

# The Role of Test and Evaluation in Defense System Requirements



Dr. J. Michael Gilmore

*Director, Operational Test and Evaluation*



# Operational Test: Mission Focused

- “Operational test and evaluation means --
  - *the field test, under realistic combat conditions, of any item of (or key component of) weapons, equipment, or munitions for use in combat by typical military users; and the evaluation of the results of such test.*” – 10 USC Section 139
- Operational testing is about assessing mission accomplishment of the unit equipped with the system under test
  - Effectiveness: can a unit equipped with the system under test accomplish the mission?
  - Suitability: can the system be used in the operational environment by the user to accomplish the mission?

**The *scope* of testing and how the system is *evaluated* should be linked to the requirements**

- ***End-to-End mission-oriented metrics are essential for evaluating operational effectiveness, suitability, and survivability***
- ***Testing must occur across the entire operational envelope***



# Mission Oriented Metrics & Key Performance Parameters (KPPs)

**KPP Def'n:** *A quantitative system attribute that the warfighter considers critical to the development of an effective military capability*

- DOT&E has seen many recent examples of KPPs that are not informative about Mission Accomplishment:
  - Systems that did not meet KPPs but were found operationally effective (e.g., F-22, USS Virginia, EA-18G)
  - Systems that did meet all KPPs but gave no operational value to the unit (e.g., E-IBCT – remnants of the Army's FCS)
- The KPPs should provide a determination of mission accomplishment and encapsulate the reasons for procuring the system
  - Metrics or Response variables for the KPPs should lend to good experimental design

## **Reporting Operational Test and Evaluation Results - - -**

*"The Director shall prepare a report stating the opinion of the Director as to whether the results of operational test and evaluation confirm that the items or components actually tested are effective and suitable for combat...."* - 10 USC Sect 2399



# Leveraging T&E knowledge

---

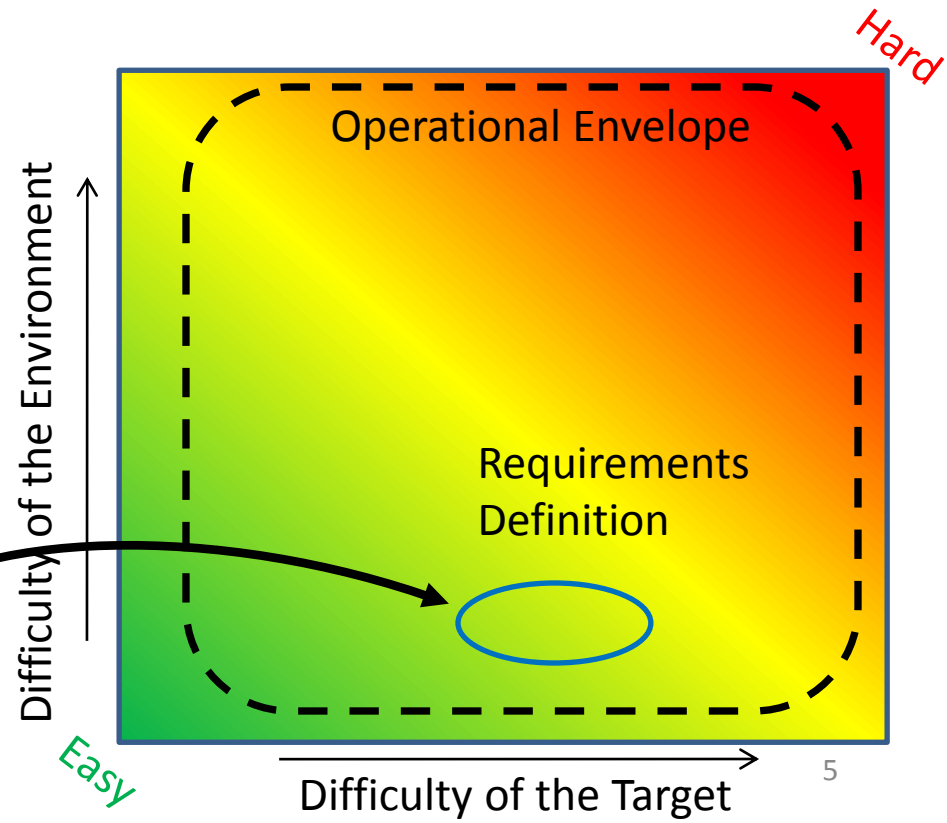
- Interaction between the requirements community and test community can help identify alternatives to hard to test requirements
  - Very high requirements are difficult and costly to test with confidence; must consider cost implications to design and test
- T&E has knowledge of current and legacy system performance; can provide input early in requirements development!
  - Can help identify unrealistic, unaffordable, untestable, and/or not technically feasible requirements
  - Current operational threat environment and what investments will be needed in test resources and infrastructure
  - Implications for size of test based on choice of metrics and operational envelope that the new system is meant to operate in



