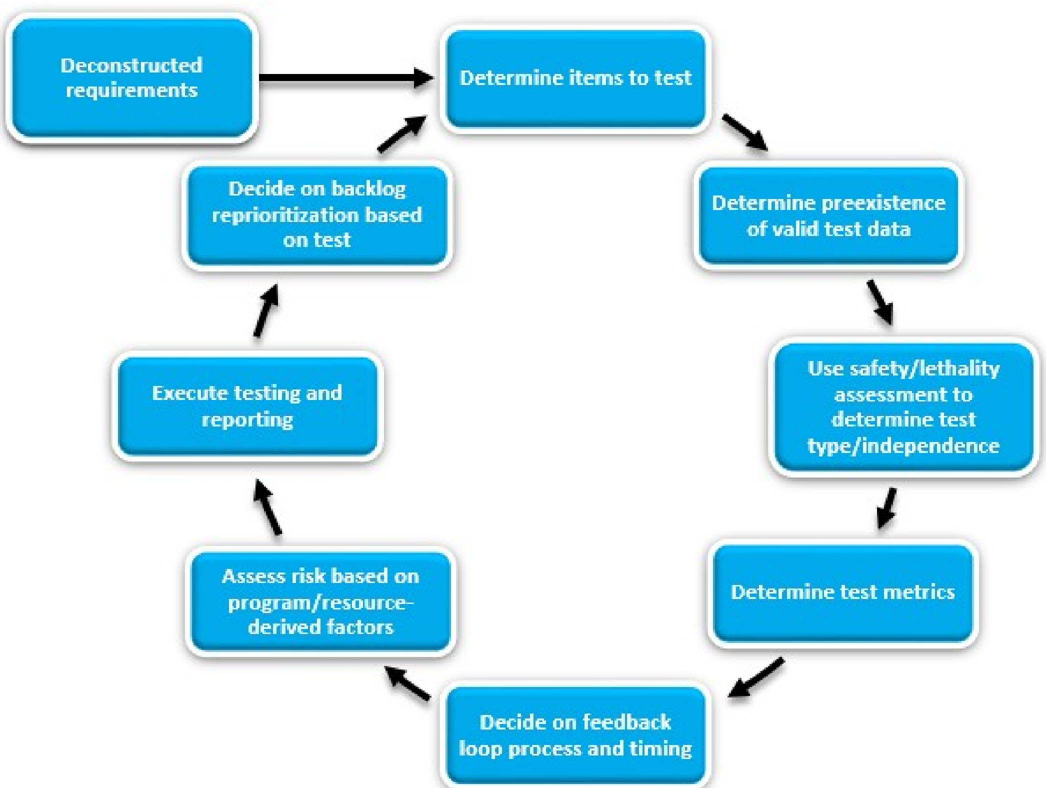


Figure 5.7: Software Iteration Test-Driven Events

Risk Analysis. Considered factors such as system complexity (to include number of subsystems) and resulting integration aspects, impact on external/unrelated systems, impact on mission critical and/or safety aspects, and enabled capabilities. Additionally, consider the evolving scope and complexity of the release(s) to be fielded. For example, an MVP may be relatively simple but eventually grows more complex. The MVCR will be more complex and should undergo more test rigor if it is designed to sunset a legacy system. Regardless, as system complexity increases, the risk of undesired/failed system interactions increases requiring a more detailed understanding of the systems architecture and operational flow. Furthermore, test planners must know capabilities and limitations of internal and external test resources, to include scheduling independent testers and security issues spanning both systems and personnel.

Independent DT and OT. Judiciously embed independent testers and maximize automated test scripts when synchronizing test with ASD. Testers are integrated either physically or remotely with the development team and user, providing continuous feedback and diminishing the risk that serious issues will not be discovered before they require extensive rework.

Execute DT or OT based on defined thresholds. If resource constraints do not allow for a continuous independent integrated DT/OT, selectively choose which type of test best fits each release cycle or proposed transition to the warfighter. These selections should consider overall risk to the iteration, release, or program as a whole.

Dedicated OT. Depending on safety/lethality/mission criticality of the software function(s) being released, test planners should consider when to sequence an end-to-end, high-volume capacity test. The timing of these events should be determined based on required confidence in the program and assess the appropriate level of operational suitability. User demand signal also drives when these dedicated events take place. Figure 8 provides an elementary depiction of the tension between releasing regularly, scheduling required dedicated OT, and software deployments based on user demand (“user pull”). End-to-end high-volume capacity testing in an accredited, operationally-representative environment becomes increasingly critical as the likelihood of unexpected system interactions increases with complexity.

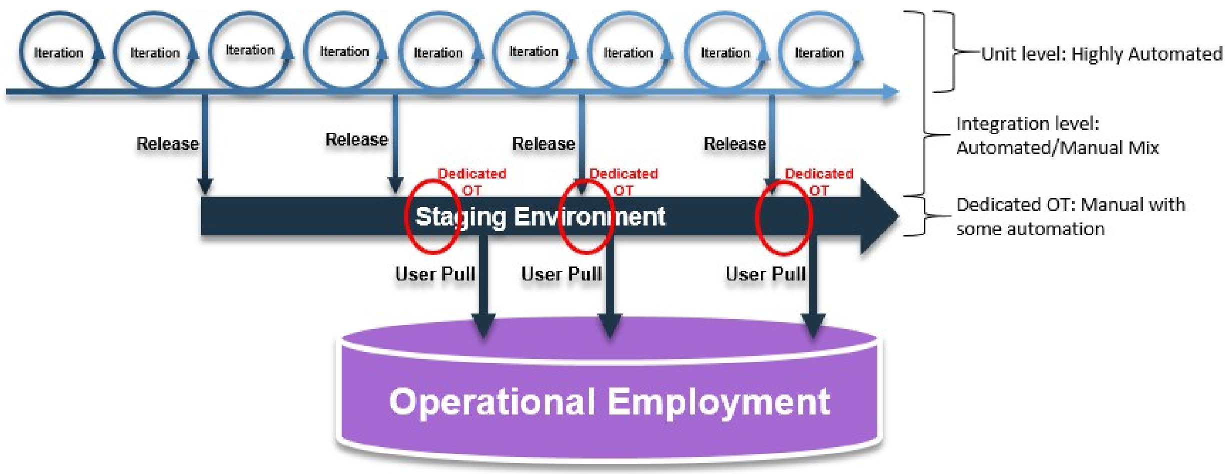


Figure 5.8: Staging Test Activities

Maximize the use of automated test and develop a process for incorporating manual test cards into automated test scripts. Automated test increases the likelihood of quality, viable code resulting from each iteration. For example, test-driven development is a software engineering technique that involves writing the test code first that the functional code must pass. Automated test pipelines also add robust cyber survivability testing to every line of code being developed and complements manual cyber survivability testing such as CVPAs and AAs. Manual testing can be minimized by a well-developed automated test plan. Moreover, manual test procedures should be incorporated into automated test scripts as soon as possible to alleviate tester burden and maintain overall ASD velocity. Finally, maintenance of automated test should be regularly performed by developers and testers to eliminate any outdated test scripts.

Assessment of the automated test environment (e.g. test tools and/or simulations) is part of all test approaches. Automated tests enable rapid agile development, but must be certified for DOD use. Certification of test scripts addressing high-risk software functionalities must involve any combination of the LDTO or OTA/OTO prior to being integrated into automated test.

ASD depends on considering and planning cyber survivability testing from the very beginning of a development cycle. However, testers should carefully assess when to bring in cyber survivability cooperative vulnerability and adversarial assessment teams as these resources will likely require advanced scheduling. Fixed release schedules greatly aid in forecasting the need for these resources.

Documentation.

Tailored test documentation must be able to keep up with the ASD process while still capturing a valid program health snapshot for stakeholders. While planning the test approach itself, testers should establish the type and frequency of reporting to match software development/release cycles. Minimum documentation should capture relevant data quickly to enable subsequent software iterations.

TEMP. At a minimum (and depending on level of oversight and program complexity), a TEMP should focus on four major areas: overall tailored test strategy to include an evaluation framework, resources, schedule, and limitations. The strategy portion includes the planned release schedule, how releases are structured with regard to test, resulting risk-based test approach, and defined program safeguards that keep the software system on target to meet requirements. Describe the pipeline from developer to released product. Release schedules, test events and correlating test measures may be captured in TEMP annexes to preserve document flexibility.

CONOPS. CONOPS should be established by program initiation and be promulgated to the program office, developer, and testers to shepherd the program's test strategy. The CONOPS may indicate user demand signal and therefore drive software delivery schedules and/or dedicated test opportunities.

MTP and OTP. The Master Test Plan (MTP) is a vehicle for generally mapping iterative and release Operational Test Plans (OTP). The MTP provides the overarching approach derived from the TEMP as well as the framework for keeping test at pace with ASD velocity. The underlying Operational Test Plans (OTP) should guide test event planning and determine time- or event-based OT looks as software system capabilities evolve. The OTP enables operational testers to formulate an adequate plan by supplying specific anticipated release capabilities, release features and a basic understanding of mission operations. It should outline the specific data to be captured, the circumstances under which those data are generated, and how the data will be collected. A test matrix (Figure 5.9) is a useful tool in tracking testing against requirements across multiple software development cycles. This strategy would allow aggregating test results to fulfill "dedicated OT" (IOT&E) while ensuring full coverage to accurately characterize the system.

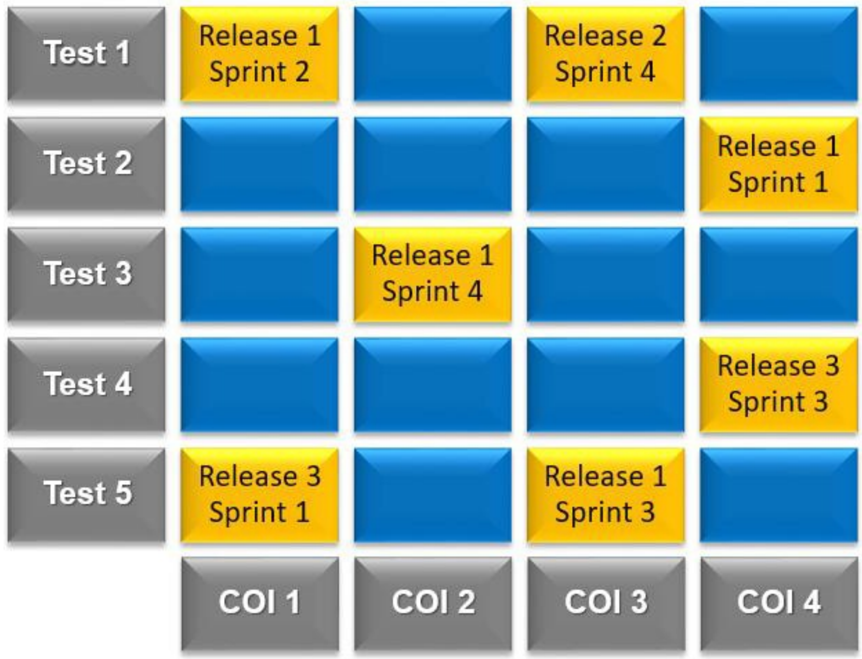


Figure 5.9: AST Matrix

Reporting.

Tailor reporting to provide a concise, accurate and relevant program assessment while avoiding undue delays to the ASD process. Keep statutory versus regulatory reporting requirements in mind while developing the test and reporting approach. Reporting can take on many different forms but should be established early during test planning in order to posture the lead command, PMO, ITT and IPTs for timely testing, feedback, and backlog reprioritization. Determine the questions needing answered or decisions the report will inform, and write it accordingly.

Testers, together with PMs, should assess the frequency and types of reporting that preserve program agility as much as practical without creating program health blind spots.

Furthermore, delegating test report authority down to the lowest level will enable reporting to keep pace with, and remain relevant to, ASD cycles. Reporting must allow flexibility in rigor and formality based on the capability under test. For example, smaller, less risky features can be reported by less formal means using content collaboration tools (e.g. Confluence) while higher risk capabilities should be written to a format acceptable to the lead command. Regardless of the format, test organization approval of these reports should be delegated as low as possible to maintain their relevance.

Automated reporting. Automated testing can generate a large amount of data that may exceed test community's capability to analyze in a timely fashion. However, some test tools can produce summary reports that can be tailored to support the test's community analysis process. Test strategy should account for what automated reporting should focus on and, if analysis resources are limited, which key data indicators trip program decisions or follow-on testing. Automated testing and reporting should be built in from the very beginning.

Quick-look reports. Quick-look reporting should take into account how quickly the feedback from one software iteration makes it into the next iteration. If quick-look reports are to be used, the time it takes for resulting feedback to affect subsequent backlogs needs to be accounted for in the MTP. Since software iterations/releases result in continually evolving software capabilities, quick-look reports should provide capabilities, limitations, and risks snapshots and provide recommendations as part of the resulting user story fed into the backlog. The quick-look report format must be established during early software release cadence road mapping; in the interest of maintaining relevance, its approval level should ideally rise no higher than the tester's flight commander.

Aggregate test data. Aggregated test data, whether resulting from automated or manual testing, must meet or exceed predetermined criteria before informing an IOT&E or equivalent report.

Metrics. The ITT should develop guidance that aid IPTs in developing metrics to accurately indicate program progress as well as gauge test effectiveness. Metrics answer whether test is doing what it needs to do in order to ultimately delivery products that work. Test, through effective metrics, should support the agile process while validating the quality of automated testing (i.e., the amount of coding errors caught through test). Other things metrics can capture are (but not limited to) defect density, amount of code covered by test, mean time to recover/repair, and velocity of releases. The Carnegie Mellon University Software Engineering Institute lists the following three key metrics:

> **Velocity**. Velocity is the volume of work accomplished in a specified period of time, by a given team. Typically, this is measured as story points accomplished per iteration. This measure is sometimes called "yesterday's weather" by agile practitioners, as if to indicate its sensitivity to local conditions as well as seasonal trends. Indeed most experts explain that velocity is "team-unique" and thinking of this

measure as a parameter in an estimating model is a mistake. The team must establish its own velocity for the work at hand.

Iteration/Sprint Burn-Down Chart. A graphical technique provides a means for displaying progress for the development team during an iteration. As items in the backlog of work are completed, the chart displays the rate and amount of progress. This chart is typically provided for viewing on a team's common wall, or electronic dashboard.

