# The Value
# of
# Operational Testing and Evaluation

# Purpose of Operational Testing

- **The purpose of operational testing is to provide realistic and objective assessments of how systems improve mission accomplishment under realistic combat conditions**
- **We do this by evaluating systems in operational scenarios when employed by units against realistic threat forces**
  - We provide acquisition executives with objective facts concerning both system performance and system contribution to mission accomplishment
  - We provide warfighters with objective information on the capabilities and limitations of systems they employ in combat
- **We provide this objective information before a system is used in combat and prior to full-rate production**
- **We evaluate the adequacy and results of the operational tests based upon sound scientific methods**
  - Several studies by the Defense Science Board and National Research Council have advocated such analytical methods
- **Our fundamental purpose has been codified in Title X and DoD 5000 for many years and has not changed**
  - Detailed language of the law and regulations contained in backups

# Themes

1. **Testing merely to determine if KPPs or requirements have been satisfied often does not capture improvements in mission accomplishment, or achievement of intended capabilities in an operational environment**

2. **Operational testing has to be performed against expected realistic operational threats to the units employing the systems**

3. **There have been many examples of operational testing identifying critical system performance problems, when employed in an operationally realistic environment, that provided opportunities for correction before systems are fielded or deployed**

4. **Operational testing can identify issues with systems' combat effectiveness. This information is provided to acquisition authorities for decision. Decisions by those authorities to fix problems can affect program schedules if they judge the associated fix worth the delay**

5. **DOT&E has been responsive and supportive of fielding systems for the current wars**

6. **Rigorous scientific and statistical methods are used to determine adequacy of operational testing in terms of miles driven, shots fired, and coverage of the operational envelope. Sound statistical analysis of operational test results provides accurate estimates of information to acquisition decision-makers**

# Theme 1

Testing only to determine if KPPs or requirements have been satisfied often does not capture improvements in mission accomplishment, or achievement of intended capabilities in an operational environment

# Theme 1: KPPs Alone Aren't Sufficient

## *P-8A Poseidon Multi-Mission Maritime Aircraft Example*

**P-8A Poseidon is a maritime patrol aircraft that will replace the P-3C Orion. P-8A is based on the Boeing 737-800 airframe and will conduct Anti-Submarine Warfare (ASW) and other missions**

- **P-8A requirements do not adequately measure the effectiveness of the P-8A's two primary missions: finding and killing enemy submarines and reconnaissance**
  - JROC-approved KPPs require only that P-8A fly a specified range while carrying a required number of sonobouys and that it communicate with certain radios
  - KPPs could be achieved without finding and killing enemy submarines or conducting reconnaissance
- **DOT&E argued for an OT that examined whether the Navy's Concept of Employment for P-8 could be executed under realistic combat conditions**
  - Testing went beyond simple verification of KPPs
  - Navy's Operational Test Agency agreed with this approach
  - Navy performed realistic testing during Fleet exercises using a full set of mission systems and crew to examine their ability to find and attack submarines and perform reconnaissance using the P-8A
  - Testing revealed important deficiencies the Navy is now working to fix through improved long-range sensors
- **DOT&E sent a memorandum to the Vice Chairman of the Joint Chiefs of Staff, who chairs the JROC, noting the lack of coverage of realistic combat conditions in the P-8A's, and other systems', approved KPPs**
  - Admiral Winnefeld moved to improve the P-8A's and other systems' KPPs, and specifically requested the continued involvement of DOT&E in assisting the JROC to do that

Acronyms this slide: Joint Requirements Oversight Council (JROC); Key Performance Parameter (KPP)

# Theme 1: KPPs Alone Aren't Sufficient
## *Virginia-class Submarine IOT&E Example*

**Virginia is a multi-mission nuclear attack submarine that will replace the existing Los Angeles class. It is intended to provide covertness equivalent to the Seawolf-class submarine at lower cost**

- **Evaluating the Virginia class solely by its threshold requirements would result in a misleading assessment of mission effectiveness**
- **DOT&E's evaluation of the Virginia class was based on its mission performance in relation to the legacy Los Angeles-class submarine that it replaces**
  - Virginia did not meet some classified KPP thresholds
  - Virginia did not meet required level of performance in some mission areas
  - However, Virginia's performance was equivalent to or better than the legacy Los Angeles class in all mission areas
- **DOT&E evaluated Virginia as effective and suitable because it performed as well or better than Los Angeles-class submarine, with improved covertness**

Acronyms this slide: Initial Operational Test and Evaluation (IOT&E); Key Performance Parameters (KPPs)

# Theme 1: KPPs Alone Aren't Sufficient
## *Virginia-class Submarine Modernization FOT&E Example*

**Virginia is a multi-mission nuclear attack submarine that will replace the existing Los Angeles class. It is intended to provide covertness equivalent to the Seawolf-class submarine at lower cost**

- **FOT&E in 2011 assessed the effect of modernizing the sonar and combat control systems on Virginia's mission performance and suitability. Evaluating strictly against requirements would have resulted in a misleading assessment**

- **Torpedo localization requirement was not adequate to assess performance**
  - The torpedo localization requirement (levied on AN/BYG-1 combat control system) was based on a Navy study that related solution accuracy and torpedo hit performance
  - The requirement assumed that a target solution (bearing, range, course, and speed estimate) within specified tolerances would result in a torpedo hit
  - Testing showed that torpedo hit performance was poorly correlated to the solution accuracy requirement
  - Many shots that failed the requirement actually resulted in hits

- **Sonar reliability requirement was lower than for the legacy system**
  - Although the sonar system met the new, lower reliability threshold, it was significantly less reliable than the system it replaced—this was noted in the evaluation

- **Although BYG-1 failed to meet its requirement, DOT&E, using a metric that was based on the Navy's actual torpedo employment tactics, assessed the system to be able to adequately place the torpedo such that it has an opportunity to hit its target**

# Theme 1: KPPs Alone Aren't Sufficient
## *MQ-1C Gray Eagle Availability Example*

**The MQ-1C Gray Eagle-equipped unit is an unmanned aircraft system that provides tactical intelligence, video, radar imagery, communications relay, and precision missile support to Army maneuver units**

- **The Gray Eagle Reliability KSA established unnecessarily high thresholds to meet the Availability KPP**

- **The JROC-approved KSA thresholds were established using an Army availability model**
  - The Army's model established the KSA thresholds, believing those thresholds necessary to achieve the Availability KPP of 80 percent
  - System performance during developmental and Limited User Testing demonstrated the system could achieve about 30 percent of its required Mean Time Between Abort KSA requirement

- **During the 2012 IOT&E a fully-equipped unit operating at a wartime operating tempo met the Availability KPP even though the major subsystems continued to fall well short of their reliability thresholds**
  - The Army's modeling assumption that most failures would cause a mission abort was demonstrated to be incorrect in operational testing

- **Recognizing that the purpose for the reliability thresholds was to achieve the Availability KPP, DOT&E de-emphasized the Reliability KSA threshold values**
  - DOT&E reported that the Gray Eagle system was Suitable
  - The Army revised their availability model and established reliability KSA thresholds consistent with the demonstrated subsystem reliability

Acronyms this slide: Key Performance Parameter (KPP), Key System Attribute (KSA)

# Theme 1: KPPs Alone Aren't Sufficient
## *CVN 78 Gerald R. Ford Example*

**CVN 78 class replaces Nimitz class aircraft carriers. CVN 78 has the Nimitz class hull form, but many ship systems are new, including the nuclear plant, catapults, arresting gear, and radar**

- **Sortie Generation Rate (SGR) is a Key Performance Parameter (KPP) for CVN 78, but the threshold is unrealistic**
  - SGR measures the number of aircraft sorties per day launched from the carrier
  - CVN 78 is intended to improve performance relative to the Nimitz class.
  - CVN 78 requirement is 160 sorties per day (12-hour fly day) in sustained operations and 270 sorties per day (24-hour fly day) in a surge
  - Nimitz class has demonstrated 120 sorties per day sustained and 240 sorties per day surge

- **CVN 78's SGR requirement includes unrealistic assumptions**
  - Assumes 35-days of fair weather and unlimited visibility
  - Assumes aircraft emergencies, failures of shipboard equipment, ship maneuvers (e.g., to avoid land), and manning shortfalls will not affect flight operations.
  - These factors affect operations on all carriers
  - CVN 78 is unlikely to achieve the unrealistic threshold in combat conditions

- **DOT&E will, of course, report against the JROC-approved threshold, as well as comparing the SGR of the Ford-class carriers against the demonstrated performance of the Nimitz-class carriers**

# Theme 1: KPPs Alone Aren't Sufficient

## *C-5 Reliability Enhancement and Re-Engining Program (RERP) Example*

**The RERP modification to legacy C-5 cargo aircraft is intended to create a C-5M with improved reliability, availability, and fleet payload capability as well as continued access to international airspace**

- **C-5M did not meet Mission Capable Rate (MCR) KPP during operational testing**
  - MCR of the C-5M alone does not capture the ability of the airlift fleet ( a combination of C-17, C-5A and C-5M aircraft) to move cargo, which is its primary mission
  - AFOTEC conducted a 30-day surge at a moderate operational tempo, including a 6-day "super-surge" at high operational tempo, to simulate wartime conditions
  - Three test aircraft achieved an MCR of 68% during the surge and 66% during the super-surge
  - Legacy C-5 fleet had an MCR of 56%
  - Consequently, C-5M did not achieve 75% requirement, but had a higher MCR than the legacy fleet

- **The availability KPP did not adequately capture the benefit of the RERP modification to C-5M capability to support mobility mission requirements**
  - DOT&E looked at fleet capacity in terms of Million Ton Miles/Day (MTM/D)
  - The operational testing results demonstrated a C-5M fleet capacity of 5.1 MTM/D, including a new capability to fly longer routes non-stop when compared to the legacy C-5
  - These results supported the Mobility Capabilities and Requirements Study (MCRS-16) strategic mobility fleet capacity for the program of record (223 C-17s, 52 C-5Ms, 37 C-5As)
  - DOT&E, therefore, did not consider the Availability KPP shortfall operationally significant

Acronyms this slide: Key Performance Parameters (KPP)

# Theme 1: KPPs Alone Aren't Sufficient
## *Early Infantry Brigade Combat Team (EIBCT) Example*

**EIBCT systems were a collection of sensors to be used in infantry brigades to detect and provide warning of enemy activities**

