2007-03-14

A Continuum of Testing (presentation)

Jacobs, Patricia A.

A Continuum of Testing

Patricia Jacobs
(pajacobs@nps.edu)
Donald Gaver
(dgaver@nps.edu)
Naval Postgraduate School
Kevin Glazebrook
(K.Glazebrook@lancaster.ac.uk)
University of Lancaster
Ernest Seglie
(Ernest.Seglie@osd.mil)
DOT&E

Pharmaceutical Testing

- Laboratory Testing
- Animal Testing: Effectiveness & Safety
- FDA review: Decision to allow clinical trials
- Clinical trials: Effectiveness & Safety
- Very costly consequences of unexpected serious side effects after distribution, patient use
- Development can be stopped at any time for cause (e.g. VIOXX)

Continuum of Testing of Military Systems: Objective

- Assure Adequate & Timely T&E Funding
- Materiel Solution(s) to fill “Capability Gap”
  - Early Operational Assessments
    - Effectiveness & suitability
    - Operational and technical risks?
- Development
  - Frequent assessment of progress
  - Test in realistic environments as early as possible
- Effectiveness and Suitability of *the entire system* after fielding
  - Lessons learned

Need Material Solution(s) for Capability Gap

- Early operational assessments of proposed solutions (M.&S.)
  - Input to design effort
  - Factors important in design
  - *Measurable* cost-effective improvements to mission capability (Evolutionary Acquisition)
  - Anticipate need for Dynamic Action by Blue to Red Adaptation

Modeling and Simulation for End-to-End System Evaluation: 1

- Size of budget & its allocation
  - Research and Development (R&D)
  - Prototypes; Developmental Testing, Subsystem Integration/DT
  - Developmental & Operational Test and Evaluation
  - Production & Procurement Costs
  - CONOPS
  - Sustainability; logistics
- Early analysis of schedule implications, including design upgrade, technology risk, and testing
  - If time requirement tight to field a block, the system more expensive (and slow) to field: problems found during OT&E, or worse, in the field (engineering design changes)
    - GEN Welch: “rush to failure” for THAAD
    - “If you don’t need it to work, I can ship it now”

Modeling and Simulation for End-to-End System Evaluation: 2

- Adequate developmental and operational T&E funding
- Preview system capability in an operational environment: Model
    - Unexpected vulnerabilities in effectiveness and suitability
    - Plan test design
        - Seed failure modes to be detected
- Preview test design
    - Use experience with similar systems
        - Government needs access to contractor performance data
    - Establish a DoD database containing lessons learned during system development
    - Red Threat
- *M&S Not a substitute for testing*: guidance

System Development and Demonstration

- Demonstrate system performance in its intended environment using competing prototypes
- T&E to assess technical progress for operational utility
- Early operational assessments
    - ID technology risks
    - Provide operational user impacts

DT results generally better than OT results

- “The PM’s rationale is that they will tell me that in past tests they have never had that problem before. Well, you never had that problem before because we, the soldier, use the equipment in the mud, and in the rain and we use it every day by the average soldier.”