Release Burn-Up Chart. A complementary graphical technique for the iteration burn- down, the release burn-up chart is also commonly used. Many cling to the convention that iterations burn down and releases burn up--though there is no mathematical principle that governs this choice. With each completed iteration, the delivered functionality grows, and the release burn-up chart depicts this progress in an intuitively logical fashion. This concept makes use of workflow management tools and other extensions of the concept.

Deficiency Reviews and Test Readiness Reviews. The traditional TRR becomes part of the release/iteration planning process and the DRB approach fragments into various iteration reviews and user stories which inform product backlogs. Essentially, there are many “mini TRRs and DRBs” peppered throughout the software development process. TMs and PMs must ensure the Agile test and deficiency reviews maintain rigor commensurate with traditional programs, particularly as they are conducted in collaborative tools versus the Joint Deficiency Reporting System (JDRS). AFI 99-103 states that “all Government testers will use JDRS for weapon systems deficiency reporting as described in TO 00-35D-54 unless a waiver is approved IAW that TO.” If JDRS is not compatible with the program’s selected Agile development method, programs must furnish the waiver request to demonstrate due diligence when developing alternate deficiency reporting methods.

Additional Considerations.

Workforce expertise. Currently the USAF T&E workforce is structured to support traditional acquisition testing. Test organizations supporting ASD programs should educate their testers on ASD concepts and coordinate closely with developers to tailor their training to the program’s selected ASD technique. ASD involves heavy user involvement so leverage user capability and knowledge when considering their role in test. Government testers are typically not coders, so they must articulate their test plans to their software developer teammates so these can be incorporated into TDD and other automated test tools.

Workforce capacity. The USAF T&E workforce is structured toward traditional acquisition workload demands. However, ASD demands a higher amount of continuous tester involvement throughout versus near the end of the program's development. T&E organizations may be tasked to persistently support multiple ASD programs and test planners must research availability and schedule appropriately, particularly if it involves high-demand, low-density capabilities (e.g., cyber red teams). This may be a key driver when assessing risk. Furthermore, determine the balance of contractor versus government test responsibility and capacity to inform the program's test strategy. Therefore, thoughtful planning should go into placing embedded testers where they have the most impact, leveraging remote collaboration tools, and maximizing automated test provided by software factories.

Local versus remote testing. As mentioned previously in this section, test planners must ascertain whether testers must be physically present (or how often) or if testing can be accomplished remotely. If the former is the case, funding and other bed down issues will have to be resolved, particularly if the test facility is not located on/near a DOD installation. If the latter is selected, planners must ensure remote access to the program's software platform is possible. Additionally, inherent latency issues must be analyzed to determine their effect on testing.

Security. Test planners face the potential of testing software on a variety of systems across escalating classification enclaves. Therefore, data access or transfer limitations between different platform enclaves, as well as security clearance disparities between developers, testers and users must be taken into account. Formal data management planning and documentation will help resolve security, network and personnel access.

Synchronization and interoperability. ASD does not happen in a vacuum; consider development pacing for any subprograms that make up the greater system. Where are the integration points? Awareness that system risk and vulnerabilities increase with system complexity is critical. When and how often can interoperability and cyber survivability be tested? How will these impact overall, end-to-end test? Will everything be ready on time? Can subprogram iterations be aligned to allow consolidated test? How early can operational users and resources be leveraged to help build automated integration tests? For programs involving Joint players that may not always have access to the CI/CD pipeline, schedule deliberate testing for established release windows (e.g. while USN vessels are in port). Answering these questions will determine the availability of test resources, especially external ones that are shared with other defense acquisition programs.

Operational Flight Program (OFP) and Hardware-in-the-loop (HWIL) testing. Improperly integrated HWIL testing (either procedurally or architecturally) can quickly turn an AST effort back into waterfall, particularly if it introduces air-gap data transfer solutions. Since Agile by nature requires regular user input, aircrew availability can stagnate delivery velocity as well. TMs should work closely with flight test organizations to forecast when aircrew collaboration is needed. Since OFP software typically enters a SIL followed by a HWIL lab prior to being loaded into aircraft (Figure 5.10), presenting multiple potential choke points. Therefore, OFP release decisions should be delegated down to the test squadron level in order to keep pace with the ASD release cadence. If possible, TMs and PMs should establish HWIL environments that can accept automated testing. Programs should also explore the ability of incorporating HWIL labs into the CI/CD pipeline.

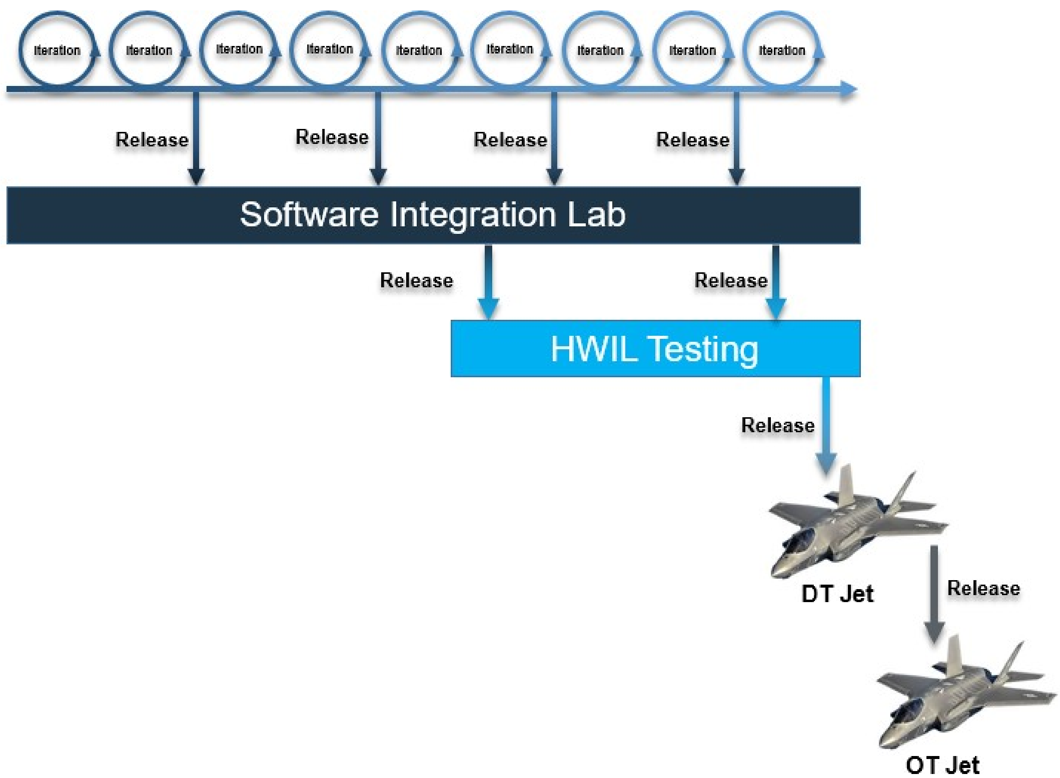


Figure 5.10: Deploying Agile Developed OFP Software

Additional Resources.

Agile Alliance Website:

https://www.agilealliance.org/

Air Force Test and Evaluation Agile Software Development Test SharePoint:

https://haf-te.sharepoint.afncr.af.mil/ASDT/SitePages/Home.aspx

DoD DevSecOps Community of Practice:

https://www.milsuite.mil/book/groups/dod-enterprise-devsecops/activity

MITRE - DevOps for Federal Acquisition:
https://www.mitre.org/sites/default/files/publications/15-2842-DevOps-for-Federal-Aquisition.pdf

CMU SEI Blog - Agile Metrics: Seven Categories:
https://insights.sei.cmu.edu/sei_blog/2014/09/agile-metrics-seven-categories.html

CMU SEI Presentation - Scaling Software Testing & Evaluation:
https://resources.sei.cmu.edu/asset_files/Presentation/2018_017_001_529045.pdf

CHAPTER 6:
TEST MANAGEMENT RESPONSIBILITIES

Roles and responsibilities. Roles and responsibilities of key personnel and organizations are defined in AFI 99-103 Chapter 2. Two key positions, CDT and Lead Government DT&E Organization (LDTO), need to be identified for each program. Please see AFI 99-103 para 4.5.4 for guidance on the Alternate LTDO option for non-oversight ACAT III programs. In addition, the CDT (MDAP and MAIS) is identified as "ITT Chair" and the OTA/OTO is designated as "ITT Co-chair".

Deficiency Reporting. Ensure the PM has established a database and a process for administering Deficiency Reports (DRs). Remember that using JDRS (or a waiver) is required per T.O. 00-35D-54. Establish who the participants will be and establish the process for prioritizing and resolving DRs.

Test and Evaluation Master Plan. TEMP coordination can be a difficult process if not accounted for early in the program's lifecycle. For MDAPs, the start of the 90-day coordination cycle for signature begins once it is received by the Air Staff. The 90 days is split between Air Staff (45 days) and OSD (45 days). Air Staff consists of AF/TE and SAF/AQ. OSD consists of DD(DTE&P) and DOT&E. Programs that have navigated this timeline successfully have completed early coordination with AO's from AF/TE, DOT&E and DD(DTE&P). Some programs not on oversight may not require the OSD review. Relevant issues should be understood and resolved before submitting for signature.

DODI 5000.02 requires a TEMP update for RFPs. RFP TEMPs must be signed by the PM, PEO, and CAE. Air Force guidance is that AF/TE will also sign RFP TEMPs that are on oversight. The coordination cycle for TEMPs at the Air Staff will also apply to RFP TEMPs (i.e. 45 days for Air Staff approval). For programs that enter post MS A, a MS A TEMP has not been accomplished. Therefore, an update of the TEMP for RFP release is effectively the MS A TEMP. Updates are required for significant changes to the test program. See AFI 99-103, para 5.15.

Rapid acquisition programs may not be subject to DODI 5000.02 requirements and this may have implications for TEMP (or equivalent document) timelines as well as content. Check with AF/TE for latest TEMP guidance regarding rapid acquisition programs. See para 8.1 for more information regarding implications for test regarding rapid acquisition programs.

Regardless of program type, useful TEMP references include:

a. AF/TE TEMP Guide

b. Defense Acquisition Guidebook (DAG), Ch 8 – TEMP Guidelines

c. DOT&E TEMP Guidebook

Detailing test program resources and limitations is critical. This data is compiled by AF/TE to provide an aggregate resources and limitations picture across the AF test community. This not only informs T&E resource decisions and budgeting, but also allows AF/TE to forecast future requirements that empower emerging test needs and associated ranges, assets, and manpower.

This period of the acquisition cycle is of the greatest importance for the T&E professional. During this timeframe, you can set the foundation for timely, technically adequate testing to allow the Program Manager to make informed decisions for the program's future milestones (post MS B & EMD phase). All aspects of the program's testing must be thought of at least in a notional way, and can be better defined as the program matures (e.g. ITT membership, TEMP generation, LDTO & OTO identification, cyber test strategy, etc.). AFI 99-103 paragraphs 4.4, 4.5, 4.6 provide in-depth guidance for early tester activities.

MS A TEMP Requirements. The MS A TEMP should address major sections of the TEMP outline in the DAG, understandably with limited detail available at MS A. The MS A TEMP should include the following: an estimate and plan for required resources to support adequate T&E, all planned T&E for phase completion including test entrance and exit criteria, and a strategy and resources for cyber test and evaluation to name a few; refer to para 4.11.3 in 99-103 for a complete list.

T&E Events Prior To MS B. Keep tabs on T&E funding, secure appropriate Modeling and Simulation resources, and know the capabilities and limitations pertaining to contractors as the program progresses toward MS B. If it is an MDAP and the MDA is the Service Acquisition Executive (SAF/AQ), a DT&E sufficiency assessment must be accomplished and staffed prior to the MS B decision briefing.

DT&E Sufficiency Assessment. Section 2366b(c)(1) of Title 10, United States Code, requires an assessment of the sufficiency of DT&E to be provided to congressional defense committees. In order to support this statutory requirement, begin coordinating with the LDTO NLT 60 days prior to the anticipated MS B decision brief for information that addresses the sufficiency of:

a. The DT&E plans within the Test and Evaluation Master Plan.

b. The DT&E schedule, including a comparison to historic analogous systems.

c. The DT&E resources (facilities, personnel, test assets, data analytics tools, and modeling and simulation capabilities).

d. The mitigation of known risks of developmental test and production concurrency.

e. The developmental test criteria for entering the production phase.

T&E Funding. One of the largest concerns for the T&E professional leading up to milestone B is securing the funding streams and the contract vehicles to perform the bulk of upcoming testing. It is crucial that the CDT and T&E action officers work with both the PM and the program's contracting officer to ensure all applicable T&E resources (test articles, ranges, facilities and/or contractor support) have been accounted for, allowing seamless completion of the required program test activities.

Modeling and Simulation (M&S) in Support of T&E. Increasingly complex battlespace environments, cross-domain systems interdependencies and increasingly capable and dynamic threats drive the necessity for capable M&S. As the DOD relies more and more on M&S as a T&E force multiplier, the full scope of a program's life-cycle needs to considered in order maximize the cost benefits to the overall T&E outlay. Two great resources for M&S are the DoD Modeling and Simulation Coordination Office (MSCO) and the Air Force Agency for Modeling and Simulation (AFAMS). These two databases allow a tester to review M&S assets and applications to reduce duplication of existing technology and products.

a. **Joint Simulation Environment (JSE)**. JSE is a high fidelity, T&E focused, M&S initiative intended as an Air Force-focused next step in the development of that program with an eye to expansion of the JSE to include other 5th generation and beyond aircraft/systems and, in the long term, be expanded to include the multiple domains e.g. Space, Cyber, Land, etc. If approved to go forward, AF M&S and acquisition policy will need to be adjusted requiring acquisition programs to actively plan and resource for the use of JSE as a development tool to include funding a high- fidelity model of their weapons system for use in JSE.

T&E Activities In Support Of Milestone C And Beyond. Acquisition Strategy and Schedules need to be coordinated with the OTA/OTO; see AFMAN 63-119 Attachment 2. If it is an MDAP and the MDA is the Service Acquisition Executive (SAF/AQ), a DT&E sufficiency assessment must be accomplished and staffed prior to the MS C decision briefing.