- **The KPPs for some of the sensors specified only that the systems produce images recognizable as human faces at specified distances**
  - DOT&E advocated and the Army agreed that the systems be tested under realistic combat conditions against a capable enemy threat
- **OT occurred in the summer of 2010 at WSMR**
  - Enemy soldiers could easily spot the large antennas needed to transmit the images back to the operations centers
  - Most of the images transmitted were false alarms as enemy soldiers spoofed the sensors
  - Unit soldiers reported during testing that they had multiple other means by which to accomplish the same objective, which they preferred
- **The Army itself determined after the OT that the sensors were not useful and unnecessary, even though they met the technical specifications required by the KPPs**
- **The Army canceled the portions of the program that were found to be unnecessary**

Acronyms this slide: Key Performance Parameters (KPP); operational testing (OT); White Sands Missile Range (WSMR)

# Theme 1: KPPs Alone Aren't Sufficient

## *Remote Minehunting System (RMS) Example*

**RMS is a system of systems that includes the Remote Multi-Mission Vehicle (RMMV) and AN/AQS-20A sonar. The Navy plans to use RMS in the LCS MCM mission package to detect and identify mines.**

- **RMS requirements do not capture the need to locate mines quickly and accurately as part of a system-of-systems onboard LCS**
  - RMS KPPs and KSAs can be achieved while system fails to support the LCS MCM mission
  - RMS search rate requirement has no limit on false alarms and excludes time required for planning, transit to and from the search area, operator assessment, follow-on actions to reduce false classifications, and efforts to recover from failures
  - In contrast, the LCS requirement includes these factors
  - RMS reliability KSA excludes the sensor, which artificially inflates the RMMV operating time and limits scoring of operational mission failures, thus yielding an unrealistic assessment of system reliability
    - » 192.5 hours MTBOMF (per CDD criteria) vice 31.3 hours (per operational criteria) during recent DT
- **DOT&E has recommended that RMS be tested as the Navy expects to employ it in the LCS MCM mission package**
  - More than a simple verification of RMS-specific KPPs
  - Navy assessed system in DT using KPPs
  - Neither the Navy nor DOT&E can assess RMS contribution to LCS performance based on this testing
- **This RMS example illustrates a recurrent problem with inconsistencies between requirements documents for different systems**

Acronyms this slide: Remote Minehunting System (RMS); Littoral Combat Ship (LCS); Mine Countermeasures (MCM); Key Performance Parameter (KPP); Key System Attribute (KSA); Capability Description Document (CDD); Mean Time Between Operational Mission Failures (MTBOMF)

# Theme 1: KPPs Alone Aren't Sufficient

## *Warfighter Information Networking – Tactical (WIN-T) Inc 2 Example*

**WIN-T Increment 2 is a mission command system that provides command and control to tactical units on-the-move via both satellite and terrestrial based transport**

- **WIN-T KPPs and KSAs don't reflect the most important requirements of a large unit command and control system: rapid, reliable and clear access to his subordinates**
  - WIN-T has 5 KPPs and 4 KSAs, none of which describe the command and control requirements of commanders. The two that come closest state:
    - » KPP 3 - Critical survival information delivery in ≤5 seconds (95% of completed messages)
    - » KSA 1 – Increment 2 shall implement a high-capacity IP routed multi-tiered network utilizing ground-to-ground and ground-to-space transmission links

- **In addition to the KPPs and KSAs, DOT&E considered connectivity (the number of subordinates that commanders could actually reach), network stability, voice call clarity and range of the systems as the key criteria indicating operational utility to unit commanders**
  - DOT&E identified that the WIN-T network was capable of forming and connecting commanders with subordinates in most cases
  - However, network instabilities, poor voice call clarity and limited range of the terrestrial waveform led to several WIN-T components being rated as not operationally effective

- **The Program Office agreed the problems identified in the DOT&E evaluation were valid, and implemented fixes to these problems that have significantly improved the system. They have scheduled a FOT&E in November 2014.**
  - DOT&E analysis was used by the Army to decide to pursue improvement in areas of network stability and voice call clarity in particular

Acronyms this slide: Warfighter Information Networking-Tactical (WIN-T), Initial Operational Test and Evaluation (IOT&E), Follow-On Test and Evaluation (FOT&E); Internet Protocol (IP); Key Performance Parameters (KPPs); Key System Attributes (KSAs)

# Theme 1: KPPs Alone Aren't Sufficient

## *Stryker Nuclear Biological Chemical Reconnaissance Vehicle Example*

**The Stryker Nuclear Biological and Chemical Reconnaissance Vehicle (NBCRV) is a Stryker vehicle augmented with a mounted suite of equipment and software designed to detect, identify, collect, analyze, and mark NBC threats, and to disseminate that information to operational commanders and forces**

- **The reliability requirement for the Stryker NBCRV was stated solely in terms of the base Stryker vehicle**
  - The KPP specifically excluded the sensors and mission equipment that enable the NBCRV to do its NBC reconnaissance mission
- **DOT&E analyzed reliability and its impact on mission accomplishment from an overall system point of view**
  - DOT&E found the Stryker NBCRV to be not suitable during the initial operational test, due to poor reliability of the overall system, including the sensors, which caused it to be unable to accomplish its NBC reconnaissance mission
  - The USD(AT&L) subsequently mandated the program undergo a reliability growth initiative prior to fielding
- **As a result of the USD(AT&L) decision, the vendor implemented corrective actions that significantly improved the reliability of both the base vehicle and the NBC sensor package**
- **The program subsequently met its reliability growth metrics and DOT&E evaluated it to be suitable following additional operational testing**
  - The Stryker NBCRV is currently being fielded

Acronyms this slide: Nuclear Biological Chemical Reconnaissance Vehicle (NBCRV), Nuclear Biological Chemical (NBC), Under Secretary of Defense (Acquisition, Technology and Logistics (USD(AT&L))

# Theme 1: KPPs Alone Aren't Sufficient
## *AN/BLQ-10 Submarine Electronic Support System Example*

**The AN/BLQ-10 system allows submarines to intercept radar and communications signals**

- **The Navy's requirements for BLQ-10 did not allow for meaningful performance assessments of the most recent operationally tested version of BLQ-10**

- **The requirements did not capture intended performance improvements**
  - The KPP for radar intercept is a roll-up measure that averages performance against the entire threat list
  - According to this metric, the previous version of BLQ-10 already passed the threshold, but it had no capability against Low Probability of Intercept radars.
  - The newest version of BLQ-10, denoted Technical Insertion 2008 (TI-08), added the capability to intercept some LPI radars
  - Adding LPI capability is operationally important, but has a minimal effect on the KPP that rolls-up performance

- **The requirements did not capture operational performance**
  - The primary operational use of BLQ-10 is threat warning, so that submarines can avoid being detected by threat radars
  - The BLQ-10 requirements do not address threat warning at all
  - DOT&E advocated a realistic test of threat warning against an advanced P-3C radar
  - The test provided valuable insight about the technical capability of the system and the operator's ability to make use of the information

- **The Fleet has recently decided to make improvement of BLQ-10 one of its top priorities**

# Theme 1: KPPs Alone Aren't Sufficient
## *Vertical Launch Antisubmarine Rocket (VLA) with the Mk 54 Mod 0 Lightweight Torpedo Example*

**The legacy VLA system for surface ships, which provided a rocket-boosted delivery of the Mk 46 torpedo, was updated to carry the new Mk 54 Mod 0 torpedo as its warhead**

- **Navy requirements focused on VLA missile accuracy, not end-to-end mission effectiveness, obscuring an assessment of system performance**
  - Approved KPPs require only that the VLA accurately deliver the torpedo to a designated water entry point, regardless of whether that point is near the target submarine
  - KPPs included minimum and maximum missile range and Circular Error Probable (CEP) for splash point relative to aim point
  - KPPs could be achieved without killing enemy submarines
- **DOT&E argued for end-to-end operational testing of the VLA**
  - Safety rules precluded firing against a manned target
  - Surrogate targets did not present appropriate acoustic signatures and threat evasions
  - DOT&E advocated using data from previous shipboard sonar testing and Mk 54 Mod 0 OT
- **DOT&E (January 2010) evaluated the system based on the system-of-systems performance**
  - Shipboard ASW sensors are incapable of detecting and localizing the target submarine with accuracy sufficient to place the torpedo in the water within reasonable attack criteria
  - Mk 54 Mod 0 OT, while incomplete, demonstrated concerns about torpedo effectiveness
  - VLA does not provide an effective end-to-end standoff ASW capability

Acronyms this slide: Key Performance Parameter (KPP); Antisubmarine Warfare (ASW)

# Theme 1: KPPs Alone Aren't Sufficient

## *MQ-1C Gray Eagle Survivability Example*

**The MQ-1C Gray Eagle-equipped unit is an unmanned aircraft system that provides tactical intelligence, video, radar imagery, communications relay, and precision missile support to Army maneuver units**

- **The Gray Eagle's Survivability and Force Protection KPP did not directly assess system survivability in an operational environment, but only stated it will be on a truck**
  - "The Ground Control System will be mounted onto an Army standard tactical vehicle with the ability to be up-armored."
- **The Army demonstrated this KPP by mounting a Ground Control Station on an FMTV during cross-country mobility testing**
- **DOT&E used all available test data to assess Gray Eagle Survivability**
  - Using existing Live Fire Test and Evaluation data, DOT&E reported on the level of ballistic protection afforded by the FMTV vehicles
  - DOT&E reported on the acoustic and visual detectability of the Gray Eagle aircraft (a CPD requirement, but not included in the Survivability and Force Protection KPP)
  - Using data collected during developmental testing, DOT&E reported on the survivability of MQ-1C air vehicles against likely infrared and radar-guided missile threats
- **Using test data from all sources, DOT&E provided decision makers with a broad survivability assessment of the Gray Eagle MQ-1C air vehicles and ground transport vehicles, well beyond the narrowly-defined KPP**

Acronyms this slide: Key Performance Parameter (KPP), Family of Medium Tactical Vehicles (FMTV), Capabilities Production Document (CPD)

# Theme 2

Operational testing has to be performed against expected realistic operational threats to the units employing the systems

# Theme 2: Realistic Threats Necessary

## *F-35 Joint Strike Fighter Example*

**The F-35 is a tri-Service, multi-national, family of strike aircraft. It will replace the F-16 and A-10 in the Air Force, the AV-8B in the Marine Corps, and augment the F-18 in the Navy**

- **To understand F-35 performance, it should be tested against existing integrated air defense systems**