# Mismatch of Requirements and Evaluation

- Evaluation is often two-fold:
  - Performance measured against specific requirements AND performance assessed across the operational envelope
- Often requirements are narrowly-focused, don't cover the envelope
  - Static in time and do not keep pace with evolving threat
  - Test scope is often limited to the system under test while the system will be operated as a system-of-systems in a joint environment
- Requirements can inform the definition of the operational envelope *by identifying multiple conditions to be considered*
  - OT will focus on examining performance under the conditions identified in the CDD, but also under the conditions operators are expected to use the system (if different)

Tests designed to requirements alone could limit examination of system performance





# The Requirement Rationale or So-What? Factor

- Testers and program managers need to understand the rationale (the *so-what* factor) of the requirement

Probability of Mission Completion	Mean Time Between Failure (MTBF)	Minimum Test Duration
99% (6-hour mission)	~600	~1,800
95% (6-hour mission)	~120	~350
95% (12-hour mission)	~230	~700

What if testing revealed 40 hours MTBF or 86% probability of completing 6-hour mission? Is this good enough?

- Emphasis should be on completing the mission not the mean time between failures



# Requirements Implications to T&E

## Metrics Influence Test Resources

- Metric choice alone can increase test resources by 50% or more.
- Converting to a continuous metric from a binary response metric maximizes test efficiency
- There are several types of quantitative data:

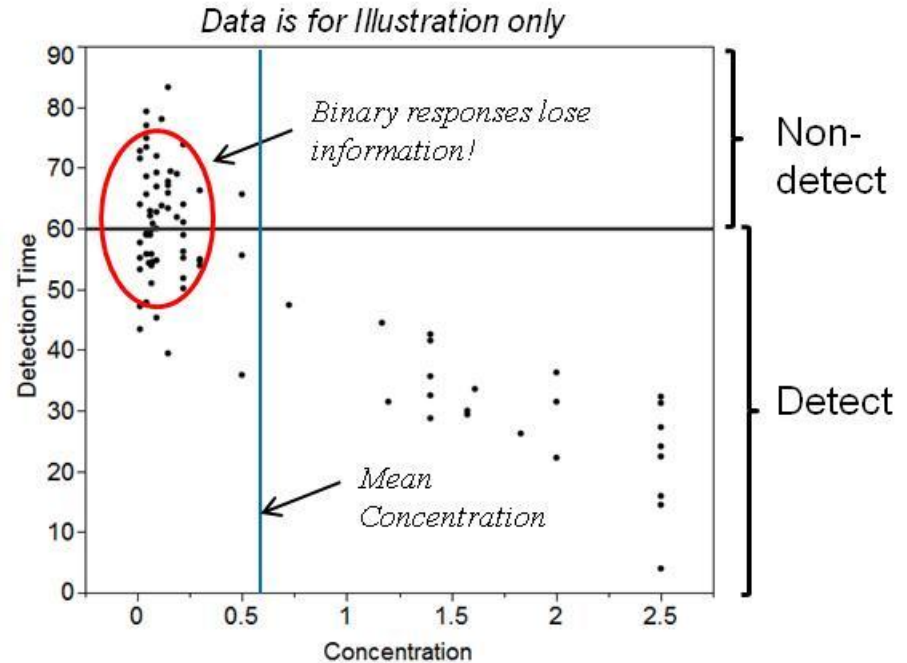
Data Type	Definition	Examples	Information Content
Binary	Data can only assume one of two values.	Pass/Fail, Hit/Miss, Detect/Non-detect	Less Information
Ordinal	Data with discrete values that imply an ordering relationship	Rank Order of preferences on a Scale of 1-5, Order in Races, Letter Grades	More Information
Continuous	Data can take on an infinite number of values	Detection range, Time until event	Most Information

*Increasing Information*  
↓  
*Decreasing Test Size*



# Chemical Agent Detector Results Comparison of Response Variables

- Estimate the probability of detection at 60 seconds at the mean concentration
- Detection times and detect/non-detect information recorded
- Binary analysis results in **300% increase** in confidence interval width



Response	Probability of Detection within 60 seconds at mean	Lower 90% Confidence Bound	Upper 90% Confidence Bound	Confidence Interval Width
Binary (Detect: Yes/No)	83.5%	60.5%	94.4%	33.9%
Continuous (Time)	91.0%	86.3%	94.5%	8.2%





# Conclusions

---

- Weak links between requirements, acquisition, and testing have been noted
  - End-to-End mission-oriented metrics are essential for determining operational effectiveness and suitability
  - Evaluation of mission accomplishment across the operational envelope is necessary
  - The So-What factor helps with trade space and assessment
- Testing community can contribute valuable information to the requirements process
  - Knowledge of current and legacy system performance; can provide input at early requirements development
  - Unrealistic, unaffordable, untestable, and/or not technically feasible
  - Current operational threat environment and what investments will be needed in test resources and infrastructure
- Constructing requirements with test implications understood can result in a significant reduction in required test resources
  - Converting to a continuous metric from a binary response metric maximizes test efficiency