DT&E Sufficiency Assessment. Section 2366c(a)(4) of Title 10, United States Code, requires an assessment of the sufficiency of DT&E to be provided to congressional defense committees. In order to support this statutory requirement, begin coordinating with the LDTO NLT 60 days prior to the anticipated MS C decision brief for information that addresses the sufficiency of:

a. The sufficiency of the DT&E completed.

b. The sufficiency of the plans and resources available for remaining DT&E.

c. The mitigation of risks identified during developmental testing to the production and deployment phase.

d. The readiness of the system to perform scheduled initial operational test and evaluation.

Do not just include a program schedule. Provide verbiage that communicates how the schedule will be achievable: Reasonable T&E cycle times with sufficient slack for rework, analysis, reports. Where the program is accepting program risk, identify mitigations in place.

Phasing is reasonable: Key performance capability evaluated before MS C and is balanced relative to that after MS C.

Miscellaneous.

Despite having a designated lead command per AFPD 10-9, some ACAT III, non-OSD oversight programs support multiple users with differing requirements. The lead MAJCOM and AFOTEC will negotiate an OT&E involvement role, or coordinate with appropriate HQ MAJCOM T&E OPR for a multi-MAJCOM/AFOTEC test approach.

AFI 99-103 and DoDI 5000.02 address the limitations on contractors during operational testing with the one exception: Contractor personnel may only participate in OT&E of Air Force programs to the extent they are planned to be involved in the operation, maintenance, and other support of the system when deployed in combat. This delineation is important to consider during the last phases of DT when approaching an Operational Test Readiness Review (OTRR) to ensure that sufficient training of OT personnel has been properly accomplished. Under-trained OT personnel can drastically affect the OTO/OTAs assessment of the system under test.

Some programs may not be clearly “owned” by a MAJCOM or sponsor with an organic operational test function. In these cases, the program’s sponsor coordinates with AFOTEC to identify an appropriate OTO, with respective MAJCOM concurrence, to complete any required operational testing. If an appropriate OTO cannot be identified, the sponsor contacts AF/TE for guidance.

If the OTO and lead HQ MAJCOM T&E OPR jointly agree that no operational testing is necessary, the LDTO provides relevant DT&E data that supports the option to not conduct operational testing. The OTO reviews the LDTO's work, assesses the risk of accepting that work, and documents their assessment with a Sufficiency of Test Report (SOTR). SOTRs can only be used for non-oversight programs of limited scope and complexity. Some clarification: the ITT recommends a SOTR when there is sufficient DT data to answer all test measures. The OTO confirms this through analysis and assesses this can support a fielding or production decision.

CHAPTER 7: TEST IN RAPID ACQUISITION

General guidance.

This guidance applies to Middle-Tier Acquisition (MTA) rapid prototyping and rapid fielding activities using authorities provided by Section 804 of Public Law (P.L.) 114-92, as amended by Sections 849(a), 864(b), 897 and 1081 of P.L. 114-328. This guidance also applies to tailored DoDI 5000.02 programs. This guidance references both MTA and tailored DoDI 5000.02 programs as "rapid acquisition activities." This guidance also applies to operational demonstrations or experiments supporting rapid acquisition activities.

Rapid acquisition requires rapid learning. Test allows us to learn about the system, correct its faults, and develop it further. We must take an approach to test that allows it to keep pace with rapid acquisitions. This preserves the ability to effectively characterize a rapidly developing system and increases probability of success as the program approaches readiness for fielding or production.

Rapid acquisition may necessitate taking greater technical or program risk, but not at the expense of safety. While a full characterization of the system considering all environmental factors may not be necessary or realistic; smart, mission-focused test tailored to the expected operational environment is required. The warfighter must be involved in the risk acceptance decision.

Independence of test is a principle that ensures objective and unbiased test results. Testers can (and in many cases should) be embedded with rapid development teams, but they should develop test reports and be supervised independently of program management chains. Independence does not preclude participation by the program office or development team in test. Contractors and system developers are important partners in test and automated software tools can enable test efficiency.

Operational effectiveness, system interoperability, safety, cybersecurity, maintainability, and other requirements are critical factors and should be evaluated by independent testers.

Speed in system development is achieved by early tester involvement. Testers can aid the PM in developing a realistic test strategy and schedule as well as identify resources and easy design decisions that enable efficient test. Integrated test teams are a best practice to realize effective test planning, efficient test execution, and timely and relevant reporting.

Roles and Responsibilities.

The Director, Air Force Test and Evaluation (AF/TE), as the senior Service official responsible for developmental and operational testing will:

a. Provide guidance, direction, and oversight for the formulation, review, and execution of T&E plans, policies, programs, and budgets.

b. Review and approve top-level test and evaluation plans to include schedule, tests, resources, i.e. the TEMP.

c. Review and prepare release of T&E reports outside the Air Force and ensure test results are available to decision makers in a timely manner.

d. Provide assessment of Sufficiency of Developmental Test Reports for DoDI 5000.02 programs to the DD(DT&E) or Service Acquisition Executive (SAE) as applicable.

Program Executive Officer (PEO)/MDA will assist the PM and ITT in identifying key government DT&E organizations and personnel, to include LDTO candidates, CDTs, and TMs.

The PM shall:

a. Prioritize early tester engagement

b. Appoint a CDT or TM as applicable

c. Assist the CDT/TM, LDTO and the OTO (or OTA) in forming an ITT

d. Develop T&E strategy with early tester collaboration

e. Request an appropriate LDTO and OTO

OTA or OTO will participate in preparation of strategies for T&E, plan and execute operational test in accordance with user-provided CONOPS.

LDTO will:

a. Function as lead integrator for all DT&E activities

b. work with the CDT/TM to plan and conduct DT&E and oversee contractor developmental test

Test Elements.

Rapid acquisition activities will include the following minimum items:

- CONOPS
- ITT
- CDT or TM
- LDTO
- OTO
- Integrated Master Test Plan (Objectives, Schedule, Resources, Evaluation Frameworks)
- Developmental Test Plan
- Operational Test Plan
- Test Review that considers adequate technical and safety review
- Test Report

Figure 7.1 illustrates relationship and flow between these test elements. Each item must be addressed, but may be tailored to each program’s need.

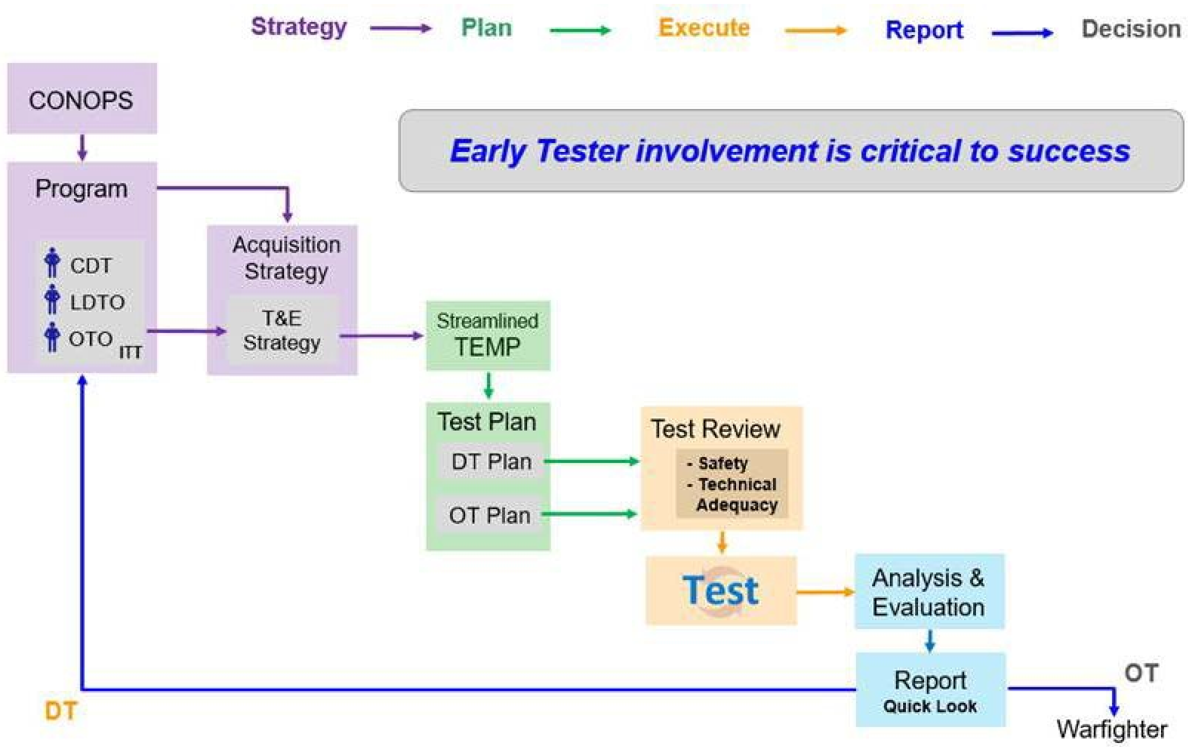


Figure 7.1 Essential Test Elements

Master Planning.

CONOPS.

The CONOPS is the over-arching, high-level vision and architecture that describes the desired system capability or function. Expected system employment informs requirements to include operations, maintenance and logistics.

Lean design, agile methodology, and continuous development that permit requirements to emerge as part of the development process are acceptable, but there must be a common, clear understanding of how the user will actually employ the capability to guide development and permit Evaluation (the objective of T&E).

T&E Strategy.

The PM should map out the test and evaluation strategy at program initiation with test agency involvement. Early, realistic planning coordinated with developmental and operational testers will set the stage for efficient and effective testing that is aligned with the CONOPS.

The T&E strategy should be mission focused and include an operationally realistic environment. This allows for early identification and correction of deficiencies based on operationally relevant data.

Early and frequent smaller-scale assessments enable rapid learning, allowing greater and earlier influence in system design. These should be small subsets of selected combinations of components, subsystems, environments, and scenarios. This ensures that critical configurations are tested and lowers the risk of unexpected system malfunctions near fielding or production decisions.

An integrated DT/OT approach will result in greater efficiency by maximizing resources and reducing the number of dedicated test periods. Early collaboration between the LDTO and OTO is key to ensuring a cohesive integrated test approach.

The T&E strategy should address safety, program requirements, technology maturity, interoperability requirements, required operator expertise, training, existing data, sustainment, and mission impact.

Prototyping and experimentation can reduce development risk. Digital engineering and modeling and simulation can fulfill some test requirements.

Integrated Master Test Plan.

The proposed T&E strategy and resources must be captured in a document to ensure all parties (testers, program office, AF/TE) are aware of the way-forward for test execution. A Master Test Plan describes the overall test objectives (“the why?”), the roles and responsibilities (“the who?”), the resources required (“the what?”), and the overall timeline for test (“the when?”). Timelines should clearly convey precedence of events as there will likely be uncertainty in the schedule driven by unknown factors.

a. Minimum requirements for this document are: Objectives, Schedule, Resources, Limitations, and Integrated Evaluation Framework. Live Fire Test and Evaluation (LFT&E) should also be described as required.

b. Alternate formats (including tailored traditional TEMP) should be considered. The objective is to provide minimum information necessary to understand the test strategy, schedule, and resources required to collect the data needed for the MDA to make a fielding or production decision.

- Signature and coordination should be limited to those with a valid stake in the T&E plan. As an example for non-oversight programs, coordination should include the decision authority, PM, LDTO, and OTO as a minimum. For programs on OSD T&E oversight, the master test plan will be coordinated IAW the TEMP signature and coordination process in AFI 99-103.

Test Design.

Planned tests must assess system capability against user CONOPS in an operationally relevant environment as well as other requirements such as interoperability, cybersecurity, and sustainability.

Testing should always be scoped to the minimum required to satisfy test objectives and inform decision-makers. This is true for all programs. In order to minimize test scope, test design must consider the relevant variables and factors (e.g. using Scientific Test and Analysis Techniques [STAT]) to determine the data needed. These factors will in turn generate necessary test measures and test events.

Priority during DT must be given to requirements or capabilities that could impact system design. Test must find critical gaps early enough to facilitate timely and cost effective system design changes.

Operational effectiveness and suitability (including Reliability, Accessibility, and Maintainability (RAM)) considerations will drive test requirements, to be recorded in the integrated master test plan.

Operational environment or test resource constraints will drive test limitations, to be recorded in the integrated master test plan.

Test Execution.

Test execution must be disciplined, relevant, and timely. A good test plan provides the framework for test execution ensuring appropriate test points are accomplished to produce useful data. Early tester involvement ("moving test left") is the best way to learn early and support rapid acquisition.

Integration of developmental and operational test should be maximized to efficiently utilize resources and reduce overall test duration.

Mission-focused test opportunities during developmental test permits early operational assessment and feedback. Tests should be designed using proposed CONOPS as appropriate.

Provide shared data to support independent analysis, evaluation and reporting by all stakeholders:

- Independent technical adequacy review of the planned test.

  a. Technical adequacy reviews of the test design and operational test readiness reviews are key to ensuring that tests are effective (i.e., the test will produce data to answer the test objectives).

  b. The only failed test or failed experiment is the one that cannot answer the objective or hypothesis because of a flaw in the test or experimental design.

  Systems failing during test is not a failure of test (unless it could have been tested earlier and was not).

- Independent safety review of the planned test.

  a. Rapid learning supports rapid acquisition. "Failing fast" does not mean being reckless. Risk management and safety are paramount in test. Test agencies will assess safety risks to personnel and property and mitigate them appropriately.

  b. Robust safety planning is based on risk awareness: the perception of uncertainty and the potential, projected outcomes resulting from uncertainty.

  c. Early involvement enables adequate time for testers to build knowledge and understand the system under test in order to characterize uncertainty (safe test execution requires understanding system boundaries).

  d. Proper test safety planning minimizes potential for program delays by identifying hazards and allowing for progress through mitigation efforts.

Test Reporting.

Test reports should be relevant, timely, factual, and concise.

- Test organizations can support rapid learning cycles with "quick look" reports and other, innovative processes.
- Test reports should be written to meet the decision maker's needs.
- Test reporting and coordination should be tailored to minimize impact to production and fielding.

Systems Engineering.

Rapid acquisition programs require systems engineering. Time and resources dedicated up- front to system design and system engineering will result in smoother development and more effective and efficient test. A CONOPS-focused approach ensures system development supports desired capability or function. Understanding the mission helps define system functional, interoperability, and security requirements.