- **Current open air test ranges cannot replicate the complexity or density of the signals environment associated with existing integrated air defense systems described in the JSF System Threat Assessment Report**
  - Modern threat systems create complex, dense electromagnetic signal environments
  - DOT&E recommended that the F-35 be evaluated in open air testing which is representative of this environment
  - Collaborating with USD AT&L, DOT&E advocated for a Department of Defense electronic warfare infrastructure improvement program to secure open- and closed-loop emitters for open air testing of F-35 and other programs

- **The near term goal is to have open air emulators available for the F-35 IOT&E so that F-35 capabilities can be evaluated against a realistic threat**
  - Emulators will represent the complexity and density of the threats
  - Open air trials will provide system performance data to support man-in-the-loop simulation development and validation

- **DOT&E worked with other OSD components and the Services for 10 months to secure approval by the Secretary of Defense for additional funding for the improvement project through the FY14 program and budget review**
  - Funding was allocated from sources outside the F-35 program

# Theme 2: Realistic Threats Necessary

## *Anti-Air Warfare Ship Self Defense (AAWSSD) Enterprise Example*

**The AAWSSD Enterprise coordinates testing across several ship classes (e.g., LHA 6) and combat system elements (e.g., Rolling Airframe Missile) to efficiently test the Navy's Probability of Raid Annihilation (PRA) requirements against anti-ship cruise missiles (ASCM)**

- **To assess ship self-defense systems against ASCMs, DOT&E advocates testing under operationally realistic conditions. This testing has identified important deficiencies**

- **To conduct realistic testing, the Navy decided to resource:**
  - The unmanned Self Defense Test Ship to examine performance against ASCM targets at short ranges that cannot be safely tested on manned ships
  - New aerial targets – Multi Stage Supersonic Target and the Subsonic Aerial Targets – are needed to represent modern ASCM threats
  - A high fidelity system-of-systems combat system model that is used to predict performance results against actual threats

- **DOT&E encouraged the Navy to develop a common Enterprise strategy for PRA assessments**
  - Allowed the Navy to reduce planned testing by 38 missiles, 28 targets, and 11 scenarios
  - Saving $240 Million out of $1.1 Billion budgeted

- **AAWSSD testing has identified major deficiencies that the Navy is working to fix:**
  - Evolved SeaSparrow Missile (ESSM) deficiencies for specific types of ASCM threats and raids
  - Rolling Airframe Missile (RAM) deficiencies against specific types of threats and raids
  - Ship Self Defense System (SSDS) combat system deficiencies with respect to sensor integration and engagement scheduling
  - Cooperative Engagement Capability (CEC) tracking problems against specific threats
  - Radar system (e.g., SPS-48E and SPQ-9B ) detection gaps for specific threats and raid types

# Theme 2: Realistic Threats Necessary
## *Paladin Integrated Management (PIM) Example*

**The PIM program upgrade the current M109A6 self-propelled howitzer and its ammunition support vehicle to improve sustainability, supportability, mobility and protection**

- **The Army's rationale for the need for PIM in its original CPD was inconsistent with the system's technical specifications which were an annex to the CPD**
  - The CPD's rationale for PIM indicated it must be capable of operating off of a FOB and able to deal with IED and other threats present in Iraq and Afghanistan
  - The PIM's technical specifications were inconsistent with that rationale and the associated threats
- **The Army agreed with the inconsistency and decided to re-write the CPD. The revised rationale indicated that PIM would always move on cleared routes; nonetheless, the Army decided to increase PIM's side and underbelly protection requirements somewhat**
- **The USD(AT&L) directed the Army to evaluate the effort and cost required to test and evaluate an objective-level underbelly kit for PIM, and subsequently directed the Army to fabricate and test five underbelly kits**
- **Consistent with the USD (AT&L) direction, the Army added plans to test PIM vehicles equipped with objective-level protection in DT and to characterize the operational impact of additional weight in operational testing**

Acronyms this slide: Paladin Integrated Management (PIM); Self-Propelled Howitzer (SPH); Capabilities Production Document (CPD)' Forward Operating Base (FOB) Improvised Explosive Device (IED)

# Theme 2: Realistic Threats Necessary
## *Stryker Double-V Hull (DVH) Family of Vehicles Example*

**The Stryker DVH provided integrated enhancements to the base Stryker vehicle to improve both crew survivability and mobility. It was a limited production effort provided to units employed in Operation Enduring Freedom**

- **The Army initially proposed limited ballistic testing of only non-functional Double-V hull chassis against threshold-level threats**
- **DOT&E advocated that the LFT&E program support a broader, more realistic evaluation**
  - Suggested addressing seven different Stryker family DVH configurations with unique mission equipment (infantry carrier, command, combat engineer, scout, mortar carrier, medical, missile launcher)
  - Each of these variants were tested under normal operating conditions and configured for combat
  - Also proposed testing "flat-bottom" Strykers in the same configurations to provide a baseline for comparison to DVH
  - Included threshold-level, objective-level, and beyond-objective threats representative of then current OEF combat experience
- **Stryker DVH LFT&E demonstrated that the Stryker DVH provides protection above requirements that was a substantial and significant improvement relative to the existing flat-bottomed Stryker**
  - Characterized protection capabilities against theater-specific threat types at multiple levels
  - Identified protection design changes for fleet-common and variant-specific vulnerabilities for DVH that the Army decided to implement
  - Identified opportunities to further increase survivability through training and employment options
- **The Army developed design fixes and provided employment recommendations to improve survivability for units deploying to Afghanistan; field commanders used the test results to plan operations tailored to the underbody threats prevalent in different regions**

Acronyms this slide: Double-V Hull (DHV); Operation Enduring Freedom (OEF); Live Fire Test and Evaluation (LFT&E)

# Theme 2: Realistic Threats Necessary

## *AH-64E Apache Attack Helicopter Example*

**The AH-64E upgrades the four-bladed, twin-engine legacy AH-64D attack helicopter with a new transmission (more power at high/hot conditions) and interoperability with unmanned aircraft system**

- **The Army initially did not include Information Assurance testing during the 2012 IOT&E, claiming that the AH-64E would be invulnerable to network attack**
- **DOT&E advocated that Information Assurance testing be conducted during the IOT&E**
  - During the last week of the Apache IOT&E, an Army computer network operations Red Team conducted limited penetration testing of AH64E computer networks
  - The Red Team considered three attack methods for gaining access to the aircraft's networked system: the Blue Force Tracker, the Aviation Mission Planning System, and aircraft maintenance ports.
  - To avoid compromising flight safety, IA penetration testing took place while the aircraft were powered up but without engines or rotors turning.
- **The computer network operations Red Team discovered multiple avenues by which AH-64E computer information could be compromised, corrupted, or exploited**
- **The program office agreed the vulnerabilities were significant, has since developed corrections to close off those vulnerabilities**
  - Another Red Team will attempt to penetrate the next lot of AH-64E during follow-on testing to confirm that the previous threat avenues have indeed been shut down

*Acronyms this slide: IOT&E: Initial Operational Test and Evaluation; IA: Information Assurance*

# Theme 2: Realistic Threats Necessary

## *Aegis Ballistic Missile Defense (Aegis BMD) Example*

**Aegis BMD provides U.S. Navy destroyers and cruisers with the capability to defeat short-, medium-, and intermediate-range ballistic missiles and provides surveillance and tracking of intercontinental ballistic missiles in support of the Ground-based Midcourse Defense system**

- **The Missile Defense Agency (MDA) proposed to deploy Aegis BMD for defense of southern Europe without flight testing against some expected threats**

- **In 2010, the President proposed a new phased adaptive approach (PAA) for ballistic missile defense in Europe, with Aegis BMD as the primary weapon element**
  - The new defense mission increased the Aegis BMD threat set to include 2,000+ kilometer threats
  - The MDA test program did not include Aegis BMD intercepts against such medium- to intermediate-range threats
  - The test program also did not include flight testing of the Aegis BMD launch-on-remote capability using a forward-based AN/TPY-2 radar. This capability is essential to Aegis BMD defense of Europe

- **DOT&E and MDA agreed it would be useful to conduct a flight test against a 3,000+ kilometer target using AN/TPY-2 radar data**
  - A successful test was conducted in April 2011

- **U.S. European Command endorsed the test and deployed the PAA architecture to Europe with the knowledge that Aegis BMD could intercept a target with a range representative of threats to southern Europe**

Acronyms this slide: Aegis Ballistic Missile Defense (Aegis BMD); Missile Defense Agency (MDA); PAA (Phased Adaptive Approach)

# Theme 2: Realistic Threats Necessary

## *Enhanced Combat Helmet (ECH) Example*

**Forces equipped with the ECH will rely on the helmet to provide ballistic protection from selected bullet and fragment threats. The ECH is more resistant to penetration than previous combat helmets**

- **The ECH can meet its ballistic protection KPP requirements without actually providing improved protection compared to existing combat helmets**
  - The program requirements document only specifies a penetration requirement against a single, direct-fire small arms threat
- **DOT&E assessed the ECH's capability to prevent injury and death from operationally realistic direct- and indirect-fire threats**
  - When hit by realistic operational threat bullets or fragments, the helmet shell rapidly deforms and impacts the soldier's head – but there is no requirement limiting this deformation
  - DOT&E coordinated with the Armed Forces Medical Examiner's Office (AFMEO) to get an expert medical opinion on the deforming shell's potential to harm the wearer following an impact with these potential threats
  - By including AFMEO's evaluation, DOT&E's assessment provides a more complete picture of the capability of the helmet against the full range of operational threats
- **The DOT&E Live Fire Test and Evaluation reports (original and follow-on) include an operationally-relevant comparison between the ECH and fielded combat helmets**
- **Army leadership considered the DOT&E assessment when they reduced their intended purchase of ECH helmets from 200,000 to 35,090**

Acronyms this slide: Enhanced Combat Helmet (ECH); Armed Forces Medical Examiner's Office (AFMEO)

# Theme 3

There have been many examples of operational testing identifying critical system performance problems, when employed in an operationally realistic environment, enabling corrections before systems were fielded or deployed

# Theme 3: Identifying Problems for Correction
## *Stryker Double-V Hull (DVH) Family of Vehicles Example*

**The Stryker DVH provided integrated enhancements to the base Stryker vehicle to improve both crew survivability and mobility. It was a limited production effort provided to units employed in Operation Enduring Freedom**

