Integrated Testing and Independent Evaluation Using Design of Experiments

mickey.quintrall.ctr@osd.mil
WBB Support for OSD
Developmental Test & Evaluation

NDIA T&E Conference

March 3, 2010

Purpose

- **DoD Policy and Guidance Regarding Integrated Testing (IT) and the application of DOE to T&E**

- **How Design of Experiments (DOE) enables Integrated T&E in the Defense Acquisition Process**

Agenda

- **Background**
  - Policy
  - OSD Initiatives
  - DoD T&E Community MOA
- **What is Design of Experiments (Testing)**
  - DoD MOA Implementation
  - MOA and DOE Test Execution
- **DOE Benefits**
  - For The Integrated Test
  - To Testers
- **Summary**

DoD Policy

DoDD 5000.01

"Test and evaluation shall be integrated throughout the defense acquisition process"

DoDI 5000.02

"Integrate, . . . successive periods of DT&E, LFT&E, and IOT&E

"The goal is early identification of technical, operational, & system deficiencies, so that appropriate & timely corrective actions can be developed prior to fielding the system"

DoD Policy

DoDI 5000.02 Enclosure 2

- "T&E shall be conducted in an appropriate **continuum of live, virtual, and constructive system and operational environments**"
- "**Developmental and operational test** activities shall be **integrated and seamless** throughout the phase"
- "Evaluations shall take into account **all available and relevant data and information** from **contractor and Government** sources"

Director, Operational Test & Evaluation Initiatives

- "… DOE provides the scientific and statistical methods needed to rigorously plan and execute tests and evaluate their results. … The DT&E and OT&E offices are working with the OTAs and Developmental Test Centers to **apply DOE across** the whole **development and operational test cycle** for a program"

- "DOE should allow DOT&E to make statements of the confidence levels we have in the results of the testing. Whenever possible, our evaluation of performance must include a rigorous **assessment of the confidence level of the test,** the **power of the test** and some measure of how well the **test spans the operational envelope** of the system"

DOT&E: ***Apply DOE Across Entire Acquisition Development Cycle***

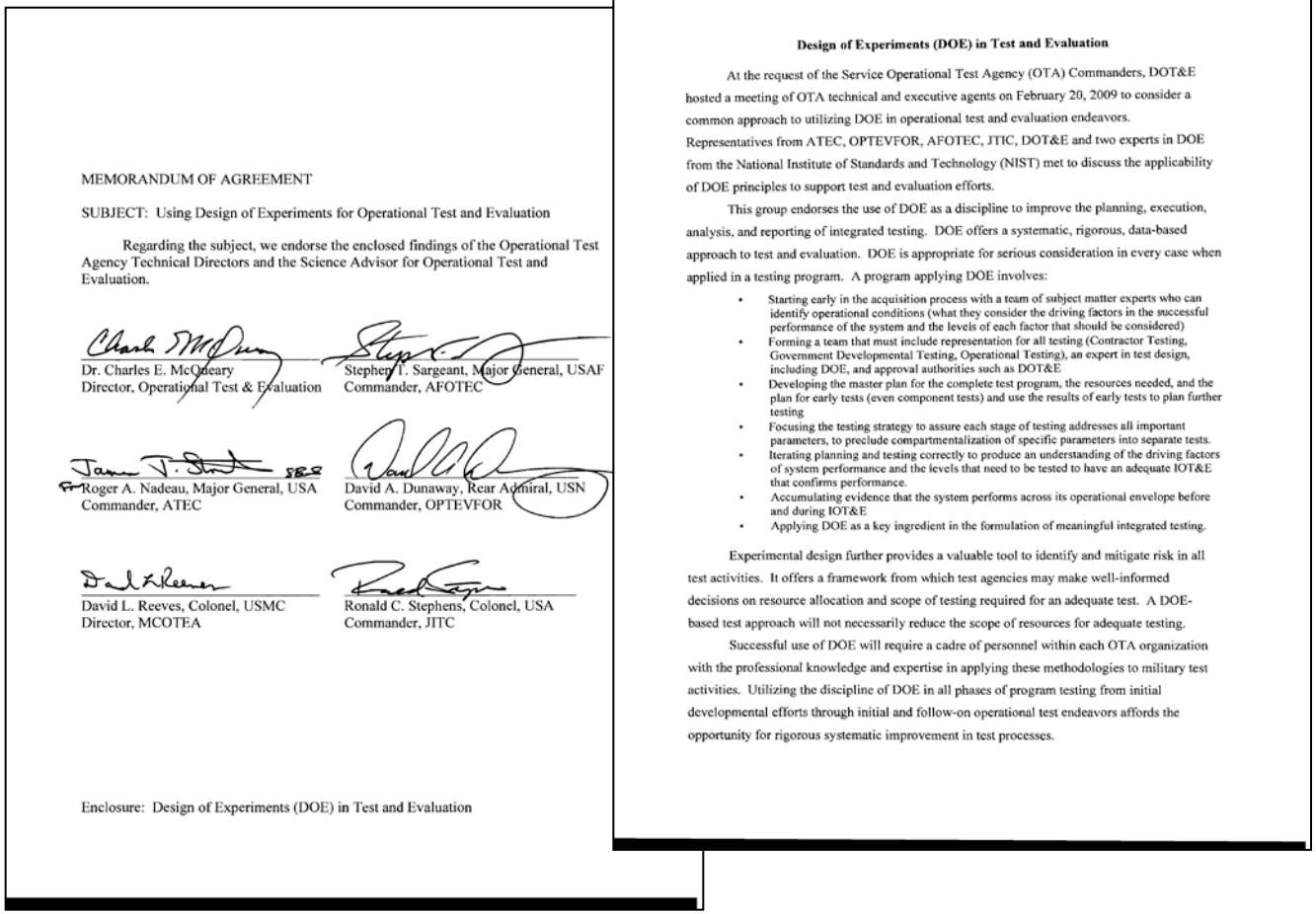
Director, Developmental Test and Evaluation Thoughts