- "Test-driven development" and "engineering-for-test" enable efficient test and rapid learning and ultimately support rapid development.
- Systems engineering informs decomposition for subsystem test and integration for end-to-end, systems-level test.

Complex systems require deliberate testing.

- By definition, system-level behavior of complex systems is non-deterministic; test rigor is more important as complexity increases.
- End-to-end, integrated testing in operationally representative environments is the best method to understand performance of complex systems. Programs can accelerate learning about complex system performance by conducting early mission-focused testing in relevant environments utilizing test strategies that examine the most stressing combinations or most likely use cases. This strategy can also expose potential operational issues early in the development cycle and reduce the likelihood of a costly and time- consuming delay towards the end of a program's development.
- Modeling and simulation and component-level testing assist in developing

answers for complex systems. While limited in their ability to provide the full complex system performance answer, they can provide answers for cases not testable any other full end- to-end testing. The test strategy must include proper resourcing and schedule consideration for modeling and simulation validation and accreditation activities.

AF/TEP will assist program offices in bringing in the developmental and operational testers early onto a program so that the program can plan, fund, and execute the most efficient, capability-based, operationally relevant testing possible.

CHAPTER 8:
TEST SUPPORT TO EXPERIMENTATION

Experimentation is currently not addressed in AFI 99-103 but warrants discussion as there is potential for its use during program development. The likelihood of experimentation under the auspices of test is increased by a DOD-wide push to accelerate defense acquisition. There are differences between experimentation and test, but also similarities; in fact, there is a close connection between the two. First let us review their definitions:

Test - A procedure intended to establish the quality, performance, or reliability of something, especially before it is taken into widespread use (Random House, 2018).

a. An African Swallow development program is designing toward a validated requirement of an unrefueled, medium-altitude, still-air range of 1,000 nautical miles carrying one coconut. A test characterizes the swallow's true capability as 998 miles, informing the warfighter acceptance decision.

Experiment - A scientific procedure undertaken to make a discovery, test a hypothesis, or demonstrate a known fact (Random House, 2018).

a. An experiment will determine whether a remote control receiver integrated with the swallow will allow in-transit commands to avoid unforeseen threats along the flight path. This experiment will validate a hypothesis and possibly result in a new CONOPS.

For additional context, AFI 99-103 defines T&E as "The act of generating empirical data during the research, development or sustainment of systems, and the creation of information through analysis that is useful to technical personnel and decision makers for reducing design and acquisition risks. The process by which systems are measured against requirements and specifications, and the results analyzed so as to gauge progress and provide feedback." T&E should be tailored based on program and/or system complexity.

Experiments (does A cause B?) generally focus on the search for solutions to address capability gaps, whereas tests seek to characterize the relationship of said solution to a defined requirement. In other words, experimentation shapes the development of a system CONOPS while test informs how well the system will perform with regard to established CONOPS. The interconnectivity of experiment and test can be illustrated by viewing an experiment as a systematic sequence of individual test events to examine a causal relationship (test a hypothesis), while a test quantifies an attribute. Furthermore, experimentation is a trial-and-error process that results in an approach to solve a problem. Test informs how well the approach works.

Demonstrations show how a process or system works and is neither an experiment nor a test. Demonstrations assume the process or system works based on previous experimentation and testing results.

The International Test and Evaluation Association (ITEA) identifies four validity requirements for rigorous experiments:

- Ability to employ the new capability
- Ability to detect change
- Ability to isolate the reason for change
- Ability to relate results to real operations

These requirements are applicable to T&E and must be considered as the PMO and ITT craft their overall strategy. Other overlapping characteristics between experimentation and test include:

a. **Apply a methodology**. Whether using the scientific method to explore a hypothesis through experimentation or following a test plan to characterize a SUT, a clearly stated methodology identifies guardrails as well as practices in order to effectively achieve results.

b. **Set clear objectives**. Clear objectives allow individuals and organizations to understand the reason behind the experiment or test, keep energy and resources on task, when to terminate efforts, and context for interpreting the results.

c. **Plan to meet objectives**. A clear plan follows pre-established methodology in order to achieve predetermined objectives. Everyone involved with the experiment or test should understand how the activity should be conducted, what the limitations and risks are, and when the effort should terminate. This is key to preventing unduly risky practices and also keeps efforts focused on achieving the aim behind the experiment or test.

d. **Execute per the plan**. Once the plan is understood by all involved it must be faithfully executed in order to mitigate any risks to personnel, assets, and other involved resources. Conduct all pertinent reviews to preclude safety violations. Additionally, following the plan allows a systematic assessment of the results and allows for valid changes to the plan, if required.

e. **Report results**. Accurate reporting allows effective assessment of a hypothesis or characterization of a system or capability. This will in turn either validates said hypothesis or capability, or results in 1) subsequent experiment that zeroes in on a CONOPS or 2) feedback resulting in system fixes or increased warfighter understanding.

A key difference between experimentation and test centers on the amount of coordination required prior to the event and the subsequent review process. Experiment plans are typically reviewed internal to an organization before proceeding. Conversely, DT/OT involves more detailed test protocols to properly characterize the SUT, and deviations may require extensive coordination with both internal and external agencies depending on test program complexity and oversight. The level of reporting after an event typically follows suit with the amount of coordination done prior. Whereas experiment feedback can stay within the organization to inform a subsequent experiment, test results can be used to inform stakeholders spanning the individual service to DOT&E.

Another key distinction between test and experimentation lies in their relation to a system's performance envelope. Experimentation's purpose is to discover the boundaries of said envelope, so uncertainty expands beyond pure system performance to encompass safety to a higher degree than test. Test normally occurs within predetermined safety boundaries (boundaries are avoided versus probed); while safety is still accounted for, the performance realm where test occurs should see less risk on the safety aspect while system performance is characterized.

Test organizations involved with experimental efforts will apply the same level of planning and execution rigor as if conducting official test activities. Test organizations will follow AFI 99-103 when planning and executing experiments. Specifically, DT and OT organizations will conduct technical, environmental and safety reviews.

a. Testers involved with experiments will not eschew regulations and other guidelines already in place to prevent mishaps. For example, aircrew responsible for conducting airborne experimentation will continue to adhere to established operational risk management (ORM) procedures. If none are established, an ORM process will be created and vetted by the local safety office prior to initiation of any experimental activities.

b. Test organizations, whether overseeing, advising, or directly involved with experimentation, will ensure all participating units adhere to AFI 91-202, The US Air Force Mishap Prevention Program and AFI 91-203, Air Force Consolidated Occupational Safety when formulating and conducting experimental activities.

DT and OT organizations will assess the nature of the potential experiment's failure to determine the risk of injury or death to personnel as well as property damage. Testers will adhere to standards prescribed by test and safety regulations, and the LDTO and participating test organization (PTO) will hold all experiment co-participants to the same standards. Although the goals of experimentation may differ from test, the approach to reaching those goals does not.

CHAPTER 9: REVIEWING CAPABILITIES-BASED REQUIREMENT DOCUMENTS

Basic Requirements Policy. This attachment covers what testers should know prior to making inputs to capabilities-based requirements documents (CBRD). It is based on policies in CJCSI 3170.01, Joint Capabilities Integration and Development System [JCIDS], CJCSM 3170.01, Operation of the Joint Capabilities Integration and Development System, and AFI 10-601, Capabilities-Based Requirements Development. These policies, along with all available studies, analyses, concepts, threat documents, and other information about the proposed system, should be reviewed prior to attending high performance team (HPT) meetings. This attachment contains checklists for use during and after HPT meetings.

Types of Requirements Documents. Requirements documents are now collectively known as capabilities-based requirements documents (CBRD). CJCSI 3170.01 and AFI 10-601 describe five main types of CBRD: the Joint Capabilities Document (JCD); the Initial Capabilities Document (ICD); the Capability Development Document (CDD); the Capability Production Document (CPD); and the DOTMLPF[1] Change Recommendation (DCR). The Air Force has several alternative methods of documenting capabilities-based requirements: the Air Force Capabilities Document (AFCD); the Combat Capability Document (CCD); and the AF Form 1067, Modification Proposal. Some older programs may still have an Operational Requirements Document (ORD) if the user has not updated it since the JCIDS process was introduced.

Traditional Tester Involvement in Requirements. Operational testers have traditionally been key contributors in the development of new operational requirements for future acquisition programs. Users drafted new requirements documents and then multiple organizations, to include the testers, reviewed them for feasibility and accuracy. Testers reviewed many areas such as projected threats, operational concepts and scenarios, operational realism, mission tasks, and system capabilities. Testers helped ensure system parameters were accurately stated so these areas could be properly tested with available T&E infrastructure. The requirements process was slow and often took up to four years to complete. Users and developers expected the test community to provide clear, decisive answers about the new system's capabilities and limitations. However, some requirements were highly problematic if they were not clearly stated, set unrealistic goals, or inadvertently set up the system for failure during test. These problems often surfaced during test planning or after testing was finished, putting the system and possibly the warfighters' lives in jeopardy when the system did not perform as expected. Good requirements were difficult to write, and once written, were difficult to change.

[1] DOTMLPF is the acronym for "doctrine, organization, training, materiel, leadership and education, personnel, and facilities."

Testers Must Be Involved Much Earlier. The need for earlier tester involvement became more urgent as the requirements and acquisition processes were transformed in 2003 to more rapidly deliver new capabilities to warfighters. Testers had to begin their involvement in developing JCIDS documents as early as possible, preferably before the Concept Refinement phase began. However, the increasing technical complexity of systems and more challenging operational environments demanded new skills and knowledge from all testers. Past policies were inadequate because they focused only on operational testers to address operational realism and "testability" issues. The new environment required developmental testers' technical knowledge of advanced technologies and test infrastructure capabilities. Today's T&E policy requires operational and developmental testers, functioning as "core" HPT members, to address the "testability," feasibility, and operational relevance of all new requirements. The goal remains to ensure systems acquired today would be effective in combat tomorrow.

Results of Inadequate Tester Involvement. These results may not appear immediately, but begin showing up as system development progresses. The results of inadequate or no tester involvement are:

- CBRDs may be challenged and/or rejected at the Air Force Requirements for Operational Capabilities Council (AFROCC), the Joint Requirements Oversight Council (JROC), or by DOT&E staff. DOT&E is allowed to challenge requirements they believe to be inadequate.

- Requirements may be misinterpreted in RFPs and contract specifications and system development goes off track. Contractors are not required to consult the Air Force for areas they do not understand.

- Requirements "creep" occurs or unforeseen late changes are needed. Testers cannot eliminate these phenomena, but they can reduce the impacts if they do occur.

- The threat evolves unexpectedly which diminishes the utility of the new system. Testers can ensure threat descriptions are properly bounded, but may need to test against critical new threats nonetheless. The risk is in fielding a new system that is obsolete in the face of this new threat.

- System development is delayed or the system is not ready for IOT&E. Non-production representative articles are tested to recoup schedule despite known limitations.

- IOT&E is halted or the system fails IOT&E. Systems do not deliver the promised mission capabilities.

- Even if IOT&E is successful, OSD may challenge the results based on unintended interpretations of an ambiguously stated requirement.

HPT Tester Membership.

The designated operational testers, as designated in the Program Management Documents (PMD) or other direction, must attend HPT meetings. Generally this is AFOTEC/AS, but will be the MAJCOM operational test organization if AFOTEC involvement is not planned. HQ USAF/TEP will also attend HPTs when possible or if the designated operational tester(s) cannot attend. AF/TEP will function as a backup HPT member if the primary operational test organization has not been determined.

The designated Center Test Authority (CTA) or Responsible Test Organization will attend as the DT&E representative. HQ AFMC/A3 or AFSPC/A3 will serve as backup in case the CTA or LDTO has not been designated.

Preparation for HPT Meetings. Long before the first HPT, testers should assist in developing the Analysis of Alternatives (AoA) to help develop properly stated measures of effectiveness (MOE) and measures of suitability (MOS). These measures will likely be the same measures used in later JCIDS documents and in T&E plans. Testers should review requirements policies along with all available studies, analyses, concepts, threat documents, and other information about the proposed system.

General Guidance.

Testers Do Not Set Requirements. While testers must advocate for changes and clarifications for poorly-stated requirements, it is up to the user to actually set the requirement. Testers should consider whether the requirements are operationally realistic, attainable, testable, and reasonable for the system. Requirements should reflect the operational capability needed, not the technical parameters of that capability.

Recognize Poorly Stated Requirements. Some examples follow:

a. Requirements that call for 100% of anything are often not reasonable and generally not testable. Even if 100% is attainable, it can never be verified in test. Simply switching to very high numbers like 99.9% can drive testing to unattainable and unaffordable numbers of trials to correctly verify the requirement. It’s a statistics thing.

b. Requirements that exceed the T&E infrastructure’s capability to test. For example, an air-to-air missile to engage targets at 300’ above ground level (AGL) may not be testable since current drones do not have the capability to fly that low. A manned target is not a feasible alternative, thus proper T&E support is not available to demonstrate the capability. Other examples that exceed our capability to test: 1) AMRAAM an "all weather missile" -- OSD wanted to know why we didn't test it in a thunderstorm; 2) AMRAAM kill a "bomber-sized target" -- we don't have any bomber drone targets.

c. Overly broad statements and generalizations such as, "AMRAAM not affected by ECM." There are too many types of ECM to ever fully test with new varieties showing up constantly, so the requirement cannot be verified. Making this requirement more specific, such as listing the threats that must be defeated, would be a better strategy.

d. Stating requirements for which other support will be provided. An example is the programmable fuze which lacked the intel support to determine the depth. The following pages contain information specific to each type of CBRD. This review sheet includes information common to reviewing all CBRDs and some information relevant to each specific type of CBRD. More information about CBRDs can be found at HQ USAF/A3/5's website, https://www.afreqs.hq.af.mil/

Initial Capability Document (ICD). The ICD development is broken into two stages. Stage I captures a capability gap and highlights the gap to senior leadership. The intent is to identify capability gaps upfront and early, in order to enhance the opportunity to investigate viable alternatives to fill the gap. The ICD Stage I encompasses the information required in Sections 1-5 of a complete ICD, as described in CJCSM 3170.01. The ICD Stage I document is not normally reviewed from a test standpoint.