- **In response to an Army Operational Need Statement in December 2009, the Department issued an Acquisition Decision Memorandum in April 2010 authorizing the Army to procure Stryker DVH vehicles**
  - DOT&E worked with the Army to develop a phased test plan that supported the production and fielding schedules
- **DOT&E advocated that the test examine emergency egress from the vehicle**
  - Reports from deployed units indicated a high frequency of injuries to the driver
  - The Army agreed with this approach
- **The operational test on the Infantry Carrier Variant occurred at Yuma Proving Ground in January 2011 and the National Training Center in February 2011**
  - Egress events demonstrated that driver emergency evacuation through both the driver's hatch and the crew compartment was impeded by changes to the driver's station
- **Based on these results, the program office and contractor developed a quick fix to the driver's seat and windshield, while concurrently re-designing the driver's seat area**
  - The fix corrected the most serious deficiency and was applied to vehicles in theater before they were issued to units beginning in June 2011
  - The re-designed driver's seat corrected numerous other issues and was retro-fitted in theater in June 2012

# Theme 3: Identifying Problems Early
## *Mk 54 Mod 0 Lightweight Torpedo Example*

**Tactical software Block Upgrade to the Mk 54 Mod 0 Lightweight Torpedo was intended to improve performance overall and address emerging threats**

- **DOT&E recommended that the Navy use a more realistic target and threat tactics during testing that ultimately uncovered a significant problem with the system**

- **Urgent Operational Needs Statement identified a classified threat of interest**
  - Navy planned to conduct Quick Reaction Assessment (QRA) to support early fielding of the Block Upgrade
  - DOT&E recommended the use of a more realistic target and threat tactics in the QRA and in operational testing that would follow

- **Testing revealed that the MH-60R helicopter could not appropriately initialize the Mk 54 Mod 0 during the more realistic scenario—the helicopter could not fire the torpedo in combat**
  - DT used less stressful scenarios and did not identify this deficiency
  - Two years after the early fielding following the QRA, the Navy had not updated tactical guidance and documentation

- **Navy investigation found numerous problems**
  - Poor communication between the weapons and platform developers and the training community
  - Critical tactical guidance was not reaching the Fleet users

- **Navy instituted new processes to address the issue**
  - Issued interim guidance to the Fleet

# Theme 3: Identifying Problems Early
## *Anti Air Warfare Ship Self Defense (AAWSSD) Enterprise Example*

**The AAWSSD Enterprise coordinates testing across several ship classes (e.g., LHA 6) and combat system elements (e.g., Rolling Airframe Missile) to efficiently test the Navy's Probability of Raid Annihilation (PRA) requirements against anti-ship cruise missiles (ASCM)**

- **Operational Testing under the AAWSSD Enterprise revealed numerous combat system deficiencies**
  - Many of these deficiencies were only found because testing was conducted under operationally realistic conditions (See theme 2 entry)
- **The AAWSSD Enterprise system-of-systems testing allows multiple systems that must work in an integrated manner to be tested at once**
- **Major Deficiencies have been found in many AAWSSD Enterprise programs**
  - Evolved SeaSparrow Missile (ESSM) deficiencies related to specific types of ASCM threats and raid types
  - Rolling Airframe Missile (RAM) deficiencies against specific raid types and threats
  - Ship Self Defense System (SSDS) combat system deficiencies with respect to sensor integration and engagement scheduling
  - Cooperative Engagement Capability (CEC) tracking problems against specific threats
  - Radar system (e.g., SPS-48E, SPQ-9B, etc. ) detection gaps associated with certain threats and raid types
- **A number of these deficiencies have been corrected prior to ship deployments and/or have lead to changes in Navy tactics**

# Theme 3: Identifying Problems Early
## *Gerald R. Ford (CVN 78) Class Aircraft Carrier*

**CVN 78 is the Navy's newest class of nuclear powered aircraft carrier and will replace the USS *Enterprise* and USS *Nimitz* Class carriers. The Full Ship Shock Trial (FSST) is a critical part of the alternative LFT&E plan for the class.**

- **With approval of the Milestone B TEMP and LFT&E Management Plan in March 2004, the program of record called for the FSST on CVN 78, the first ship of the class.**

- **The Navy proposed delaying the FSST to CVN 79 in an 8 July 2011 memo to DOT&E**
  - Delaying the FSST to CVN 79 would cause a delay of 5 to 7 years in obtaining data critical to evaluating the survivability of the CVN 78 class and precludes timely modification of subsequent ships to assure their survivability
  - CVN 78 will be operational for several years and will deploy before the FSST occurs

- **What occurred:**
  - DOT&E memo to PEO Carriers on 3 August 2011 indicated concern with the delay but willing to consider Navy arguments
  - On 10 November 2011, PEO Carriers briefed DOT&E on low risk and high cost of delay
  - In a 28 November 2011 memo to ASN(RDA), DOT&E indicated the delay was unacceptable and Navy should fund and schedule the FSST on CVN 78
  - ASN(RDA) replied on 18 June 2012, stating that CVN 79 was the best option
  - DOT&E rescinded approval of the CVN 78 Class LFT&E Management Plan on 12 July 2012, because the Navy made unilateral changes to the approved alternative LFT&E plan

- **To obtain DOT&E approval of latest TEMP and LFT&E Management Plan revisions, the Navy should schedule and fund the conduct of a shock trial on CVN 78**

# Theme 3: Identifying Problems Early

## *T-AKE 1 USNS Lewis and Clark Example*

**T-AKE 1 is the Navy's newest class of dry cargo and ammunition ships**

- **The Navy initially did not plan to conduct an LFT&E program for a commercial-designed support ship operated by civilian mariners (Military Sealift Command)**
- **The Navy argued there was no requirement to assess the survivability of T-AKE since it would never operate in harms way**
  - Using the Navy's concept of operations, DOT&E showed that the T-AKE's mission would expose it to the same threats as the ships it supplies
  - The Navy and DOT&E agreed to an LFT&E program that included surrogate testing, modeling and simulation, selected analyses, and a total ship survivability trial
- **Surrogate testing and analyses demonstrated that:**
  - The legacy Heat Sensing Device found in magazines throughout the fleet failed to function properly
  - The design of Intermediate Lateral Supports in the magazines interfered with the new magazine sprinkler system
  - The Self-Contained Breathing Apparatus bottle charging station was inadequate
  - Magazines were vulnerable to heat from weapons-induced fires
- **As a result of the LFT&E program, the Navy decided to correct the critical system problems for T-AKE and to establish new fleet-wide requirements for magazine fire detection systems**

# Theme 3: Identifying Problems Early
## *DoD Automated Biometrics Identification System (ABIS)*

**DoD ABIS is the result of a Joint Urgent Operational Need request and provides receive, process, store, match, and share capabilities for positively identifying known and potential adversaries**

- **Since 2010, there have been four failed attempts to deploy the ABIS 1.2 upgrade, all resulting in roll-back decisions to the ABIS 1.0 operational baseline**
  - The program established a pattern of fixing problems found in developmental testing in a non-operational environment, and then failing in operational deployment
  - In the last deployment attempt, CENTCOM and SOCOM reported 42 high priority deficiencies impacting critical warfighter missions
- **DOT&E is now requiring an initial operational test and evaluation of ABIS 1.2 prior to any further deployment attempts. DOT&E is advocating for**
  - An ABIS 1.0 baseline assessment prior to the ABIS 1.2 IOT&E. No independent baseline assessment has been performed on ABIS 1.0
  - A rigorous DT&E in an operationally representative environment to include demonstration of end-to-end information exchanges under various load and threat conditions
  - Entrance criteria to the IOT&E including demonstration of adequacy of tools, processes and procedures for metrics reporting, system monitoring, and configuration management
  - Establishment of a formal standards conformance program listing external systems verified to be interoperable with the ABIS biometrics repository
- **Using DOT&E's Evaluation concept, the Army Test and Evaluation Command is planning for a rigorous DT&E, supported by the Electronics Proving Ground, following by an independent IOT&E**
  - DOT&E is proposing a 30-day parallel operations IOT&E construct to ensure ABIS 1.2 is effective, suitable, and survivable before ABIS 1.0 is stood down
  - The IOT&E will include a Design of Experiments approach to scope the data collection and resulting analyses according to high risk factors most likely to affect performance

# Theme 3: Identifying Problems Early
## *Mine Resistant Ambush Protected Vehicles Example*

**Mine Resistant Ambush Protected (MRAP) vehicles are a family of vehicles designed to provide increased crew protection against battlefield threats, such as Improvised Explosive Devices (IED), mines, and small arms**

- **The MRAP Joint Program Office originally intended to conduct live fire testing against only the KPP threshold value level of explosive underbelly and side attack threats**
- **In addition to threats consistent with the KPPs, DOT&E advocated testing against larger explosive threats which were actually being seen in the theater where MRAPS were to be deployed**
- **The JPO agreed to the need for this testing against the larger threats**
  - Testing was conducted rapidly, and revealed significant vulnerabilities against the larger threats
- **The JPO decided to and was able to make rapid changes to the vehicle design prior to deployment**
- **Both Secretary Gates and Deputy Secretary Carter endorsed the testing and the rapid corrections it enabled prior to the deployment of the MRAPs**

# Theme 3: Identifying Problems Early

## *Airborne Laser Mine Detection System (ALMDS) Example*

**ALMDS is employed on the MH-60S helicopter to detect, classify, and localize shallow moored mines and floating and drifting mines on the surface. The system is included in the LCS MCM mission package**

- **During DT, Navy tested to requirements, which specify a narrow range of environmental conditions (including water clarity), this testing missed important system limitations**
  - Navy conducted all DT in the Gulf of Mexico near Panama City, Florida
- **OT community advocated testing in a more diverse set of operationally realistic environments**
  - Argued that operational testing should consider the threat environment and should not be limited to conditions specified in requirements documents
- **Navy conducted an operational assessment in a more challenging environment (poor water clarity) consistent with conditions that would be encountered in actual operations in the Persian Gulf**
  - System performed poorly
  - Pre-test predictions of water clarity were inaccurate
  - System water clarity measurements did not agree with values computed using tactics guide that reflected experience in DT

Acronyms this slide: Airborne Laser Mine Detection System (ALMDS); Developmental Testing (DT); Littoral Combat Ship (LCS); Mine Countermeasures (MCM); Operational Test (OT)

# Theme 3: Identifying Problems Early
## *Virginia-class Modernization FOT&E Example*

**Virginia is a multi-mission nuclear attack submarine that will replace the existing Los Angeles class. The Wide-Aperture Array (WAA) is a sonar array that provides estimates of the range to a target**