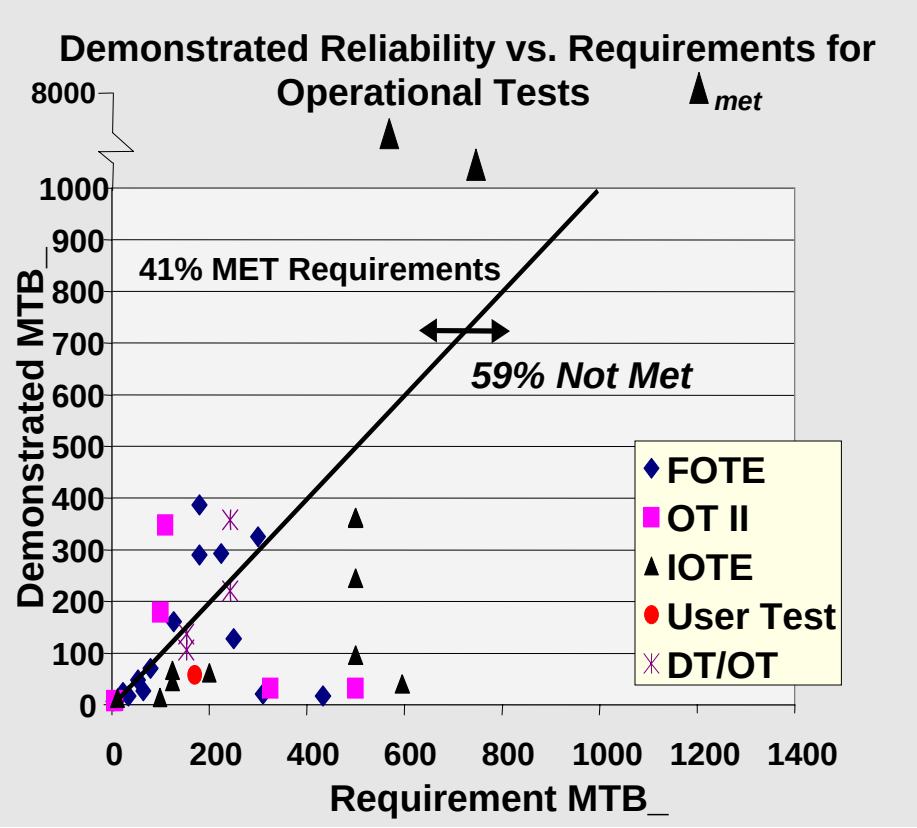
BG Honoré

1st CAV

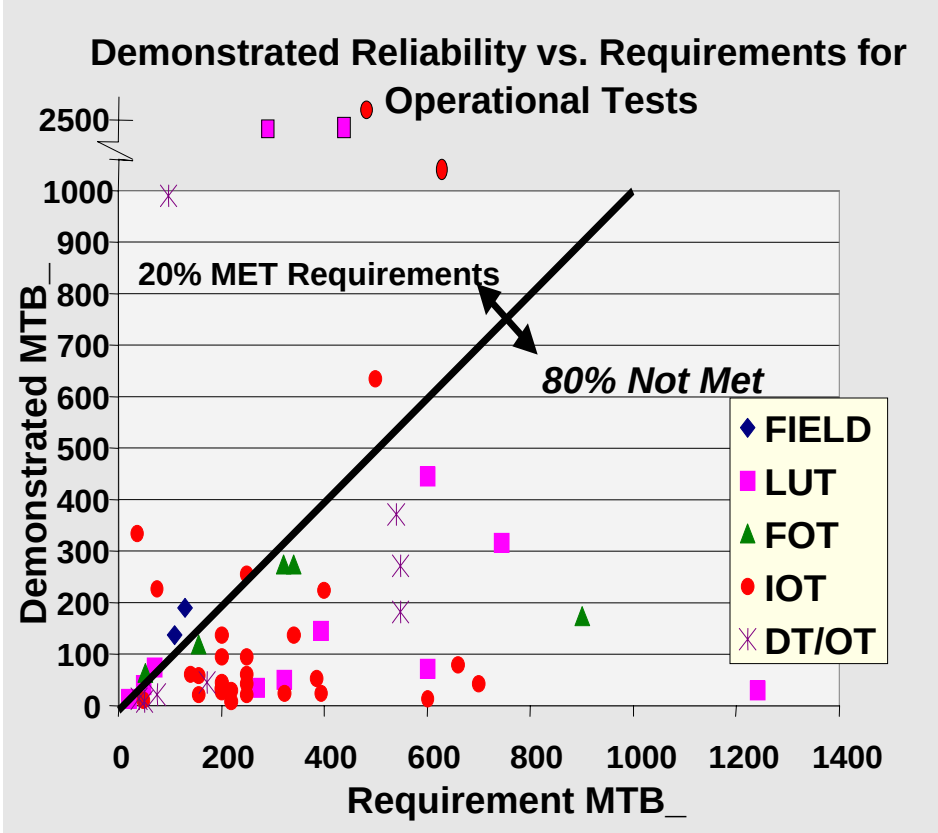
Bring Mature Systems to Operational Test