ICD Stage II builds upon ICD Stage I, captures the results of the FSA, provides a final recommendation for a materiel approach(es), and enters the JCIDS process as a complete ICD (Sections 1-7). The ICD Stage II supports the AoA, the Technology Development Strategy (TDS), the Milestone A decision, and subsequent Technology Development activities. This is usually the first CBRD reviewed from a test perspective. ICDs are generated very early in the acquisition process. Most of the time an acquisition program has not been identified before the ICD is approved. This makes the identification of test issues difficult. One area of interest surrounds testability. In this case the concern is not limited to unrealistic or unattainable situations, but whether the test infrastructure is in place to support the anticipated testing for the program. In this sense, the testers may "take away" more information than they provide to the HPT. It is important that necessary test infrastructure issues are communicated to AF/TER as soon as possible to maximize the opportunity for developing or otherwise acquireing the needed resources.

Capability Development Document (CDD). AF/XOR may direct a MAJCOM/Agency to develop a CDD however, in most cases, sponsors will have already developed an ICD Stage I/ICD Stage II prior to development of CDD.

CDD Strategy Development Initiated. The requirements strategy lays the foundation for CDD development and supports the System Development and Demonstration phase for a single increment. The sponsor, along with operators, continues the collaboration initiated in ICD development with Air Force acquisition, test, and logistics communities (and other appropriate SMEs). The preferred materiel solution is based on mature technologies demonstrated during the Technology Development phase. The sponsor applies lessons learned during the Concept Refinement and Technology Development phases plus any other appropriate risk reduction activities such as experimentation, T&E, and capability/schedule tradeoffs.

CDD High Performance Team (HPT). The HPT is the preferred method to develop a CDD and is used unless waived by AF/XOR at the RSR. A CDD HPT consists of a lead (normally the sponsor), core and support team members. During the RSR, AF/XOR approves the core team (ideally 8-10 members) and consists of SMEs from the Air Force, government agencies, and other Services as required. Support team membership provides "reach-back" expertise in areas not represented by the core team. The HPT accelerates the documentation process and increases the potential for a quality document. Its overarching objective is to capture, articulate, and document the operator's operational requirements in minimum time, while achieving stakeholder buy-in. The HPT leverages the expertise of all stakeholders by inviting them to participate in the development of the document. Although the sponsoring MAJCOM/Agency maintains ownership of the document, the HPT lead maintains responsibility for writing and ensuring document coordination until approval. One major benefit of a document generated by an AF/XORD-facilitated HPT is the approval to conduct simultaneous Air Force, Joint Staff, Service, and Agency coordination; whereas, non-HPT documents are staffed sequentially.

CDD Development. The CDD, guided by the ICD, the AoA, and the TDS, captures the information necessary to initiate an acquisition program to develop a proposed capability, normally using an evolutionary acquisition strategy. The CDD outlines an affordable increment of capability using mature technology and supports Milestone B.

Capability Development Document
(CDD)

Review Procedures/Checklist and
Lessons Learned

Para 1 - Capability Discussion

- ❑ Review this section to make sure it is an overarching discussion that encompasses the KPPs, thresholds and objectives. No new capabilities should be introduced/discussed in this section that are not addressed in section 6.

Para 3 - Concept of Operations Summary

- ❑ Review this section to make sure the concepts discussed are directly tied to requirements in section 6. Make sure there are no new concepts that could be construed by DOT&E as testable requirements.

Para 4 – Threat Summary

- Make sure that any discussion regarding Threats to be Countered addresses threats to the proposed system that the system is expected to counter. Many times this section erroneously addresses enemy systems that the proposed system will neutralize. For example, a new bomber might be expected to counter surface to air missiles with flares or chaff. This is the appropriate discussion for this section, not that the bomber is expected to counter massed enemy artillery or centers of gravity. The point is that this section help identify what, if anything, needs to be considered during LFT&E.

Para 6 - System Capabilities Required for the Current Increment

- This section is of primary importance during tester review
- Note the focus is on the current increment
- AFI 10-601 also governs this section
    - Increases requirement by adding KSA's (key system attributes)
    - If this section contains an effectiveness requirement that includes reliability it should be commented on. Not having a separate reliability requirement makes it difficult for AFOTEC and DOT&E to assess suitability. It can also confound the T&E results by allowing a low reliability to be compensated by high effectiveness. This can cause problems down the road leading to additional testing.
    - Review this section making sure only a minimum number of requirements are KPPs. Failing KPPs in test is bad for the program and threshold/objective requirements leave the contractor with room for trade space, which will impact testing. [2]
    - The Requirements Correlation Tables (RCT's) must include paragraph numbers and rationale with analytical references.
    - Attributes should be validated for testability and check rationale with analytical reference.
        - Review this section for requirements that include All, 100%, or even 99.9% as these have proven difficult if not impossible to test.
    - Know the difference between KPP / KSA / Attribute (AFI 10-601).
    - Performance attributes apply only to a single increment so ensure testing can be accomplished on current planned increment.
    - Follow on increments require new CDD.

Para 12 - Schedule and IOC/Full Operational Capability (FOC) Definitions

[2] There are now several mandatory KPPs including net-readiness, survivability, and force-protection.

- ❑ Make sure IOT&E/FOT&E is discussed here if appropriate.

Para 14 - Other System Attributes

- ❑ Check this section for additional testability of System Attributes.

Capability Production Document (CPD). Like the CDD, AF/XOR may direct a MAJCOM/Agency to develop a CPD although sponsors will, in most cases, have developed an ICD Stage I/ICD Stage II, and CDD prior to development of CPD.

CPD Strategy Development Initiated. AF/XOR approves the requirements strategy before initiating the CPD. The requirements strategy lays the foundation for CPD development and supports the Production and Deployment Phase for a single increment. The sponsor, along with operators, continues the collaboration initiated in CDD development with Air Force acquisition, test, and logistics communities (and other appropriate SMEs). Strategy development includes the sponsor's interaction with other Services and agencies (as required), including the appropriate lead FCB working group. The sponsor applies lessons learned, findings of design reviews, test results to refine performance attributes for a single increment. The requirements strategy establishes operational performance expectations for the capability to be produced and fielded.

CPD Development.

- CPD Template

  Again, the HPT is the preferred method to develop a CPD and is used unless waived by AF/XOR at the RSR.

  The CPD supports Milestone C and is developed after the Design Readiness Review (DRR). The CPD must be approved before Low Rate Initial Production (LRIP) and Initial Operational Test & Evaluation (IOT&E).

Capability Production Document
(CPD)
Review Procedures/Checklist and
Lessons Learned

Para 6 - System Capabilities Required for the Current Increment

- This section is of primary importance during tester review
- Note: The CPD must be written to the current increment – that is, regardless of how the program has defined it, the portion of the program about which the fielding or production decision is about. So if the program says they are fielding Spiral 1.0 of Increment 1, the CPD must be written strictly to Spiral 1.0 as this is what will be operationally tested.
- AFI 10-601 also governs this section
  - ❑ Increases requirement by adding KSA's (key system attributes)

- If this section contains an effectiveness requirement that includes reliability it should be commented on. Not having a separate reliability requirement makes it difficult for AFOTEC and DOT&E to assess suitability. It can also confound the T&E results by allowing a low reliability to be compensated by high effectiveness. This can cause problems down the road leading to additional testing.
- Review this section making sure only a minimum number of requirements are KPPs. Failing KPPs in test is bad for the program and threshold/objective requirements leave the contractor with room for trade space, which will impact testing.[3]
- The Requirements Correlation Tables (RCT's) must include paragraph numbers and rationale with analytical references.
- Attributes should be validated for testability and check rationale with analytical reference.
    - Review this section for requirements that include All, 100%, or even 99.9% as these have proven difficult if not impossible to test.
- Know the difference between KPP / KSA / Attribute (AFI 10-601).
- Performance attributes apply only to a single increment so ensure testing can be accomplished on current planned increment.
- Follow on increments require new CPD.

[3] There are now several mandatory KPPs including net-readiness, survivability, and force-protection.

CHAPTER 10: REVIEWING PROGRAM MANAGEMENT DOCUMENTS

Direction and policy for developing PMDs is in Headquarters Operating Instruction (HOI) 63-1, *Headquarters Air Force Guidance for Preparing Program Management Directives*, 20 Nov 03. This document is designed to point out areas of the PMD that are of special interest to the test community as well as highlight areas that have been problematic in the past.

Section II, Program Information: This table should be checked for accuracy with special attention paid to the following blocks.

- **Item m, System Survivability**: The information in this block must be based on the definition of covered systems or covered product improvement programs as spelled out in Title 10 §2366 and DODI 5000.02. Not all systems and modifications used in combat require survivability testing and could be inadvertently subjected to unnecessary OSD T&E oversight depending on how program direction is worded. Check if the system will be used in combat (as stated in the operational requirements) and consult with the program office and users to be sure if LFT&E rules apply.

- **Item u, OT&E Agency**: This block must identify all organizations that will be conducting operational testing. Each organization listed must have a respective paragraph in Section III that states which kind(s) of "operational testing" will be conducted as defined in AFI 99-103.

  a. At a minimum, AFOTEC will always be listed unless a non-involvement decision has been made in which case the OT&E agency listed will be the organization documented in the AFOTEC non-involvement letter.

  b. For systems in sustainment undergoing multiple modifications, more than one type of operational testing involving more than one operational test organization may be required. In this case each organization should be listed.

Section III, Program Specific Management Direction: This block must separately address each of the acquisition programs or modifications embedded in the PMD. Each program or modification is different and has specific T&E requirements to support various senior leader decisions (e.g., fielding, full-rate production (FRP), continue development, or declare IOC). Lumping all programs or modifications together is unsatisfactory because T&E organizations and responsibilities may vary for each one, and key decisions are different for each one.

The PMD must direct the CDT/TM and the operational tester (i.e., AFOTEC or MAJCOM tester) to co-chair the integrated test team (ITT). The term "test plan working group" is no longer used. Multiple programs or modifications may come under a single ITT (e.g., the same ITT could have charge of all programs in the PMD). The ITT must:

a. Design a T&E strategy that uses integrated testing as much as possible.

b. Write a test and evaluation master plan (TEMP) if required. Small programs or modifications may not require a TEMP.

c. Identify the responsible test organization (LDTO) if one has been selected.

The PMD should state if acquisition programs will use an evolutionary acquisition (EA) strategy. These programs will be in multiple phases of development at the same time, thus a more complex T&E strategy will be required. If the PMD covers multiple increments of a program, the operational tester for each increment must be clearly stated.

Tester involvement must be clearly directed for each program or modification according to AFI 99-103. Simply citing "OT&E" is inadequate because there are numerous types of operational tests that could be conducted. One or more operational test organizations (as spelled out in Section II, item u) may be directed to conduct specific types of operational tests such as IOT&E, QOT&E, FOT&E, MOT&E, OUE, and FDE.

All required testers and their T&E activities must be logically integrated and support each program or modification with the right kinds of tests at the right times for that particular phase of system development. You may need to check the T&E Strategy, TEMP, or individual test plans to figure this out.

Operational testing with an evaluation is required if the system will be fielded or go to FRP. If AFOTEC is not involved, MAJCOMs must be directed to conduct operational testing to support a FRP and/or fielding decision. Check the kinds of decisions supported, and check with AFOTEC about their involvement in each program or modification.

The words "as required" must **not** be used when directing testing because they are vague and promote indecision. A specific kind of operational testing must be clearly directed. Using "as required" could allow test organizations to "move the goal posts after the ball is kicked."

If used, the LDTO must be listed. LDTOs must be government T&E organizations identified in accordance with AFI 99-103 and MAJCOM guidance.

Participating Test Organizations (PTOs) and their roles must be listed.

Distribution List: Check the distribution lists to ensure AF/TEP and other testers that have responsibilities listed in the PMD are included.

Remember, the overarching goal is to ensure that T&E logically supports the acquisition program or modification, and that T&E is as integrated as possible.

CHAPTER 11: INTEGRATED TEST TEAMS (ITT) AND ITT CHARTERS

The ITT is the overarching test management team that is co-chaired by the CDT/TM and the Operational Test Director. It is responsible for the T&E grand strategy that supports the program's acquisition process. The ITT must begin building the overall T&E strategy during the Concept Refinement and Technology Development Phases well before a program is officially started at Milestone B. If these early phases are bypassed, the ITT must come together as soon as the new program is identified so that key T&E decisions can be made and strategies developed. Overarching direction and policy for forming ITTs and developing ITT charters is in AFI 99-103, *Capabilities-Based Test and Evaluation*.

The ITT charter must establish the working relationships among people so they can work together efficiently as a team. It must create a partnership between the program manager (PM), the testers, and other organizations needed to support T&E and the acquisition program. The ITT charter should be short (i.e., 10 pages or less), focused on organizational responsibilities, non-controversial, and free of extraneous material that could delay coordination. It should be completed as soon as the need for an ITT becomes evident so the ITT can begin working together as early as possible. All potential organizations should assign temporary members to early ITT meetings until permanent members are designated.

ITT Charters, Structure, and Members. A formal, signed ITT charter will describe ITT membership and structure, responsibilities, ITT resources, and the products for which the ITT is responsible. ITT charters must properly balance two highly important mandates: ensuring each organization's needs are properly represented; and providing the acquisition program with responsive, high quality T&E inputs. Since acquisition programs vary so much in size, content, and complexity, the ITT's membership and structure must be tailored to fit the particular conditions of individual programs, organizations, and other contingencies. Well-written ITT charters will help achieve all these factors.

ITT Structure. The ITT will tailor their membership, structure, and protocols as necessary to help ensure program success. To achieve flexibility and efficiency, the ITT may decide to use **two levels,** such as an Executive Level consisting of O-6s and GS-15s and a Working Group Level consisting of necessary experts to fulfill ITT. See Figure A5.1 below. Only the key organizations would be at the Executive Level as agreed at the initial ITT startup meeting. Such a tiered approach would permit the Executive Level to meet less frequently (i.e., every six months) while allowing the Working Group level to handle day-to-day operational and tactical matters with more frequent meetings (i.e., monthly). One or both of these management levels may be most appropriate to deal with the issues and tasks at hand. For example, the Executive Level should handle policy and strategic issues, document approvals, etc., and the Working Group Level would write test documents such as the TEMP. Figure 11.1 shows a notional ITT structure.