- **Insufficient at-sea developmental testing resulted in flawed Wide-Aperture Array (WAA) software**
- **Improving the WAA was a major focus of the Virginia modernization effort**
  - The WAA is installed on all Virginia-class submarines. It uses wave-front curvature to estimate the range to a sonar contact
  - The Virginia modernization effort included new processing algorithms for the WAA to improve its performance in the presence of loud, interfering contacts
  - Most development of the modernized WAA software was done in the laboratory; the final software version had not been fully tested at sea before the operational test
- **At-sea operational testing in March 2011 discovered that the new WAA provided inaccurate range estimates**
- **Based on the performance observed during operational testing, the Navy stood up a tiger team to address the problem, and issued a software fix in September 2011**

# Theme 3: Identifying Problems Early
## *Miniature Air Launch Decoy (MALD) Example*

**MALD is an expendable, air-launched, low-cost decoy designed to present the radar signature and flight characteristics of fighter, attack, or bomber aircraft against threat radar defense systems**

- **MALD is required to be successfully deployed after up to 60 hours of aircraft carriage time, operational testing highlighted a problem in this area**
  - This requirement derives from potentially long B-52 missions to transport MALD into the theater of operations
  - MALD is also required to withstand 30 days of uncovered/60 days of covered storage
- **The program office initially decided not to conduct long carriage time tests during DT**
- **DOT&E advocated long carriage time flights be conducted as part of the IOT&E**
- **During IOT&E, two missiles lost propulsion prematurely after being launched from a B-52 after each was carried more than 40 hours and spent a week on the wing of a B-52 in Guam, consistent with the way they would be employed in combat**
  - Failure analysis found water in the fuel bladder believed to be from a rainstorm while MALD was loaded on the wing of a B-52 sitting on the ramp
  - When the B-52 flew at high altitude, the water froze and blocked the MALD fuel filter during free flight leading to the flameouts
  - The Air Force incorporated a new fuel filter that could block ice crystals without restricting fuel flow

*Acronyms this slide: Developmental Test and Evaluation (DT&E); Initial Operational Test and Evaluation (IOT&E); Miniature Air-Launched Decoy (MALD)*

# Theme 3: Identifying Problems Early

## *Department of the Navy Large Aircraft Infrared Countermeasures System Example*

**Department of the Navy Large Aircraft Infrared Countermeasures (DoN LAIRCM) is an infrared missile warning and countermeasures system designed to provide protection of rotary-wing aircraft against shoulder-fired and vehicle-fired infrared-guided missiles**

- **Prior to its IOT&E, DoN LAIRCM had not been tested in mountainous terrain or during typical rotorcraft maneuvers. IOT&E testing uncovered a significant deficiency**
- **DOT&E advocated testing in more operationally realistic environments with multiple rotorcraft maneuvers, as would occur during operations in Afghanistan**
- **IOT&E was conducted in April 2009**
  - DoN LAIRCM did not declare threats for certain relevant engagement scenarios involving mountainous terrain and rotorcraft maneuvers
  - A major deficiency was identified with the missile declaration algorithms causing the system to ignore valid threat engagements due to threat range estimate errors
  - The Navy implemented a software upgrade that mitigated the issue. This was demonstrated to be effective through additional testing.
- **This deficiency and its correction were particularly relevant to operations in Afghanistan**

*Acronyms this slide: Department of the Navy (DoN); Initial Operational Test and Evaluation (IOT&E); Large Aircraft Infrared Countermeasures (LAIRCM)*

# Theme 3: Identifying Problems Early
## *Aegis Weapon System Example*

**The Aegis Weapon System (AWS) is the centerpiece of the combat system installed in Navy cruisers and destroyers**

- **OT conducted under realistic combat conditions identified critical system deficiencies that had not been discovered during developmental testing**

- **Testing examined air defense of an Aegis cruiser with newest AWS software in 2011**
  - Test ship was challenged with simultaneous cruise missile, slow aircraft, and small boat attacks
  - Because of radar and gun casualties that occurred before the test event, the ship had to maneuver at high speed to keep available sensors and weapons on approaching threats while slowing the closure rate of the attacking boats

- **Standard Missile engagement against cruise missile target failed when AWS performance problems caused the Standard Missile to self-destruct**
  - Post-event analysis identified three AWS performance deficiencies
  - Classified problem number 1 was caused by a software coding error; the Navy corrected this problem and re-tested the AWS
  - Classified problems 2 and 3, also software problems, were revealed as a consequence of problem 1. Correction could not be verified during at-sea testing because conditions could not be duplicated
  - Navy plans to verify correction of problems using M&S

- **Critical AWS deficiencies found during stressing OT event might not otherwise have been discovered until the ships were engaged in combat operations**

Acronyms this slide: Aegis Weapon System (AWS); operational testing (OT); anti-ship cruise missile (ASCM); modeling and simulation (M&S)

# Theme 3: Identifying Problems Early

## *Warfighter Information Networking – Tactical (WIN-T) Inc 2 Example*

**WIN-T Increment 2 is a mission command system that provides command and control to tactical units on-the-move via both satellite and terrestrial based transport**

- **WIN-T Increment 2 exhibited numerous performance problems in operational testing that were not identified in developmental testing**
  - Both the satellite and terrestrial networks were unstable
  - Range of the terrestrial waveform was very limited
  - The WIN-T user interface is overly complex. When troubleshooting, soldiers had to stop the vehicles and spend upwards of 30 minutes resolving problems, during which time they were out of the fight
  - System reliability was very low
- **Developmental testing failed to identify any of these issues**
  - Scaling the network to an operational size introduced these issues
  - Developmental testing used small numbers of configuration items, while operational testing fielded a full brigade
- **Poor performance in two operational tests (IOT&E in May 2012 and FOT&E in May 2013) has led USD(AT&L) and the Army to plan for additional operational testing (FOT&E2 in November 2014) prior to a full-rate production decision**
  - The Program Office agreed the problems identified during operational testing are significant, and is continuing efforts to resolve the network stability and system complexity issues

Acronyms this slide: Warfighter Information Network-Tactical (WIN-T)

# Theme 3: Identifying Problems Early
## *Cooperative Engagement Capability on the E-2D Hawkeye Example*

**The Navy's Cooperative Engagement Capability (CEC) facilitates data fusion from multiple surface (e.g., DDG 51s) and airborne platforms (e.g., E-2D) for a more accurate tactical air picture**

- **Operational testing of the CEC on E-2D led to the discovery of important problems**
- **DOT&E recommended that the USG-3B CEC variant on the E-2D be tested to satisfy E-2D requirements and to demonstrate performance of the upgraded USG-3B signal data processor**
- **Operational testing revealed several CEC deficiencies that degraded the USG-3B CEC's performance, relative to the version installed on the legacy E-2C Hawkeye, and the E-2D's capability**
  - CEC's ability to maintain consistent air contacts (i.e., tracks) on each CEC node was degraded relative to the E-2C version
  - CEC created dual tracks well in excess of historic values
  - CEC interoperability problems with E-2D mission computer degraded the single integrated air picture presented to the various participating combat systems (e.g., the other E-2Ds and CEC equipped ships)
- **Using DOT&E's analysis, ASN(RD&A) decided to delay the Full Rate Production Decision for USG-3B until the Navy identified the root causes of the degrade CEC performance and associated corrections**

# Theme 3: Identifying Problems Early
## *Advanced Anti-Radiation Guided Missile (AARGM)*

**AARGM is a Navy program that replaces the HARM missile for use in suppressing enemy air defenses. AARGM improves HARM performance by attacking defense units using shutdown countermeasures. It also provides geo-location techniques that HARM does not have**

- **Operational mission failures and deficiencies were discovered during several operational testing phases**
  - Developmental testing did not uncover any of the major problems discovered in the first operational testing period, which were so severe that the Navy directed halting the operational test
  - After corrections were made, results from a developmental testing phase indicated the system was ready for operational testing; this second phase of operational testing lead to further discoveries of deficiencies
  - After further corrections of deficiencies were accomplished, a series of operational tests was conducted; based on the test results, the Navy proceeded with full-rate production and fleet introduction of the weapon
- **The weapon system still does not provide adequate suppression of enemy air defenses in some cases**
  - Testing of one key performance parameter has been deferred
  - Although the weapon performs adequately against some air defense capabilities, it has proven inadequate in suppressing threats operating in specific modes
  - There are additional deficiencies in other areas (classified)
- **Congress has provided funds for a Block 1 Upgrade program that is to fix all known deficiencies**
  - A follow-on operational test will verify that the weapon can suppress the entire set of threat systems that it is required to engage

Acronyms this slide: Advanced Anti-Radiation Guided Missile (AARGM); High Speed Anti-Radiation Missile (HARM)

# Theme 3: Identifying Problems Early
## *B-2 Bomber Radar Modernization Program (RMP) Example*

**The B-2 RMP program was designed to replace the original mechanical scan radar with a modern electronically steered array radar**

- **The program planned to change the antenna pattern of the B-2 radar without testing or detailed analysis of how it may affect the vulnerability of the B-2 to enemy radars**
- **The program office planned to change the antenna pattern of the B-2 radar**
  - The intent of the change was to put more power in the main antenna beam
  - Changing the main antenna beam, however, changes the power in antenna side lobes. These side lobes can make the B-2 more vulnerable to enemy radars
- **DOT&E and AFOTEC advocated side-by-side comparative testing to ensure the B-2 with RMP was not more vulnerable than the legacy B-2**
- **Further detailed analysis by the program office revealed that the new antenna pattern could be more vulnerable than the legacy pattern**
- **As a result, the program office decided to retain the legacy antenna pattern**

Acronyms this slide; Radar Modernization Program (RMP); Air Force Operational Test and Evaluation Center (AFOTEC)

# Theme 3: Identifying Problems Early
## *Patriot Example*

**Patriot is a U.S. Army air and missile defense system that protects deployed forces and critical assets from short- and medium-range ballistic missiles and aircraft threats**