- **"Integrated Testing is important to institute in order to attain test data that can be used across the acquisition processes… Early Planning for Integrated Testing sets up complimentary independent [DT & OT] evaluation"**

- **"DOE is one of many testing methodologies, which work to maximize economies of scale in the T&E effort…DOE considerations are one set structured processes within the T&E tool-bag working for many OSD Integrated Testing efforts"**

DDT&E: ***Integrated Testing and Independent Evaluation Can be Aided by Applying DOE Across Entire Acquisition Development Cycle***

T&E Community MOA, May 1, 2009



MEMORANDUM OF AGREEMENT

SUBJECT: Using Design of Experiments for Operational Test and Evaluation

Regarding the subject, we endorse the enclosed findings of the Operational Test Agency Technical Directors and the Science Advisor for Operational Test and Evaluation.

Dr. Charles E. McQueary
Director, Operational Test & Evaluation

Stephen T. Sargeant, Major General, USAF
Commander, AFOTEC

Roger A. Nadeau, Major General, USA
Commander, ATEC

David A. Dunaway, Rear Admiral, USN
Commander, OPTEVFOR

David L. Reeves, Colonel, USMC
Director, MCOTEA

Ronald C. Stephens, Colonel, USA
Commander, JITC

Enclosure: Design of Experiments (DOE) in Test and Evaluation

Design of Experiments (DOE) in Test and Evaluation

At the request of the Service Operational Test Agency (OTA) Commanders, DOT&E hosted a meeting of OTA technical and executive agents on February 20, 2009 to consider a common approach to utilizing DOE in operational test and evaluation endeavors. Representatives from ATEC, OPTEVFOR, AFOTEC, JTIC, DOT&E and two experts in DOE from the National Institute of Standards and Technology (NIST) met to discuss the applicability of DOE principles to support test and evaluation efforts.

This group endorses the use of DOE as a discipline to improve the planning, execution, analysis, and reporting of integrated testing. DOE offers a systematic, rigorous, data-based approach to test and evaluation. DOE is appropriate for serious consideration in every case when applied in a testing program. A program applying DOE involves:

- Starting early in the acquisition process with a team of subject matter experts who can identify operational conditions (what they consider the driving factors in the successful performance of the system and the levels of each factor that should be considered)
- Forming a team that must include representation for all testing (Contractor Testing, Government Developmental Testing, Operational Testing), an expert in test design, including DOE, and approval authorities such as DOT&E
- Developing the master plan for the complete test program, the resources needed, and the plan for early tests (even component tests) and use the results of early tests to plan further testing
- Focusing the testing strategy to assure each stage of testing addresses all important parameters, to preclude compartmentalization of specific parameters into separate tests.
- Iterating planning and testing correctly to produce an understanding of the driving factors of system performance and the levels that need to be tested to have an adequate IOT&E that confirms performance.
- Accumulating evidence that the system performs across its operational envelope before and during IOT&E
- Applying DOE as a key ingredient in the formulation of meaningful integrated testing.

Experimental design further provides a valuable tool to identify and mitigate risk in all test activities. It offers a framework from which test agencies may make well-informed decisions on resource allocation and scope of testing required for an adequate test. A DOE-based test approach will not necessarily reduce the scope of resources for adequate testing.