- Systems fail because they are immature in design or manufacture, not (always!) because the fundamentals are bad

1985-1990



1996-2000



source: ATEC/AEC

Operational Testing

- ***EARLY*** (and Affordably Often) under representative real world conditions
  - Early discovery of problems permits fixes sooner & less costly; e.g. system weighs too much for (transport to) intended use in the field
  - Learn from training and exercise opportunities

Current Model

- Single-use system possibly containing design defects (DDs)
  - Each system used for one field mission
- Tests may discover DDs which are then removed (Reliability Growth)
  - Each remaining DD survives a test independently of previous tests
  - When DD activates during a test, the DD is removed (intention! Not always successful)
  - Test conditions→ Probability of DD test survival

Fixed Budget

- Includes
  - Testing & Removing discovered DDs
    - reliability growth
  - Buying copies for fielding
    - Copies may have DDs remaining
    - Remaining DD activates during field mission→ mission fails
    - Cost of modification
- Tradeoff:
  - Spend less on reliability growth→ can buy more systems, but it is more likely remaining DDs will activate during a mission (mission failure): redesign of the fielded system
  - Spend more on reliability growth→ can buy fewer systems but it is more likely a system will finish a mission without DDs activating

Design Defects (DDs) (Simplified Distinction!)

- Two types of DDs: DD1 & DD2
  - DD2s more difficult to activate during a test than DD1s
- When a DD is discovered during a test, it is removed

Tests

- Type 1 (Early)
  - Less expensive
  - Not as effective at activating DD2s; as effective at activating DD1s
  - Less expensive to remove activated DDs of both types
  - Can choose effectiveness of early tests
    - More expensive early tests are more effective at activating DD2s
- Type 2 (Late)
  - Expensive
  - More effective at activating DD2s (as effective at activating DD1s)
  - More expensive to remove activated DDs of both types

Learning from Testing

- Continue to do Early Tests until there are r1 Early tests in a row during which 0 DDs activated (r1 successful tests in a row)
  - Not learning more
- Then Late Tests until there are r2 Late tests in a row during which 0 DDs are activated
  - Not learning more
- Use remaining budget to buy systems for the field
- Choose r1 and r2 so as to maximize the mean number of fielded systems (missions) for which no remaining DDs activate

Number of Successes in a Row that Maximizes the Expected Number of Field missions in which No DDs activate (maximum run length=5)

Cost per Early Test	Prob. Surv. DD2 in One Early Test	Best Success Run Length: Early Tests	Best Success Run Length: Late Tests (more expensive)	Mean # Field Missions with 0 DDs activating
smaller	0.99 (less effective)	2	4	617
larger	0.85 (more effective)	5	3	726

Remarks

- More effective Early Tests →
  - Longer best run of successes for early tests (despite additional expense)
  - More reliability growth early
  - Fewer Late (more expensive) Tests
  - More field missions completed without remaining DDs activating

Example: Naval Special Warfare Rigid-Hull Inflatable Boat (NSW RIB)

- Close teamwork early testing: OPTEVFOR and combat users
- Competing Vendor Prototypes developed & tested until acceptance trials and source selection
- Early testing operationally realistic: OT&E completed in source selection phase before production contract
- All operational requirements met or exceeded during combined DT/OT
- Early reliability experience for NSW RIB much better than for craft it is replacing

Continuum of Testing

- Ensure adequate funding for T&E
- Provide operational insights throughout the development process
- Realistic testing environments as soon as possible
  - Mature systems to operational test
- Result: Field an effective and suitable system as early as possible and with less cost