- **The U.S. Army wanted to use only one Patriot interceptor in flight tests against ballistic missiles, even though their tactical doctrine was to fire two interceptors**
- **DOT&E advocated that all flight tests against ballistic missiles in DT/OT and OT use the tactical ripple firing doctrine of two Patriot interceptors per threat missile**
  - The Army agreed to this and used two interceptors in the last Patriot DT flight test prior to DT/OT in 2001
- **During IOT&E flight testing in 2002, a PAC-3 missile ripple fire mission was attempted using missiles from the same launch tube in different canisters for the first time**
  - The first missile launched correctly from the lower-left launch tube in the upper-left canister, but the second missile failed to launch from the lower-left launch tube of the lower-right canister
  - The failure to launch was caused by a built-in test software code in the guidance processor unit that only manifested itself when the two interceptors in a ripple fire engagement were located in the same launch tube in their canister (in this case, the lower-left)
- **This particular built-in test software code was leftover by the contractor from early missile development and was not needed for operational use; so, it was removed from the missile after this problem was revealed**

Acronyms this slide: Development Test /Operational Test (DT/OT); Initial Operational Test and Evaluation (IOT&E); Operational Test (OT); Patriot Advanced Capability-3 (PAC-3)

# Theme 3: Identifying Problems Early
## *Key Management Infrastructure (KMI) Example*

**KMI generates, distributes, and manages cryptographic products that enable assured identification of users of weapon systems and weapon systems components**

- **KMI Increment 2, Spiral 1 had two rounds of developmental testing, including one round executed by an independent Developmental Test Team from the Joint Interoperability Test Command (JITC)**
  - Both rounds of DT met test objectives by the test organizations in functional testing
  - Results of the second, independent evaluation led to a decision to proceed to Operational Assessment
- **Operational Assessments, executed by JITC prior to deployment for Initial Operational Test and Evaluation (IOT&E) and using mission-based scenarios, discovered nearly 1,000 software defects**
  - Over 200 defects were discovered on the first day of the Operational Assessment, many of which affected mission-critical capabilities
  - Most, if not all, defects should have been discovered during a robust DT effort
- **Deploying the system for operational use after the final round of DT, without operational test and evaluation, would have rendered users unable to deliver critical cryptographic products to the field**
  - All mission-critical defects were corrected and re-tested prior to the deployment for IOT&E allowing users to accomplish missions using the KMI system

Acronyms this slide: Key Management Infrastructure (KMI) ; Joint Interoperability Test Command (JITC); Initial Operational Test and Evaluation (IOT&E)

# Theme 4

Operational testing can identify issues with systems' combat effectiveness. This information is provided to acquisition authorities for decision. Decisions by those authorities to fix problems can affect program schedules if they judge the associated fix worth the delay

Additional examples can be found in the *Reasons Behind Program Delays* briefing

# Theme 4: Fixes, Not Testing, Delays Programs
## *Cooperative Engagement Capability on the E-2D Hawkeye Example*

**The Navy's Cooperative Engagement Capability (CEC) facilitates data fusion from multiple surface (e.g., DDG 51s) and airborne platforms (e.g., E-2D) for a more accurate tactical air picture**

- **Operational testing of the CEC on E-2D led to the discovery of important problems**
- **DOT&E recommended that the USG-3B CEC variant on the E-2D be tested to satisfy E-2D requirements and to demonstrate performance of the upgraded USG-3B signal data processor**
- **Operational testing revealed several CEC deficiencies that degraded the USG-3B CEC's performance, relative to the version installed on the legacy E-2C Hawkeye, and the E-2D's capability**
  - CEC's ability to maintain consistent air contacts (i.e., tracks) on each CEC node was degraded relative to the E-2C version
  - CEC created dual tracks well in excess of historic values
  - CEC interoperability problems with E-2D mission computer degraded the single integrated air picture presented to the various participating combat systems (e.g., the other E-2Ds and CEC equipped ships)
- **Using DOT&E's analysis, ASN(RD&A) decided to delay the Full Rate Production Decision for USG-3B until the Navy identified the root causes of the degraded perofrmance and associated corrections**

# Theme 4: Fixes, Not Testing, Delays Programs
## *MH-60S Helicopter with AQS-20A Towed Mine Hunting Sonar*

**The AQS-20A sonar towed by the MH-60S helicopter was intended to provide Naval combatant ships with an organic mine hunting capability**

- **Testing of the AQS-20A on the MH-60S identified important problems**
- **Towing a sonar is a challenging mission for the MH-60S helicopter**
  - Testing examined tow cable reeling mechanism
  - MH-60S was operating at the edge of its power envelope
- **DT and Early OT through 2007 revealed numerous deficiencies**
  - Cable reeling mechanism failures led to jettisoning the AQS-20A
  - Helicopter lacked sufficient power to tow effectively
  - AQS-20A mine detection deficiencies under certain operationally realistic conditions
- **Navy decertified the system from test in 2008 in order to fix the problems**
  - Testing did not resume until 2011
  - Testing was downgraded from an OT to an Operational Assessment
  - System continued to exhibit problems
- **DOT&E issued an Operational Assessment report in 2012**
  - Assessed the MH-60S to be underpowered
  - The Navy and DOT&E agreed that the helicopter cannot safely perform the tow mission
- **Navy cancelled plans to deploy the AQS-20A from the MH-60S helicopter**

# Theme 4: Fixes, Not Testing, Delays Programs
## *LPD 17 San Antonio Class Example*

**LPD-17 is the Navy's newest class of amphibious transport dock ships, replacing four classes of amphibious ships**

- **Operational and Live Fire testing of the first ship of the class was delayed because of problems during construction**

- **LPD 17 was delivered to the Navy in December 2005, but was in such an incomplete state of construction that she was not available for testing**
  - LPD 19 was shock tested in August 2008
  - The Total Ship Survivability Trial was conducted on LPD 18 in September 2008
  - LPD 17 made the first class deployment starting in August 2008, without the benefit of this critical testing

- **Design and reliability issues identified during testing resulted in an assessment of not operationally survivable in a combat environment**
  - Excessive loss of critical systems including chill water, radars, steering, and diesel generators resulted from a poor plan to provide backup power when needed
  - Firemain, chilled water, potable water, and JP-5 systems lack sufficient isolation valves and sensors to segregate damaged section of the systems

- **The Navy has identified a number of corrective actions for implementation on future ships and for back fit to existing ships**
  - For example, results of tests were the basis for changes to Navy surface ship firefighting doctrine for ships with well decks

# Theme 4: Fixes, Not Testing, Delays Programs
## *F-35 Joint Strike Fighter Example*

**The F-35 is a tri-Service, multi-national, family of strike aircraft. It will replace the F-16 and A-10 in the Air Force, the AV-8B in the Marine Corps, and augment the F-18 fleet in the Navy**

- **F-35 testing was planned to determine whether the F-35 system was ready to start training pilots in a limited, early configuration of the aircraft, but it was not clear that this testing could be safely conducted**

- **DOT&E identified concerns prior to the start of the F-35A Ready-for-Training Operational Utility Evaluation and recommended delaying the start of the evaluation until the concerns were addressed**
  - The observed air abort rate – a metric used to assess system maturity and safety – was three times higher than objective rates identified by the Air Force to begin training
  - Aircraft configuration shortfalls in the integrated power package grounded the fleet earlier in the year and needed to be corrected
  - Maintenance and sustainment operations were still immature making sortie generation difficult
  - The water-activated parachute release system had not been qualified nor fielded yet in the ejection seats of the fielded aircraft. Lack of the release system added safety risk to overwater flights, as would be flown during the OUE
  - Updating and re-certification of the pilot simulator were necessary
  - Pilot workload assessments and deficiencies in the Integrated Caution and Warning System had not been adequately addressed

- **The Air Force agreed with many of DOT&Es concerns and delayed the OUE until a rigorous plan for conducting the OUE safely was developed for approval by the Air Force Vice Chief of Staff**

# Theme 4: Fixes, Not Testing, Delays Programs

## *Warfighter Information Networking – Tactical (WIN-T) Inc 2 Example*

**WIN-T Increment 2 is a mission command system that provides command and control to tactical units on-the-move via both satellite and terrestrial based transport**

- **WIN-T had an IOT&E in May 2012, with a planned full-rate production decision (FRP) in September 2012**
- **Problems identified in the IOT&E resulted in a USD (AT&L) decision to proceed with a 2nd LRIP decision instead of FRP, and plan an FOT&E. The identified problems included:**
  - Poor network stability of both the terrestrial and satellite networks. The terrestrial network also showed very limited range
  - Poor performance of the company based Soldier Network Extension
  - Poor reliability of all configuration items
- **The FOT&E was conducted in May 2013. Continuing problems led the USD (AT&L) to decide upon a 3rd LRIP instead of an FRP, and plan for an FOT&E2 in November 2014. Identified problems included:**
  - Poor performance of the company based Soldier Network Extension
  - Poor range of the terrestrial based network
  - Overly complex and non-intuitive interfaces on leader vehicles
- **The agreed need to correct the problems cited above caused USD(AT&L) and the Army to plan for a full-rate production decision in May 2015, vice an original goal of September 2012**

Acronyms this slide: Warfighter Information Networking-Tactical (WIN-T), Initial Operational Test and Evaluation (IOT&E), Follow-On Operational Test and Evaluation (FOT&E), Low Rate Initial Production (LRIP)

# Theme 4: Fixes, Not Testing, Delays Programs
## *Expeditionary Fighting Vehicle Example*

**The Expeditionary Fighting Vehicle (EFV) was a high water speed, fully tracked, armored amphibious personnel carrier intended to support amphibious landings from over the horizon and mechanized land operations**

- **EFV was progressing toward its 2007 Milestone C review with no significant known issues based on reported developmental test results**
- **Because of poor system performance during OT&E in 2002, DOT&E advocated the 2006 pre-Milestone C OT to address realistic combat loads, reliability, and water operations. The 2006 OT identified critical system and operational shortfalls not seen during developmental testing because developmental testing was not conducted in an operationally representative context**
  - In the OT, the EFV-equipped unit completed its assigned mission in two of 14 attempts
  - Low reliability, poor system availability, and high maintenance burden were the major contributing factors to unsuccessful mission performance. During OT missions, the Mean Time Between Operational Mission Failures estimate was 4.5 hours, which was significantly below both the 15- hour estimate from developmental testing and the 43.5-hour KPP threshold.
  - Water performance and armor protection (both KPPs) were compromised because 1,800 pounds of armor had to be removed to allow the EFV to achieve high water speed when combat loaded
- **DOT&E's analysis of OT results and briefings to OSD, Department of the Navy, and Service acquisition leaders – coupled with a Nunn-McCurdy breach – led USD(AT&L) to conclude in June 2007 that the program was not ready to proceed into low-rate initial production**
  - The EFV's Engineering and Manufacturing Development phase was extended at a cost of approximately $1B and a 4-1/2 year delay to allow redesigned prototype vehicles to be acquired and additional testing to be conducted, including another OT event
  - The Secretary of Defense cancelled the EFV program in January 2011 after the Secretary of the Navy and Commandant of the Marine Corps concluded that EFV procurement and operations & maintenance costs would be unaffordable