Successful use of DOE will require a cadre of personnel within each OTA organization with the professional knowledge and expertise in applying these methodologies to military test activities. Utilizing the discipline of DOE in all phases of program testing from initial developmental efforts through initial and follow-on operational test endeavors affords the opportunity for rigorous systematic improvement in test processes.

"This group endorses the use of DOE as a discipline to improve the planning, execution, analysis, and reporting of integrated testing"

MOA & DOE Test Execution Process

1. Start Early

2.Form Team

3.Develop Master Plan

- Define the Problem

4. Focus Strategy to Ensure all Parameters are Covered

- Determine the Dependent Variables
- Determine the Independent Variables
- Determine the Levels of Indep. Variables

5. Iterate Planning and Testing

- Determine the possible combinations
- Determine the number of observations
- Redesign for Optimum Execution
- Randomize to Eliminate Errors

6. Accumulate Evidence Across Operational Envelope

- Comply with Ethical & Legal Requirements
- Develop a Model to Validate Experiment
- Collect Data
- Reduce Data
- Verify Data to Eliminate Anomalous Results
- Analyze Results
- Report Findings

7. Apply DoE To Integrated Testing

What is Design of Experiments?

- **The goal of DOE is to: Learn about process factors and their interaction with each other, so that we can accurately predict the outcome of the process.**
  - Responses—What are the Desired/Expected Outcomes?
  - Factors—What Measures are Important?
  - Levels—What are the Possible Ranges/Extents for the Factors?

The wise investigator expends his effort not in one grand design (necessarily conceived at a time when he knows least about unfolding reality), but in a series of smaller designs, analyzing, modifying, and getting new ideas as he goes.

— *G. E. P. Box*

What is Design of Experiments?

- **Allocates Testability for All Requirements to a Test Sequence**
  - Contractor Test, DT, OT
  - Component—Subsystem—System
  - Brass-board—Prototype—Preproduction—Production
- **Determines Optimum Test Runs & Test Points**
  - Based on Factors, Levels & Interactions
  - Utilizes Statistical Tools
- **Iterative Process**
  - Based on Test History
  - Utilization of Test Resources
  - Modeling and Simulation to Support Predictions

"I contend that all experiments are designed. Some are designed by intuition and gut feel. … Other experiments … according to a rigorous statistical protocol …. In either case, experiments are designed." *— Gregory Alexander*

DOE Benefits in an Integrated Test Environment

- **Everyone Understands the Test Problem, the Test Environment and How the System is Tested**
- **Statistical Tools Identify Optimum Factors, Test Points and Conditions to be Tested**
- **Performances being Assessed are Allocated to Specific Tests in Sequence**
- **Allows Comprehensive Body of Data to be Accumulated to Support Findings**
- **Facilitates Coordination of Test Events**

DT Results Can Be Used to Support OT Findings and Help Scope OT Events

DOE Benefits to Testers

- Allows Independent Evaluation
- Efficient Utilization of Test Resources
- Ability to Project Test Resource Requirements
- More and Better Data to Support Analysis and Findings
- Potential for Schedule Acceleration
- Savings may Accrue from Shortened Schedule and Avoided Rework—not Necessarily from Reduced Testing

Higher Confidence in Test Results

Summary

Background

- Policy
- OSD Initiatives
- DoD T&E Community MOA

What is Design of Experiments (Testing)

- DoD MOA Implementation
- MOA and DOE Test Execution

DOE Benefits

- For Integrated Test
- To Testers

Current OSD Project: Generate OSD T&E guidance, which explains acceptable and possible use of designed experimentation (DOE).

DOE Resources for Testers

- **USAF DoE Community of Practice**
  - Web-ex Mondays 1400 CT
  - Contact: https://connect.dco.dod.mil/eglindoe
    Gregory T. Hutto: Gregory.Hutto@Eglin.af.mil
- ***Design and Analysis Of Experiments*, 6th Ed., 2004**
  - Douglas C. Montgomery, ISBN 0-471-15746-5
- ***Design of Experiments*, 2nd Ed., 1957**
  - Cochran and Cox, Wiley and Sons
- ***Response Surface Methodology*, Process and Product Optimization Using Designed Experiments, 3rd Ed., 2009**
  - Raymond H. Myers and Douglas C. Montgomery
- ***Joint Test and Evaluation Program Handbook***
  - DOT&E, December 2008
- ***Efficient Simulation Using DOE Methods***
  - Dr. Tom Donnelly, SAS Institute: Tom.Donnelly@jmp.com
- **Sample Size, Confidence and Designed Experiments**
  - Dr. Mark Kiemele, President, Air Academy Associates: aaa@airacad.com