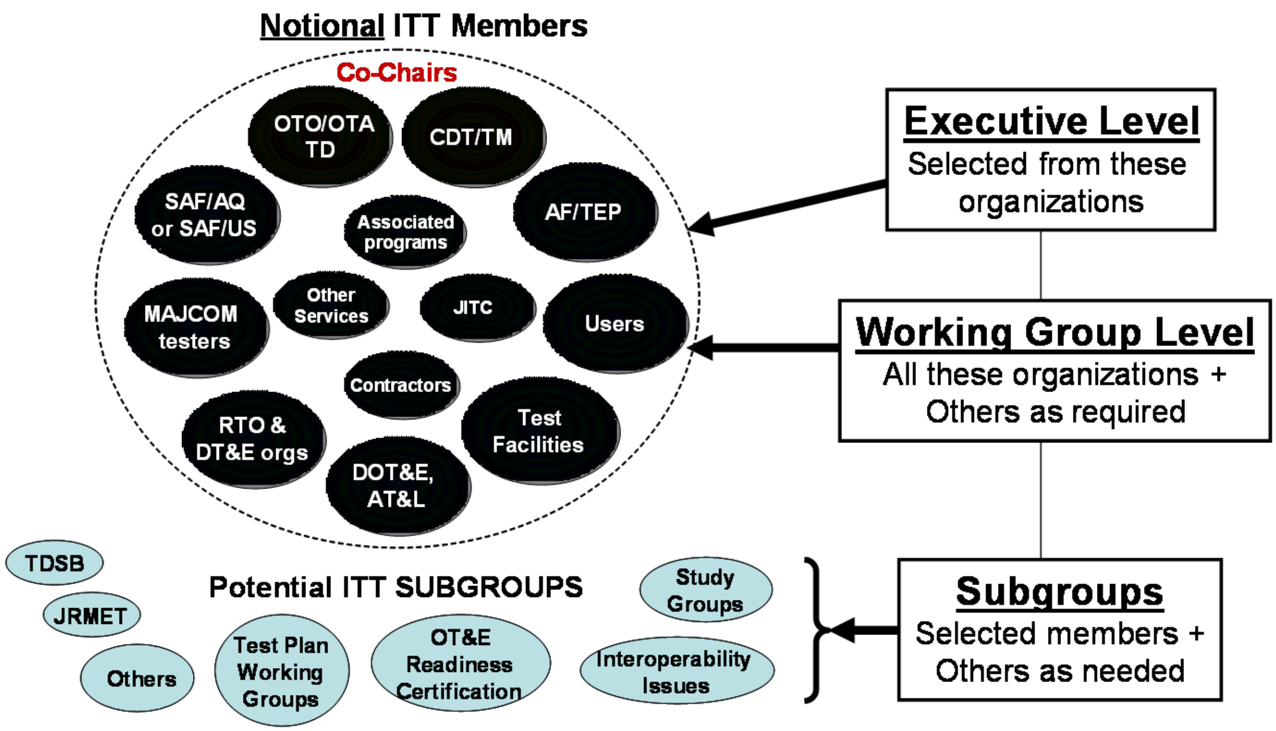


Figure 11.1: Notional ITT Structure and Members

The Executive Level. A limited number of key organizations to provide overall direction and sign the ITT charter at the O-6 or GS-15 level. These signatures should not exceed the O- 6 or GS-15 level. Limiting the level and number of signatories to the charter will speed up coordination and approval and focus on the key organizations. Signing above this level may be counter-productive if imbalances are created among Executive Level members and/or coordination is delayed.

The Working Group Level. Working Group Level organizations are not signatories on the ITT charter, but will be accountable to and receive their authority through their Executive Level member (if they have one). This does not mean Working Group Level members are less valuable. On the contrary, they bring special skills and technical expertise to the ITT that Executive Level members may not have.

Subgroups. The ITT may create any number of permanent or temporary subgroups to accomplish specific tasks. Test integrated product teams (TIPT) can be assigned for writing test plans, combined test forces (CTF) for test conduct, working groups for test readiness reviews and test data reviews, and study groups for specific issues and problems.

- The ITT charter should very briefly describe which subgroups will be formed (if known) and outline their functions. Do not embed charters for any subgroups in the ITT charter because they could unnecessarily complicate and delay the ITT charter's coordination. Additional subgroups are permitted as the circumstances warrant.

- The ITT construct is not the same as the CTF construct. Depending on the program, some CTFs may operate more or less independently from the ITT, but the ITT and CTF must work collaboratively in these cases. The Joint Strike Fighter CTF is an example. Some ITT members may also be part of a CTF or TIPT due to the limited number of personnel available. Management of T&E from an ITT perspective will be different than conduct of T&E at the CTF level.

Member Organizations. ITT membership (at the Executive Level and Working Group Level) may vary depending on program needs. The ITT should include expertise from the organizations listed in AFI 99-103, paragraph 4.4.3. Potential member organizations are listed at the 2-letter level only to indicate the type of expertise from anywhere within that organizational context. The intent is to allow these organizations flexibility to select their best qualified representatives. The ITT does not require anyone higher than an O-6 or GS-15 to attend. Thus, AFOTEC may select their Executive Level representative (an O-6 or GS-15) from AFOTEC/XP or XO or AFOTEC, Det 5/CC, and select someone else for the Working Group Level. HQ USAF/TE will select an O-6 for the Executive Level and an AF/TEP action officer for the Working Group Level. Include or exclude organizations as appropriate.

Multiple Programs Under a Single ITT. To create greater efficiency, a single ITT may cover a number of related acquisition and/or sustainment programs. For example, an aircraft system may have many ongoing modifications from small to very large, each of which requires separate test plans and test activities. A single ITT could cover all these sub- programs to ensure more efficient allocation and scheduling of limited T&E resources. Another example could be to charter an ITT to manage T&E for a family of systems or a group of similar information technology (IT) systems. If these programs have similar or interoperable components, then a single "umbrella ITT" may be the proper venue for managing T&E for all. In each case, the ITT's span of control should be tailored to its capacity for overseeing the T&E grand strategy and assigning responsibilities to subgroups. To conserve resources, attendees should consider using video or telephone conferencing capabilities when appropriate.

Avoid Duplication. The ITT Charter should not duplicate the contents and direction cited in other T&E documents such as the TEMP, test plans, or AFIs, but only reference them for the sake of brevity. The primary focus should be on items or tasks that are generally not suitable for a TEMP, test plan, or AFI. It must address the tasks that AFI 99-103 and the DOD 5000-series leave to the discretion of PMs, testers and others, and any other unique attributes of the program. The ITT charter should not be another list of previously published roles and responsibilities, nor should it include details about how the system will be tested. Extraneous materials, such as lists of items beyond the discretionary control of ITT members, should not be attached to ITT charters. If there is any doubt about including any material in the charter, leave it out and place it in more appropriate documents.

Conflict Resolution. When ITT members disagree on problems or issues, they should broker agreements in the spirit of compromise for the good of the program. If an agreement cannot be readily reached, the conflict resolution flowchart shown below in Figure 11.2 should be used. If the co-chairs cannot resolve the issue within the ITT, the issue should be raised to the organizational leadership for resolution. In all instances the ITT will comply with governing guidance and directives.

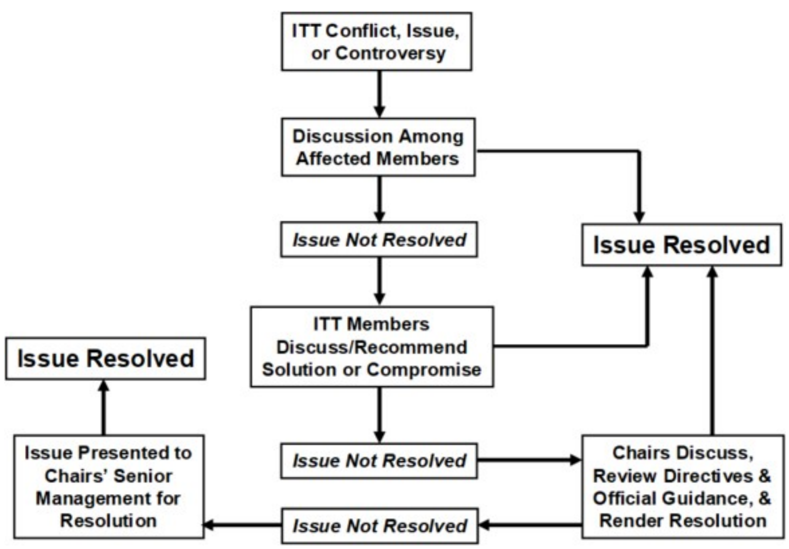


Figure 11.2 Conflict Resolution Flowchart

Recommended ITT Charter Outline. The following outline covers the primary subject areas for an ITT charter. The charter should concisely cite information necessary to understand the program and how the ITT will support that program. This list is not all-inclusive and may be modified as necessary.

ITT Charter for
the XXX
Program
(Recommended)

Introduction.

Program(s) Covered. List the program(s) the ITT will oversee and give a brief history of the ITT since program inception. If this is a standing ITT, list when additional programs were added to the original ITT charter.

Authority. Cite the document(s) directing formation of the ITT and its responsibilities such as:

- AFI 99-103
- Acquisition Decision Memorandum or other documents directing formation

Program Description. Briefly describe the program(s) covered by the charter.

- **Other Key Program Information.** Acquisition category (ACAT); on OSD T&E Oversight List; etc.
- **Acquisition Strategy Overview.** Briefly describe the acquisition strategy and how the test and evaluation (T&E) strategy supports it.

ITT Mission, Scope, and Overarching Goals. Outline the reasons for having the ITT (reference AFI 99-103, paragraph 3.14).

ITT Membership and Responsibilities. List the ITT member organizations and how they will support development of the T&E strategy and other T&E matters. Build on the list of responsibilities in AFI 99-103 (but don't repeat them), and list those that are unique to this ITT and program.

- **System Program Office (SPO).** CDT or Test Manager if the program is not an ACAT.
- **Operational Test Organization.** This is AFOTEC unless it has been determined (IAW AFI 99-103) they will not be involved in conducting operational testing. AFOTEC will be involved with all ACAT I, ACAT II, and OSD Oversight programs until the AFOTEC Involvement Determination Process determines otherwise. Operational testers from the other Services and relevant Air Force MAJCOMs should also attend.
- **Lead Developmental Test & Evaluation Organization (LDTO).** DT&E community representatives should attend early meetings until the LDTO is formally designated. Various sources are available such as the appropriate Product or Logistics Center Test Authority (CTA), or as designated by

AFMC/A3F, or the Test Representative from the DT&E test organization most likely to conduct government DT&E.

- **Associated System Program Offices.** List any SPOs for associated systems that must be interoperable with the chartered system(s).

- **MAJCOM Operational Test Organization(s).** MAJCOM attendance is required if AFOTEC is not the designated operational tester, and may attend at their discretion if AFOTEC is the designated operational tester. As the program progresses, a transition from AFOTEC to MAJCOM should be anticipated and planned for.

- **Participating Test Organizations (PTO).** Describe how JITC, AFC2ISRC, etc., will support the program's T&E activities.

- **Operational User(s).** Assist with clarification of JCIDS documents and the development of CONOPS, strategies, and other operational plans.

- **HQ USAF Offices.** Describe how AF/TE, AF/XO, SAF/AQ or SAF/US, SAF/XC, and others as required will support the ITT's efforts.

- **OSD Offices.** Describe how DOT&E, USD(AT&L) and others as required will support the ITT's efforts.

Formation of Sub-Groups. Briefly describe the ITT subgroups that will support or conduct T&E such as CTFs, study groups, writing teams, data scoring boards, certification boards, etc. These groups will likely draw upon the same members and organizations as the ITT, but will have distinctly different functions and may have their own charters. Do not embed these charters within the ITT charter.

Administrative Matters.

- Frequency of Meetings
- Attendance
- Meeting Minutes
- Action Items

ITT Charter Updates. Review the ITT charter for currency soon after each milestone or major decision review, for each new increment that is started, and when additional associated systems are added to or taken from the ITT.

Coordination and Signatures. The level of signature on the ITT charter should generally be at the O-6 or GS-15 level.

Conflict Resolution. Describe how conflicts will be resolved according to Figure A5.2.

APPENDIX 1:
GLOSSARY

A common understanding of terms is essential to effectively implement instructions and guidance. In some cases, definitions from multiple sources are offered where they may be of value. Italicized words and notes in brackets are not part of the definition and are offered only for clarity.

Acquisition Category (ACAT)—Acquisition categories determine the level of review, decision authority, and applicable T&E policies and procedures. They facilitate decentralized decision making and execution, and compliance with statutorily imposed requirements. See DODI 5000.02, Enclosure 2 for details.

Advanced Concept Technology Demonstration—A demonstration of the military utility of a significant new technology and an assessment to clearly establish operational utility and system integrity. (CJCSI 3170.01D)

Availability (Ao)—A measure of the degree to which an item is in the operable and committable state at the start of a mission when the mission is called for at an unknown (random) time. (*Defense Acquisition Guidebook*)

Capability Based Testing—A mission-focused methodology of verifying that a capabilities solution will enable operations at an acceptable level of risk. Capabilities-oriented evaluations are emphasized throughout system testing in addition to traditional evaluations of system performance measured against specification-like requirements. It requires understanding Concept of Operations and involves developing T&E strategies and plans to determine whether a capability solution option merits fielding.

Combined Testing—See Integrated Testing.

Covered System—1. A vehicle, weapon platform, or conventional weapon system that includes features designed to provide some degree of protection to users in combat; and this is a major system within the meaning of that term in Title 10 §2302(5). (Title 10 §2366). **2.** All categories of systems or programs identified in Title 10 §2366 as requiring live fire test and evaluation. In addition, non-traditional systems or programs that do not have acquisition points referenced in Title 10 §2366, but otherwise meet the statutory criteria. ***NOTE*:** The definitions of "covered system," "major munitions program," and "covered product improvement program" are encompassed in the single

DoD term "covered system." (*Defense Acquisition Guidebook*, Appendix 3, which includes conventional munitions programs for which more than 1,000,000 rounds are planned to be acquired; or a modification to a covered system that is likely to affect significantly the survivability or lethality of such a system.)

Covered Product Improvement Program—See Covered System.

Critical Operational Issue (COI)—1. Operational effectiveness and operational suitability issues (not parameters, objectives, or thresholds) that must be examined during operational testing to determine the system's capability to perform its mission (paraphrased from DAU's *Test and Evaluation Management Guide).* **2.** A key question that must be examined in operational test and evaluation to determine the system's capability to perform its mission. Testers normally phrase a COI as a question to be answered in evaluating a system's operational effectiveness or suitability.