 Acronyms this slide: KPP - Key Performance Parameter OSD - Office of the Secretary of Defense USD(AT&L) - Undersecretary of Defense (Acquisition, Technology, & Logistics)

# Theme 4: Fixes, Not Testing, Delays Programs

## *AH-1Z Attack and UH-1Y Utility Helicopters Example*

**This program replaces legacy 2-bladed attack and utility helicopters with 4-bladed attack and utility helicopters with identical twin engines, drive trains, tail sections, digital cockpits, and helmet-mounted displays**

- **In the IOT Phase I in 2006, DOT&E reported that both aircraft were not effective because of poor helmet performance**
  - The helmet was not qualified (by developmental testers) for use at night
  - As a result, night operations were not attempted with either aircraft
  - To be operationally effective, both aircraft, per their Concepts of Operations, must be employable at night
- **In the IOT Phase II in 2008, the UH-1Y was effective and suitable, but the AH-1Z targeting sensor was so unreliable that the attack aircraft could only complete 37 percent of assigned missions successfully**
  - With helmet redesign and other reliability improvements, the UH-1Y helicopter effectively completed all assigned daytime and nighttime missions
  - The Navy withdrew the AH-1Z from testing due to its targeting sensor issues
  - Development and integration of the AH-1Z targeting sensor continued
- **In the IOT Phase III in 2010, DOT&E reported the AH-1Z effective, suitable, and survivable.**
  - The reliability of the AH-1Z targeting sensor and other components improved
- **Both aircraft have since been deployed in support of operations in Iraq, Afghanistan, and other Navy operations worldwide**

# Theme 4: Fixes, Not Testing, Delays Programs
## *Ground-Based Midcourse Defense (GMD) Example*

**GMD defends the U.S. Homeland from intercontinental ballistic missile threats from North Korea and Iran**

- **In 2010, the GMD interceptor – using an upgraded Exoatmospheric Kill Vehicle (EKV) called the Capability Enhancement-II (CE-II) EKV – experienced a second consecutive failure in its flight test program**
  - GMD interceptors using the previous CE-I EKV version had been successful in three prior flight tests, but these tests were conducted within a less challenging engagement scenario
- **DOT&E and MDA agreed to conduct an intercept flight test with an interceptor using a CE-I EKV within the more challenging engagement scenario**
  - MDA conducted this test in 2013, and the interceptor using the CE-I EKV failed within this engagement scenario
- **DOT&E advised that the test results raised questions regarding the robustness of the EKV design and the adequacy of the system engineering used in the design and recommended that MDA consider some redesign of the EKV**
- **USD (AT&L) determined that problems exist with the GMD interceptor. MDA will be taking an initiative to address these problems**
  - USD (AT&L) has stated that the interceptor problems were caused by "bad engineering" that was attributable to decisions to rush deployment of technologies that had not been completely and thoroughly tested

Acronyms this slide: Capability Enhancement-I (CE-I); Capability Enhancement-II (CE-II); Exoatmospheric Kill Vehicle (EKV); Ground-Based Midcourse Defense (GMD); Missile Defense Agency (MDA)

# Theme 4: Fixes, Not Testing, Delays Programs
## *Key Management Infrastructure (KMI) Example*

**KMI generates, distributes, and manages cryptographic products that enable assured identification of users of weapon systems and weapon systems components**

- **Operational Assessments executed by an Operational Test Team from JITC prior to deployment for Initial Operational Test and Evaluation (IOT&E) and using mission-based scenarios discovered nearly 1,000 software defects**
  - Over 200 defects were discovered on the first day of the first Operational Assessment.
  - Most defects should have been found during earlier developmental testing
  - Defects were subsequently fixed by the developer, but new defects were introduced; in some cases functions that worked in the previous release no longer worked in the “fixed” release
- **Numerous rounds of regression testing to find and fix defects in critical capabilities resulted in a nearly one-year slip in the development schedule**
  - Program Manager declared a critical program breach to have time to fix all of the problems.
  - Program Manager also directed the developer to use the JITC-developed scenarios in regression testing to ensure mission threads could be executed
- **Initial Operational Test and Evaluation (IOT&E) had few critical software problems, and the majority of test issues were in configuration management, training, and other suitability areas**
  - Follow-on Operational Test and Evaluation to verify correction of defects found in IOT&E resulted in DOT&E finding the system operationally effective, suitable, and survivable (from a cybersecurity perspective)

Acronyms this slide: Key Management Infrastructure (KMI) ; Joint Interoperability Test Command (JITC); Initial Operational Test and Evaluation (IOT&E)

# Theme 5

DOT&E has been responsive and supportive of fielding systems for the current wars

# Theme 5: Support to Wars
## *Mine Resistant Ambush Protected Vehicles Example*

**Mine Resistant Ambush Protected (MRAP) vehicles are a family of vehicles designed to provide increased crew protection against battlefield threats, such as Improvised Explosive Devices (IED), mines, and small arms**

- **Because of the urgent operational need for increased crew protection against battlefield threats in Iraq and Afghanistan, multiple MRAP vehicle configurations had to be procured, tested, and fielded on a highly accelerated basis**

- **From November 2007 to June 2013, DOT&E provided oversight and reporting of 10 operational test (OT) events of 18 different MRAP vehicle configurations**
  - To identify vehicle design aspects that could compromise mission performance and to surface failure modes likely to be experienced in combat, DOT&E advocated greater OT attention to conducting the types of combat missions that units equipped with MRAPs were expected to do and to vehicle reliability
  - Based on data from the 10 OT events and other information sources, DOT&E reported that 10 of 18 MRAP vehicle configurations were operationally effective and operationally suitable
  - Other classified findings regarding survivability affected two variants

- **For the 8 configurations that were not effective or not suitable, OT findings contributed to:**
  - Essential vehicle modifications, such as the retrofit of an independent suspension system to the widely used Navistar MaxxPro Dash to support operations in Afghanistan type terrain
  - Additional systems engineering and developmental testing to investigate problems found during OT, such as the weapon system failures that significantly compromised the reliability of the Special Operations Forces version of the MRAP-All Terrain Vehicle (SOF M-ATV)
  - Information to support Service and Department decisions on which vehicle types to divest and which to retain as part of the Services' enduring fleet of MRAP vehicles. For example, the Navistar MRAP Recovery Vehicle demonstrated very limited mobility and poor reliability during OT, and was not retained in the Services' enduring fleet. As another example, the Dash Ambulance variant was found not operationally effective primarily because it could not safely accommodate litter patients taller than 5 feet 11 inches; it was not retained either.

- **Key survivability corrections were made to two variants at the direction of Secretary Gates**

# Theme 5: Support to Wars
## *Stryker Double-V Hull (DVH) Family of Vehicles Example*

**The Stryker DVH provided integrated enhancements to the base Stryker vehicle to improve both crew survivability and mobility. It was a limited production effort provided to units employed in Operation Enduring Freedom**

- **In response to an Army Operational Need Statement in December 2009, the Defense Acquisition Executive issued an Acquisition Decision Memorandum in April 2010 authorizing the Army to procure the Stryker DVH vehicles**
  - DOT&E worked with the Army to develop a phased test plan that supported the production and fielding schedule
  - Testing focused on the minimum necessary to support intended operations in Afghanistan, establish required levels of survivability, and no degradation to operational performance
- **Operational and live fire testing was synchronized with production and fielding schedules**
  - Initial test assets were delivered in August 2010
  - Initial live fire testing on the infantry carrier variant was complete on 20 December 2010
  - Operational testing on the infantry carrier variant was complete on 25 February 2011
  - The Army reviewed the emerging results test data and recommended continued production on 2 March 2011
  - The Defense Acquisition Executive approved continued production on 12 May 2011
  - Fielding of the infantry carrier variant began in Afghanistan in June 2011
- **DOT&E found the DVH operationally effective and suitable (except for the driver's compartment), and survivable**
  - DOT&E recommended that the Army correct deficiencies noted in the design of the driver's station before deployment
- **The Army fielded corrections to the driver's station before fielding in June 2011**

# Theme 5: Support to Wars
## *Common Missile Warning System Example*

**The Army's Common Missile Warning System (CMWS) provides warning of infrared-guided surface-to-air threats launched against aircraft and cues a countermeasure**

- **Results from the 2005 Initial Operational Test and Evaluation showed that CMWS was not effective against all of the required threats in the worldwide threat environment**
  - Because of the urgent need to field CMWS to protect Army aircraft and soldiers operating in OIF/OEF, the Army had optimized the CMWS software for that threat environment
  - Test results showed that CMWS performed well against the OIF/OEF threats in the relevant operational environment and using operational tactics
  - Test results showed that CMWS did not perform well against additional threats outside of those employed in OIF/OEF
- **DOT&E reported that CMWS was effective for employment in OIF/OEF in its report to Congress even though it did not perform well against threats in other theaters**
  - Rapid testing and reporting supported rapid deployment of CMWS to OIF/OEF
- **The Army proceeded with a Full Rate Production decision and began to quickly field many systems**
  - CMWS was repeatedly credited with saving aircraft and aircrew from MANPADS

Acronyms this slide: Common Missile Warning System (CMWS); Operation Iraqi Freedom (OIF); Operation Enduring Freedom (OEF)

# Theme 5: Support to Wars
## *MQ-1C Gray Eagle Example*

**The MQ-1C Gray Eagle-equipped unit is an unmanned aircraft system that provides tactical intelligence, video, radar imagery, communications relay, and precision missile support to Army maneuver units**

- **DOT&E supported rapid fielding of ISR assets to support operations in Afghanistan**
- **In response to a SECDEF initiative to rapidly field ISR assets to support wartime operations, DOT&E worked with the Army to structure Gray Eagle operational testing to combine it with training for deploying units**
  - The Army agreed to a Limited Users Test in 2009 to support testing, training and fielding of a Gray Eagle platoon to Iraq
  - The Army agreed to a second Limited Users Test in 2010 to support testing, training and fielding of a Gray Eagle platoon to Afghanistan
- **Test results were used by the Army to prioritize and correct shortfalls in performance and training; units participating in the tests praised the training opportunities the testing provided**
- **Deployed MQ-1C units achieved distinction in both theaters and remain deployed in support of operations in Afghanistan**