Critical Technical Parameter (CTP)—Measurable critical system characteristics that, when achieved, allow the attainment of operational performance requirements. They are technical measures derived from operator requirements. Failure to achieve a critical technical parameter should be considered a reliable indicator that the system is behind in the planned development schedule or will likely not achieve an operational requirement (paraphrased from *Defense Acquisition Guidebook*).

Cyber test—Characterizes and evaluates systems and sub-systems operating in the cyberspace domain, and the access pathways of such systems.

Cyberspace—A domain characterized by the use of electronics and the electromagnetic spectrum to store, modify, and exchange data via networked systems and associated physical infrastructures.

Cybersecurity Testing —Testing of the systems' and sub-systems' ability to protect or defend against a cyber-attack. Cybersecurity testing focuses on identifying and eliminating or mitigating system cyber vulnerabilities.

Cyber Resiliency Testing — Evaluates a system's ability to meet operational requirements while subjected to cyber-attacks

Dedicated Operational Testing—Operational test and evaluation that is conducted independently from contractors, developers, and operators and used to support production or fielding decisions.

Deficiency Report (DR)—The report used to identify, document, and track system deficiency or enhancement data while a system is in advanced development, operational test, or operational transition.

- **Category I** DRs are those that could cause death, severe injury, severe occupational illness, major loss or damage, or directly restrict combat or operational readiness if left uncorrected.
- **Category II** DRs are those that do not meet the criteria of a Cat I DR. They are attributable to errors in workmanship, nonconformance to specifications, drawing standards, or other technical requirements; or identify a problem for potential improvement or enhancement.
- **Enhancements** are a type of Category II DR that identifies conditions that complement, but are not absolutely required for successful mission accomplishment. The recommended condition, if incorporated, will improve a system's operational effectiveness or suitability (paraphrased from TO 00-35D-54).

Deployment—1. The movement of forces within operational areas. **2.** The relocation of forces and materiel to desired operational areas. Deployment encompasses all activities from origin or home station through destination. (JP 1-02)

Developmental Test and Evaluation (DT&E)—Test and evaluation conducted to evaluate design approaches, validate analytical models, quantify contract technical performance and manufacturing quality, measure progress in system engineering design and development, minimize design risks, predict integrated system operational performance (effectiveness and suitability) in the intended environment, and identify system problems (or deficiencies) to allow for early and timely resolution. DT&E includes contractor testing and is conducted over the life of the system to support acquisition and sustainment efforts. (*Defense Acquisition Guidebook*)

Early Operational Assessment (EOA)—An operational assessment (OA) conducted before MS B. An EOA assesses the design approach sufficiently early in the acquisition process to assure it has the potential to fulfill operator requirements. See Operational Assessment.

Evaluation Criteria—Standards by which the accomplishment of required technical and operational effectiveness and/or suitability characteristics, or resolution of operational issues, may be addressed. (*Defense Acquisition Guidebook*)

Evolutionary Acquisition—Evolutionary acquisition is the preferred DoD strategy for rapid acquisition of mature technology for the user. An evolutionary approach delivers capability in increments, recognizing, up front, the need for future capability improvements. The objective is to balance needs and available capability with resources, and to put capability into the hands of the user quickly. The success of the strategy depends on consistent and continuous definition of requirements, and the maturation of technologies that lead to disciplined development and production of systems that provide increasing capability towards a materiel concept. The approaches to achieve evolutionary acquisition require close collaboration between the user, tester, and developer. (DODI 5000.02) They include:

- **Spiral Development**—In this process, a desired capability is identified, but the end-state requirements are not known at program initiation. Those requirements are refined through demonstration and risk management; there is continuous user feedback; and each increment provides the user the best possible capability. The requirements for future increments depend on feedback from users and technology maturation. (DODI 5000.02)

- **Incremental Development**—In this process, a desired capability is identified, an end-state requirement is known, and that requirement is met over time by developing several increments, each dependent on available mature technology. (DODI 5000.02)

Fielding—The decision to acquire and/or release a system to operators in the field.

First Article Test (FAT)—Production testing that is planned, conducted, and monitored by the materiel developer. FAT includes pre-production and initial production testing conducted to ensure that the contractor can furnish a product that meets the established technical criteria. (DAU's *Test and Evaluation Management Guide*)

Follow-on Operational Test and Evaluations (FOT&E)—The continuation of IOT&E or QOT&E activities past the full-rate production decision. FOT&E answers specific questions about unresolved COIs or completes areas not finished during the IOT&E or QOT&E. It ensures the initial system acquisition process is complete.

Force Development Evaluation (FDE)—The operational test and evaluation of fielded, operational systems during the sustainment portion of the system life cycle after acceptance for operational use. The focus is on maintaining or upgrading operational systems after the initial acquisition process is complete. An FDE also supports acquisition of MAJCOM-managed systems.

Foreign Comparative Test (FCT)—A T&E program centrally managed by OSD which provides funding for U.S. T&E of selected equipment items and technologies developed by allied or friendly countries when such items or technologies are identified as having good potential to satisfy valid DoD requirements. (DoD 5000.3-M-2)

Full-Up System-Level Testing—Testing that fully satisfies the statutory requirement for "realistic survivability testing" or "realistic lethality testing" as defined in Title 10 §2366. (*Defense Acquisition Guidebook*, Appendix 3)

Increment—A militarily useful and supportable operational capability that can be effectively developed, produced or acquired, deployed, and sustained. Each increment of capability will have its own set of threshold and objective values set by the user. (CJCSI 3170.01D and AFI 10- 601) ***NOTE***: An increment may contain multiple spirals. Generally, only increments are fielded according to DODI 5000.02, CJCSI 3170.01D, and AFI 63-101.

Information Support Plan (ISP)—[The plan] used by program authorities to document the IT and NSS needs, objectives, interface requirements for all non-ACAT and fielded programs. (CJCSI 6212.01C)

Initial Operational Test and Evaluation (IOT&E)—See Operational Test and Evaluation.

Integrated Testing—Any combination of two or more types of testing used to achieve greater test efficiency, reduced cost, and schedule savings without compromising the objectives and needs of the participating test organizations.

Integrated Test Team (ITT)—A cross-functional team of empowered representatives from multiple disciplines and organizations and co-chaired by the Chief Developmental Tester/Test Manager and the Operational Test Director. The ITT is responsible for developing the T&E strategy and TEMP, assisting the acquisition community with T&E matters, and guiding the development of integrated test plans. There is one ITT for each acquisition program.

Joint Test and Evaluation (JT&E)—An OSD-sponsored T&E program conducted among more than one military Service to provide T&E information on combat operations issues and concepts. JT&E does not support system acquisition. (DoDD 5010.41)

Lead Developmental Test & Evaluation Organization (LDTO)—The lead government developmental test organization on the ITT that is qualified to conduct and responsible for overseeing DT&E.

Lethality—The capability of a munition or directed energy weapon to cause damage that will cause the loss or a degradation in the ability of a target system to complete its designated mission(s). (*Defense Acquisition Guidebook*, Appendix 3)

Live Fire Test and Evaluation (LFT&E)—The firing of actual weapons (or surrogates if actual weapons are not available) at components, subsystems, sub-assemblies, and/or full-up, system- level targets or systems to examine personnel casualties, system vulnerabilities, or system lethality; and the evaluation of the results of such testing. (*Defense Acquisition Guidebook*, Appendix 3)

Logistics Support Elements—1. A composite of all support considerations necessary to ensure the effective and economical support of a system for its life cycle. It is an integral part of all other aspects of system acquisition and operation. (JP 1-02) ***NOTE:*** The ten logistics support elements are: maintenance planning; manpower and personnel; supply support; support equipment; technical data; training and training support; computer resources support; facilities; packaging, handling, storage, and transportation; and design interface. Formerly known as Integrated Logistics Support. (AFI 10-602)

Logistics Supportability—The degree to which the planned logistics support allows the system to meet its availability and wartime usage requirements. Planned logistics support includes the following: test, measurement, and diagnostic equipment; spare and repair parts; technical data; support facilities; transportation requirements; training; manpower; and software. (*Defense Acquisition Guidebook*)

Logistics Test and Evaluation—The test methodology, criteria, and tools for evaluating and analyzing the ten logistics support elements as they apply to a system under test. The objective is to influence the design through applying the logistics support elements as early as possible in the acquisition cycle. This testing integrates the evaluation and analysis efforts of RM&A, human factors engineering, and logistics test, and is an integral part of the DT&E report.

Lot Acceptance Test (LAT)—A test based on a sampling procedure to ensure that the product retains its quality. No acceptance or installation should be permitted until this test for the lot has been successfully completed. (*Glossary, Defense Acquisition Acronyms and Terms*, and DAU's *Test and Evaluation Management Guide*)

Low-Rate Initial Production (LRIP)—Production of the system in the minimum quantity necessary (1) to provide production-configured or representative articles for operational tests pursuant to §2399; (2) to establish an initial production base for the system; and (3) to permit an orderly increase in the production rate for the system sufficient to lead to full-rate production upon the successful completion of operational testing. ***NOTE:*** The LRIP quantity should not exceed 10 percent of the total number of articles to be produced as determined at the milestone B decision. (Title 10 §2400)

Maintainability—The capability of an item to be retained in or restored to a specified condition when maintenance is performed by personnel having specified skill levels, using prescribed procedures and routines, at each prescribed level of maintenance and repair. (*Defense Acquisition Guidebook*)

Major Munitions Program—See Covered System.

Measurable—Having qualitative or quantitative attributes (e.g., dimensions, velocity, capabilities) that can be ascertained and compared to known standards. (See Testable.)

Measure of Effectiveness (MOE)—A qualitative or quantitative measure of a system's performance or a characteristic that indicates the degree to which it performs the task or meets a requirement under specified conditions. MOEs should be established to measure the system's capability to produce or accomplish the desired result.

Measure of Performance—A quantitative measure of a system's capability to accomplish a task. Typically in the area of physical performance (e.g., range, velocity, throughput, payload).

Military Utility—The military worth of a system performing its mission in a competitive environment including versatility (or potential) of the system. It is measured against the operational concept, operational effectiveness, safety, security, and cost/worth. Military utility estimates form a rational basis for making management decisions. (*Glossary, Defense Acquisition Acronyms and Terms*)

Mission Based Cyber Risk Assessments (**MBCRA) —** Key component to developing a cyber- survivable system. The MBCRA process helps identify, assess, and prioritize cyber-risks based on potential impacts to the system's capability to perform its operational mission(s).

Multi-Service—Involving two or more military Services or DoD components.

Multi-Service Operational Test and Evaluation (MOT&E)—OT&E conducted by two or more Service OTAs for systems acquired by more than one Service. MOT&E is conducted according to the T&E directives of the lead OTA, or as agreed in a memorandum of agreement between the participants.

Objective—An operationally significant increment above the threshold. An objective value may be the same as the threshold when an operationally significant increment above the threshold is not significant or useful. (AFI 10-601)

Operational Assessment (OA)—An analysis of potential operational effectiveness and suitability made by an independent operational test activity, with operator support as required, on other than production systems. The focus of an operational assessment is on significant trends noted in development efforts, programmatic voids, areas of risk, adequacy of requirements, and the ability of the program to support adequate operational testing. Operational assessments may be made at any time using technology demonstrators, prototypes, mockups, engineering development models, or simulations, but will not substitute for the dedicated OT&E [sic] necessary to support full production decisions. (*Defense Acquisition Guidebook*)

Operational Effectiveness—Measure of the overall ability to accomplish a mission when used by representative personnel in the environment planned or expected for operational employment of the system considering organization, doctrine, tactics, supportability, survivability, vulnerability and threat. (CJCSI 3170.01D)

Operational Suitability—The degree to which a system can be placed and sustained satisfactorily in field use with consideration given to availability, compatibility, transportability, interoperability, reliability, wartime usage rates, maintainability, safety, human factors, habitability, manpower, logistics, supportability, logistics supportability, natural environmental effects and impacts, documentation, and training requirements. (CJCSI 3170.01D)

Operational Test Agency (OTA)—An independent agency reporting directly to the Service Chief that plans and conducts operational tests, reports results, and provides evaluations of effectiveness and suitability on new systems. (DoDD 5000.1) ***NOTE***: Each Service has one designated OTA: The Air Force has the Air Force Operational Test and Evaluation Center (AFOTEC). The Navy has the Operational Test and Evaluation Force (OPTEVFOR). The Army has the Army Test and Evaluation Command (ATEC). The Marine Corps has the Marine Corps Operational Test and Evaluation Agency (MCOTEA).

Operational Test and Evaluation (OT&E)—**1.** The field test, under realistic combat conditions, of any item of (or key component of) weapons, equipment, or munitions for the purpose of determining the effectiveness and suitability of the weapons, equipment, or munitions for use in combat by typical military users; and the evaluation of the results of such test. (Title 10 §139(a)(2)) **2.** Testing and evaluation conducted in as realistic an operational environment as possible to estimate the prospective system's operational effectiveness and operational suitability. In addition, OT&E provides information on organization, personnel requirements, doctrine, and tactics. It may also provide data to support or verify material in operating instructions, publications, and handbooks.

Operational Testing—A generic term describing the test and evaluation options and levels of effort available to an operational test organization.

Operational Utility Evaluation (OUE)—OUEs are evaluations of military capabilities conducted to demonstrate or validate new operational concepts or capabilities, upgrade components, or expand the mission or capabilities of existing or modified systems.

Operator—Refers to the operating command which is the primary command operating a

system, subsystem, or item of equipment. Generally applies to those operational commands or organizations designated by Headquarters, US Air Force to conduct or participate in operations or operational testing, interchangeable with the term "using command" or "user." In other forums the term "warfighter" or "customer" is often used. (AFI 10-601)

Oversight—Senior executive-level monitoring and review of programs to ensure compliance with policy and attainment of broad program goals.

Oversight Program—A program on the OSD T&E Oversight List for DT&E, LFT&E, and/or OT&E. The list includes all ACAT I (MDAP) programs, ACAT II (major system) programs, and any other programs selected for OSD T&E oversight. These programs require additional documentation and have additional review, reporting, and approval requirements.

Participating Test Organization (PTO)—Any test organization required to support a lead test organization by providing specific T&E data or resources for a T&E program or activity.