Acronyms this slide: Intelligence, Surveillance, and Reconnaissance (ISR), Secretary of Defense (SECDEF)

# Theme 5: Support to Wars
## *Mk 83 Joint Direct Attack Munition Example*

**The Mk 83 Joint Direct Attack Munition (JDAM) is a 1,000-pound class conventional gravity bomb with a Global Positioning System-aided inertial guidance system and movable tail fins to help steer it accurately to the target**

- **Wartime needs would not have been met following the normal acquisition cycle**
  - Original initial operational capability was scheduled for Fall 2003
  - The events of 9/11 introduced an urgent need for the weapon much earlier
- **DOT&E supported accelerated operational testing to provide timely delivery of the weapon to the warfighter**
  - An operational readiness review was held in December 2001
  - A quick reaction weapon assessment (testing) was conducted in January and February 2002
  - A DOT&E evaluation was issued in March 2002 supporting an early operational capability of the weapon's employment in combat
  - Additional operational testing was conducted from July 2002 to February 2003 to fully characterize the weapon and its proper employment during conflict
- **The weapon was used effectively during Operation Iraqi Freedom, with 1,400 weapons expended**

Acronyms this slide: Joint Direct Attack Munition (JDAM)

# Theme 6

Rigorous scientific and statistical methods are used to determine adequacy of operational testing in terms of miles driven, shots fired, and coverage of the operational envelope. Sound statistical analysis of operational test results provides accurate estimates of information to acquisition decision-makers

# Theme 6: Rigorous Statistical Methods
## *Mine-Resistant Ambush-Protected (MRAP) Vehicles Example*

**Mine Resistant Ambush Protected (MRAP) vehicles are a family of vehicles designed to provide increased crew protection against battlefield threats, such as Improvised Explosive Devices (IED), mines, and small arms**

- **The MRAP Joint Program Office (JPO) initially proposed testing with limited instrumentation that would not have provided an adequate assessment of crew casualties**
  - Only two crew positions would have anthropomorphic test devices (ATD) to measure the forces that cause accelerative injury; injuries for other positions would be interpolated
  - Injury reporting would be focused on its impact on mission completion (personnel incapacitation), rather than injury severity
- **DOT&E advocated, and the JPO readily agreed to, improvements in test conduct and injury assessment methodologies**
  - ATDs were provided for all crew positions and test range data acquisition systems were upgraded to handle the increased amount of data
  - Army Research Laboratory coordinated with military and civilian medical community to improve injury assessments reflecting injury severity
  - The Services and DOT&E used the improved injury assessments to assess overall vehicle protection capabilities in a combat engagement
- **LFT&E provided detailed time-history data from ATDs in all crew positions to quantify injuries to the crew**
  - Instrumenting all personnel positions revealed unexpected, counterintuitive patterns of injuries – patterns that would have been missed or misconstrued through interpolation.
  - Individual injury severity and measures of overall morbidity were more directly responsive to user requirements
  - The investments in range improvements and assessment methodologies advanced the state-of-the-art testing capabilities
- **The results obtained enabled the JPO to rapidly redesign the MRAP vehicles to reduce injuries discovered during LFT&E, results which were directly correlated with improved outcomes in combat operations**

Acronyms this slide: Anthropomorphic Test Device (ATD), Joint Program Office (JPO)

# Theme 6: Rigorous Statistical Methods
## *Enhanced Combat Helmet (ECH) Example*

**Forces equipped with the ECH will rely on the helmet to provide ballistic protection from selected threats. The ECH is more resistant to penetration than previous combat helmets**

- **DOT&E approved a test protocol to ensure statistically principled ballistic testing of combat helmets**
  - The protocol specifies the number of helmets tested, the statistical measures of merit, and the criteria for acceptance.
  - Statistically-principled testing assures it is unlikely that the government will accept and field a helmet that performs poorly
  - The ECH was the first combat helmet to follow the DOT&E-approved protocol
- **Following an engineering change proposal intended to increase manufacturing capacity, the ECH failed the small arms component of the DOT&E-approved protocol**
  - The helmet failed because of too many small arms penetrations, which demonstrated that the helmet did not provide the desired protection
  - Multiple attempts to implement and verify corrective actions within the manufacturing process failed to produce a helmet with acceptable performance
- **The manufacturer ultimately decided it was necessary to use different ballistic shell laminate material to provide for an acceptable helmet against the small arms threat**

Acronyms this slide: Enhanced Combat Helmet (ECH)

# Theme 6: Rigorous Statistical Methods

## *Integrated Defensive Electronic Countermeasures (IDECM) Block 4 Example*

**The IDECM Block 4 system is a modernization of the F/A-18 Radiofrequency jammer for the F/A-18E/F and provides carrier-capable jamming against advanced threats for the first time for the F/A-18C/D.**

- **Modern statistical test design techniques allowed the number of test events to be reduced**
- **Follow-on Test and Evaluation (FOT&E) test planning examined how many repetitions of each flight test condition against a given threat radar are needed to prove effectiveness**
  - COTF proposed 10 replicates per condition, for a total of 95 runs, based on single roll-up for each factor
  - Initial DOT&E experimental designs for the FOT&E indicated 3 to 4 replicates, for a total of 29 to 38 runs, were sufficient
- **DOT&E recommended a design of experiments (DOE) approach to create an adequate OA and FOT&E**
  - All parties agreed to use the Operational Assessment (OA) to evaluate the DOE approach
- **The results of the OA showed the flexibility and strength of DOE**
  - Testing used 38 planned runs, and although multiple test points were missed during test execution, enough data under the proper test conditions were collected during the OA to make preliminary effectiveness determinations for a sub-set of required threats
  - The test also indicated significant run-to-run and day-to-day variations, reinforcing the importance of randomization on test condition order
  - COTF saw a benefit in the DOE approach and is redesigning the FOT&E. This design will allow for flexibility in test execution and clear determination of primary and secondary effects on IDECM Block 4 performance

*Acronyms this slide: Commander, Operational Test and Evaluation (COTF); Design of Experiments (DOE); Follow-on Operational Test and Evaluation; Integrated Defensive Electronic Countermeasures (IDECM); Operational Assessment (OA)*

# Theme 6: Rigorous Statistical Methods
## *F-35 Joint Strike Fighter Example*

**The F-35 Joint Strike Fighter is a tri-Service, multi-national, family of strike aircraft of three variants. It will replace the F-16 and A-10 in the Air Force, the AV-8B in the Marine Corps, and augment the F-18 fleet in the Navy**

- **An efficient test design to examine the operational effectiveness of the Block 2B F-35 across mission areas and vast array of operating conditions was developed**

- **The multi-mission capabilities of the F-35 result in a complex and large operational battle space for possible combat operations**
  - With initial war fighting capability – designated Block 2B capability – the Services state that the F-35 will be expected to conduct a variety of missions including: air-to-ground strike, close air support, suppression/destruction of enemy air defense, reconnaissance, and limited air-to-air operations
  - If it is conducted, the Block 2B operational testing will occur across a variety of complex conditions including varying threat levels (both ground and air), times of day, and a variety of targets.

- **DOT&E advocated a scientific test planning process that assures efficient and adequate investigation of all missions and operational conditions**
  - The operational test team identified mission-oriented measures that reflect the operational capability of the system
  - The operational test team created a highly efficient test design to cover the planned operational space within each mission area. For example, for air-to-ground missions against preplanned strike targets, the test team identified 128 test operating conditions; the test design uses 21 trials to adequately cover those conditions.

- **If it is executed, the resulting F-35 Block 2B test design will systematically cover the operational battle space with efficient use of test resources**

# Theme 6: Rigorous Statistical Methods
## *AIM-9X Sidewinder Missile Example*

**The AIM-9X Sidewinder is the Department's short range, infrared guided air-to-air missile. The latest version, Block II, provides lock on after launch capability along with guidance improvements**

- **As a result of IMU failures that occurred during the AIM-9X Block II IOT&E, the Services decided to pause the test in 2013**
- **The pause in testing led to a review of the original IOT&E plan size**
  - AIM-9X Block II IOT&E was paused on 29 July 2013 after multiple IMU flight test failures caused by the shock that the missile experiences during launch
  - During the pause, the program office developed a new manufacturing process for the IMU, along with software to improve guidance
  - Also during the pause, COTF, AFOTEC, and DOT&E examined what portion of the original IOT&E needed to be reexamined once testing recommenced
  - Power of test calculations revealed that the total number of captive carry tests could be reduced by about half while maintaining a statistically significant test
- **The restarted IOT&E will have about half the captive carry tests required in the original plan**

Acronyms this slide: Air Intercept Missile (AIM); Inertial Measurement Unit (IMU); Initial Operational Test and Evaluation (IOT&E); Commander, Operational Test and Evaluation Force (COTF); Air Force Operational Test and Evaluation Center (AFOTEC); Director, Operational Test and Evaluation (DOT&E)

# Theme 6: Rigorous Statistical Methods
## *Ballistic Missile Defense System (BMDS) Example*

**The BMDS is a collection of weapons, command and control systems, and sensors that over time is intended to protect the United States, deployed forces, allies, and friends against ballistic missiles of all ranges and in all phases of flight**

- **Traditional statistical approaches to quantify the performance of the BMDS have been hampered by the severely-limited data environment of the BMDS. One key metric, in particular, in the evaluation of BMDS performance is the probability of engagement success (PES)**
- **PES is the probability of successfully negating a ballistic missile threat and is extremely difficult to estimate through flight testing**
  - Individual intercept flight tests can cost up to $300 million each
  - Complex tests can take years to plan
  - Few test locations are suitable for intercept testing
- **In response, DOT&E developed a novel, new methodology for assessing PES that provides a means for including end-to-end tests as well as partial and component tests that other methods cannot accommodate**
  - The ability to incorporate partial and component tests is crucial because the cost and logistics of performing end-to-end system tests places significant restrictions on the amount of data that is available
  - Inclusion of partial tests allows for greater sampling of the BMDS battle space and, therefore, a more robust capability to quantify operational performance
- **Since 2010, DOT&E has incorporated the methodology into its annual BMDS assessment reports to Congress, providing Congress with the first quantitative, vice subjective, estimates of BMDS performance over the tested battle space**



Acronyms this slide: Ballistic Missile Defense System (BDMS); Probability of Engagement Success (PES)