Pre-Production Qualification Test (PPQT)—The formal contractual tests that ensure design integrity over the specified operational and environmental range. These tests usually use prototype or pre-production hardware fabricated to the proposed production design specifications and drawings. Such tests include contractual reliability and maintainability demonstration tests required prior to production release. (*Glossary, Defense Acquisition Acronyms and Terms*, and DAU's *Test and Evaluation Management Guide*)

Production Acceptance Test and Evaluation (PAT&E)—Test and evaluation of production items to demonstrate that items procured fulfill requirements and specifications of the procuring contract or agreements. (DAU's *Test and Evaluation Management Guide*)

Production Qualification Test (PQT)—A technical test conducted prior to the full rate production decision to ensure the effectiveness of the manufacturing processes, equipment, and procedures. These tests are conducted on a number of samples taken at random from the first production lot, and are repeated if the manufacturing process or design is changed significantly, or when a second source is brought on line. (*Glossary, Defense Acquisition Acronyms and Terms*, and DAU's *Test and Evaluation Management Guid*e)

Program Manager (PM)—**1**. The designated individual with responsibility for and authority to accomplish program objectives for development, production, and sustainment to meet the user's operational needs. The PM shall be accountable for credible cost, schedule, and performance reporting to the MDA. (DoDD 5000.1) **2**. Applies collectively to system program directors, product group managers, single managers, acquisition program managers, and weapon system managers. Operating as the single manager, the PM has total life cycle system management authority. ***NOTE***: This AFI uses the term "PM" for any designated person in charge of acquisition activities prior to MS A (i.e., before a technology project is officially designated an acquisition program).

Prototype—1. A model suitable for evaluation of design, performance, and production potential. (JP 1-02) ***NOTE:*** The Air Force uses prototypes during development of a technology or acquisition program for verification or demonstration of technical feasibility. Prototypes may not be representative of the final production item.

Qualification Operational Test and Evaluation (QOT&E)—A tailored type of IOT&E performed on systems for which there is little to no RDT&E-funded development effort. Commercial-off-the-shelf (COTS), non-developmental items (NDI), and government furnished equipment (GFE) are tested in this manner.

Qualification Test and Evaluation (QT&E)—A tailored type of DT&E for which there is little to no RDT&E-funded development effort. Commercial-off-the-shelf (COTS), non- developmental items (NDI), and government furnished equipment (GFE) are tested in this manner.

Recoverability—Following combat damage, the ability to take emergency action to prevent loss of the system, to reduce personnel casualties, or to regain weapon system combat mission capabilities. (*Defense Acquisition Guidebook*, Appendix 3)

Reliability—The capability of a system and its parts to perform its mission without failure, degradation, or demand on the support system. (*Defense Acquisition Guidebook*)

Research, Development, Test and Evaluation (RDT&E)—The type of funding appropriation (3600) intended for research, development, test and evaluation efforts. (DoD 7000.14-R, Vol 2A, and AFI 65-601, Vol I) ***NOTE****:* The term "research and development" (R&D) broadly covers the work performed by a government agency or the private sector. "Research" is the systematic study directed toward gaining scientific knowledge or understanding of a subject area. "Development" is the systematic use of the knowledge and understanding gained from research for the production of useful materials, devices, systems, or methods. RDT&E includes all supporting test and evaluation activities.

Risk—1. A measurable probability of consequence associated with a set of conditions or actions. (*Glossary, Defense Acquisition Acronyms and Terms*) **2.** Probability and severity of loss linked to hazards. (JP 1-02) **3**. A subjective assessment made regarding the likelihood or probability of not achieving a specific objective by the time established with the resources provided or requested. It also refers to overall program risk. (*Defense Acquisition Guidebook*)

Seamless Verification— A concept for structuring test and evaluation (T&E) to more effectively support the requirements and acquisition processes so new capabilities are brought to operators more quickly. Seamless verification promotes using integrated testing procedures coupled with tester collaboration in early requirements definition and system development activities. It shifts T&E away from the traditional "pass-fail" model to one of providing continuous feedback and objective evaluations of system capabilities and limitations throughout system development.

Specification—A document intended primarily for use in procurement which clearly and accurately describes the essential technical requirements for items, materials, or services, including the procedures by which it will be determined that the requirements have been met.

Specifications may be prepared to cover a group of products, services, or materials, or a single product, service, or material, and are general or detail specifications. (*Glossary, Defense Acquisition Acronyms and Terms*)

Spiral—One subset or iteration of a development program within an increment. Multiple spirals may overlap or occur sequentially within an increment. ***NOTE***: Generally, spirals are not fielded according to DODI 5000.02, CJCSI 3170.01D, and AFI 63-101.

Survivability—The capability of a system and crew to avoid or withstand a man-made hostile environment without suffering an abortive impairment of its ability to accomplish its designated mission. Survivability consists of susceptibility, vulnerability, and recoverability. (*Defense Acquisition Guidebook*, Appendix 3)

Susceptibility—The degree to which a weapon system is open to effective attack due to one or more inherent weaknesses. (Susceptibility is a function of operational tactics, countermeasures, probability of enemy fielding a threat, etc.) Susceptibility is considered a subset of survivability. (*Defense Acquisition Guidebook*, Appendix 3)

Sustainment—1. The provision of personnel, logistic, and other support required to maintain and prolong operations or combat until successful accomplishment or revision of the mission or of the national objective. (JP 1-02) **2.** The Service's ability to maintain operations once forces are engaged. (AFDD 1-2) **3.** Activities that sustain systems during the operations and support phases of the system life cycle. Such activities include any investigative test and evaluation that extends the useful military life of systems, or expands the current performance envelope or capabilities of fielded systems. Sustainment activities also include T&E for modifications and upgrade programs, and may disclose system or product deficiencies and enhancements that make further acquisitions necessary.

Tactics Development and Evaluation (TD&E)—TD&E is a tailored type of FDE specifically designed to further exploit doctrine, system capabilities, tactics, techniques, and procedures during the sustainment portion of the system life cycle. TD&Es normally identify non-materiel solutions to tactical problems or evaluate better ways to use new or existing systems.

Testable—The attribute of being measurable with available test instrumentation and resources.

NOTE: Testability is a broader concept indicating whether T&E infrastructure capabilities are available and capable of ***measuring*** the parameter. The difference between testable and measurable may indicate a test limitation.

Some requirements may be ***measurable*** but not ***testable*** due to T&E infrastructure shortfalls, insufficient funding, safety, or statutory or regulatory prohibitions.

Test and Evaluation (T&E)—The act of generating empirical data during the research, development or sustainment of systems, and the creation of information through analysis that is useful to technical personnel and decision makers for reducing design and acquisition risks. The process by which systems are measured against requirements and specifications, and the results analyzed so as to gauge progress and provide feedback.

Test and Evaluation Master Plan (TEMP)—Documents the overall structure and objectives of the T&E program. It provides a framework within which to generate detailed T&E plans and it documents schedule and resource implications associated with the T&E program. The TEMP identifies the necessary developmental, operational, and live-fire test activities. It relates program schedule, test management strategy and structure, and required resources to: COIs; critical technical parameters; objectives and thresholds documented in the requirements document; and milestone decision points. (DAU's *Test and Evaluation Management Guide)*

NOTE**:** Where the word "TEMP" appears in this AFI, the SAMP T&E annex is also implied. The TEMP may be included in a SAMP as a T&E annex.

Test and Evaluation Organization—Any organization whose designated mission includes test and evaluation.

Test and Evaluation Strategy—The overarching integrated T&E plan for the entire acquisition program that describes how operational capability requirements will be tested and evaluated in support of the acquisition strategy. Developed prior to Milestone A, the T&E strategy addresses modeling and simulation, risk and risk mitigation, development of support equipment, and identifies how system concepts will be evaluated against mission requirements, among other things. The T&E strategy is a precursor to the test and evaluation master plan.

Test Deferral—The delay of testing and/or evaluation of a specific critical technical parameter, operational requirement, or critical operational issue to a follow-on increment.

Test Integrated Product Team (TIPT)—Any temporary group consisting of testers and other experts who are focused on a specific test issue or problem. There may be multiple TIPTs for each acquisition program.

Test Limitation—Any condition that hampers but does not preclude adequate test and/or evaluation of a critical technical parameter, operational requirement, or critical operational issue during a T&E program.

Test Team—A group of testers and other experts who carry out integrated testing according to a specific test plan. ***NOTE***: A combined test force (CTF) is one way to organize a test team for integrated testing.

Threshold—A minimum acceptable operational value below which the utility of the system becomes questionable.

User—See Operator.

Verification, Validation and Accreditation (VV&A)—VV&A is a continuous process in the life cycle of a model or simulation as it gets upgraded or is used for different applications. (AFI16-1002)

- ***Verification***: Process of determining that M&S accurately represent the developer's conceptual description and specifications.
- ***Validation***: Rigorous and structured process of determining the extent to which M&S accurately represents the intended "real world" phenomena from the perspective of the intended M&S use.
- ***Accreditation***: The official determination that a model or simulation is acceptable for use for a specific purpose.

Vulnerability—The characteristic of a system that causes it to suffer a definite degradation (loss or reduction of capability to perform its designated mission) as a result of having been subjected to a certain (defined) level of effects in an unnatural (man-made) hostile environment. Vulnerability is considered a subset of survivability. (*Defense Acquisition Guidebook*, Appendix 3)

APPENDIX 2: ACRONYMS

AATC – Air National Guard – Air Force Reserve Test Center

ACAT – Acquisition Category

AEDC – Arnold Engineering Developmental Complex

AFAMS – Air Force Agency for Modeling and Simulation

AFCD – Air Force Capabilities Document

AFLCMC – Air Force Life Cycle Management Center

AFMC – Air Force Materiel Command

AFNWC – Air Force Nuclear Weapons Center

AFOTEC – Air Force Operational Test & Evaluation Center

AFRL – Air Force Research Laboratory

AFSC – Air Force Sustainment Center

AFSIT – Air Force System Interoperability Testing

AFTC – Air Force Test Center

APDP – Acquisition Professional Development Program

ASD – Agile Software Development

AST – Agile Software Test

ATD – Advanced Technology Demonstration

B-LRIP – Beyond Low Rate Initial Production

CBRD – Capabilities-Based Requirements Documents

CCD – Combat Capability Document

CD – Continuous Delivery

CDD – Capabilities Development Document

CDT – Chief Developmental Tester

CI – Continuous Integration

COI – Critical Operational Issue

CONOPS – Concept of Operations

COTS – Commercial-Off-The-Shelf

CPD – Capability Production Document

CSE – Cyber Survivability Endorsement

CSEIG – Cyber Survivability Endorsement Implementation Guide

CTA – Center Test Authority

CTF – Combined Test Force

CTP – Critical Technical Parameter

DAES – Defense Acquisition Executive Summary

DAG – Defense Acquisition Guidebook

DAMIR – Defense Acquisition Management Information Retrieval

DCR – DOTMLPF Change Recommendation

DD (DTE&P) – Deputy Director, Developmental Test, Engineering and Prototyping

DOT&E – Director, Operational Test and Evaluation

DOTMLPF – Doctrine, Organization, Training, Materiel, Leadership and Education, Personnel, and Facilities

DR – Deficiency Report

DRU – Direct Reporting Unit

DSO – DevSecOps

EA – Evolutionary Acquisition

EMSEC – Emission Security

EOA – Early Operational Assessment

EWIR – Electronic Warfare Integrated Reprogramming

FAT – First Article Test

FCT – Foreign Comparative Testing

FDE – Force Development Evaluation

FOT&E – Follow-on Operational Test and Evaluation

FRP – Full Rate Production

GFE – Government Furnished Equipment

HPT – High Performance Team

HWIL – Hardware In-the-Loop

ICD – Initial Capabilities Document

IPT – Integrated Product Team

ISP – Information Support Plan

ITEA – International Test and Evaluation Association

ITT – Integrated Test Team

JCD – Joint Capabilities Document

JCIDS – Joint Capabilities Integration and Development System

JCTD – Joint Capability Technology Demonstration

JOTT – Joint Operational Test Team

JROC – Joint Requirements Oversight Council

JSE – Joint Simulation Environment

JT&E – Joint Test and Evaluation

JTIC – Joint Interoperability Test Command

LAT – Lot Acceptance Test

LDTO – Lead Developmental Test and Evaluation Organization

LFT&E – Live Fire Test & Evaluation

LRIP – Low Rate Initial Production

MBCRA – Mission Based Cyber Risk Assessment

MD – Mission Data

MDA – Milestone Decision Authority

MDAP – Major Defense Acquisition Program

MOE – Measures of Effectiveness

MOT&E – Multi-Service Operational Test and Evaluation

MRAP-C – Mission-based Risk Assessment Process for Cyber

M&S – Modeling and Simulation

MSCO – Modeling and Simulation Coordination Office

MTA – Middle-Tier Acquisition

MTP – Master Test Plan

MVCR – Minimum Viable Capability Release

MVP – Minimum Viable Product

NDI – Non-Developmental Items

OA – Operational Assessment

OFP – Operational Flight Program

ORD – Operational Requirements Document

ORM – Operational Risk Management

OT – Operational Test

OTA – Operational Test Agency

OT&E – Operational Test and Evaluation

OTO – Operational Test Organization

OTP – Operational Test Plan

OUE – Operational Utility Evaluation

PAT&E – Production Acceptance Test and Evaluation

PEO – Program Executive Officer

PM – Program Manager

PMD – Program Management Documents

PMO – Program Management Office

PPQT – Pre-Production Qualification Test

PQT – Production Qualification Test

PTO – Participating Test Organization

QOT&E – Qualification Operational Test and Evaluation

QT&E – Qualification Test and Evaluation

RAM – Reliability, Accessibility, and Maintainability

RCT – Requirements Correlation Tables

RDT&E – Research, Development, Test and Evaluation

RMF – Risk Management Framework

SIL – Software Integration Lab

SOTR – Sufficiency of Test Report

SPO – System Program Office

STAT – Scientific Test and Analysis Techniques

SUT – System Under Test

TD&E – Tactics Development and Evaluation

T&E – Test and Evaluation

TEMP – Test and Evaluation Master Plan

TIPT – Test Integrated Product Teams

TM – Test Manager

UC – Unified Capabilities

VV&A – Verification, Validation and Accreditation

WIPT – Working-level Integrated Product Teams

XFA – Cross Functional Authority