OSD T&E - DOE Points of Contact

DOT&E: Dr E. Seglie
ernest.seglie@osd.mil

DT&E: Mr M. Quintrall
mickey.quintrall.ctr@osd.mil

Back-Up

Lessons Learned & Best Practices

1. Start Early

- Tester Participates in Requirements Process
  - Focus on Mission Accomplishment
  - Ensure Testability
- Flow Mission-based Test Design Down to DoE Test Design

Discussion:

- JCIDs “Requirements” are Often “System Attributes” & not Mission-focused, Operational Requirements
- Permits Team-designed Mission-based Analysis Structure.
- Early T&E Involvement Could Produce Better JCIDs Documents

Lessons Learned & Best Practices

2. Form team

- Representatives Must Include User, PM, DT, OTA, DOT&E
- Ensure Team Accepts DoE Methodology
- Provide Consistent Oversight/Guidance

Discussion:

- Involves Commitment to Test Design
- Eliminate Excursions Not Indicated by Prior Test Results

Lessons Learned & Best Practices

3. Develop Master Plan for Complete Test Program

- Evolve Test Plan from Evaluation Plan
- Single Test Data Management Scheme
  - Supports All Events (Data Source Matrix)
  - Documents Testing of All Requirements
- Iterative Through Design & Development Process
  - Test & Evaluation Strategy (TES), to
  - Test and Evaluation Master Plan (TEMP)

Discussion:

- Possible conflicts—Plan Adjustments After Approval
  - Adjust Test Objectives Based on Prior Test Results
  - Allow Changes to Previously-approved Resource Requirements
- Requires “Flexibility” on Part of Testers and PM

Lessons Learned & Best Practices

4. Focus Testing Strategy

- Leverage “Screening” to Determine Scope of Individual Events
- Use Data from Completed DT to Support OT
- Execute All Tests to Get Needed Data

Discussion:

- Discounts Factors with no Impact on Mission Performance
- Uses DT Findings for Factors not Significantly impacted by OT Environment (operator, field conditions)
- OT Entry will Depend on Completion of DT

Lessons Learned & Best Practices

5. Iterate Planning & Testing

- Apply Lessons Learned
- Adjust Test Plans Based on Earlier Tests
- “Plug-in” to DT Results Early
- Dynamic Planning versus Static Plans

Discussion:

- Eliminates Unimportant Factors Based on Early Test Results, Focus on High-risk Elements
- Defines Test Baselines for System Upgrades
- Uses Data to “Fill-in” DT Design

Lessons Learned & Best Practices

6. Accumulate Evidence Across Operational Envelope

- Treat each test as building on previous tests
- Use Master Data Source Matrix.
- Evaluate Effectiveness & Suitability Based on All Previous Data.
- Use Knowledge to Identify Unknown and High Risk Items.
- Use DT events to explore operational conditions, include system operators
- Examine TEMP Timeline, Process & Approvals

Discussion:

- Applies to Component & System Testing
- Encourages DT to Test to Operational Conditions

Lessons Learned & Best Practices

7. Apply DoE when Formulating Integrated Testing

- "Successful use of DoE will require a cadre of personnel…"

Discussion:

- Will Require Comprehensive Workforce Education & Training
  - Selected Workforce Education & Training Resources are Currently Available
- Workforce must Focus on Presentation of Plans, and Test Results, to Assure Understanding by Decision-makers.

DOE Test/Experiment Process

1. Define the Problem
2. Determine the Dependent Variables
3. Determine the Independent Variables
4. Determine the Number of Levels of Independent Variables
5. Determine the possible combinations
6. Determine the number of observations
7. Redesign for Optimum Execution
8. Randomize to Eliminate Errors
9. Comply with Ethical and Legal Requirements
10. Develop a Mathematical Model to Validate Experiment
11. Collect Data
12. Reduce Data
13. Verify Data to Eliminate Anomalous Results
14. Analyze Results
15. Report Findings

DOE in U.S. DoD T&E—*Army*

- **ATEC**
  - DoE used for planning system evaluations and individual data-collection events
  - Single table depicts how the individual test events will manage each factor
  - Be able to reconfigure for unforeseen events
  - Manage tradeoffs between operational realism and sufficient data
  - Requires detailed front-end planning

Four Requirements To Design Rigorous Warfighting Experiments

Internal Validity

1. **Capability Used**
2. **Detection of Change in Effect**
3. **Isolation of Reason for Change**

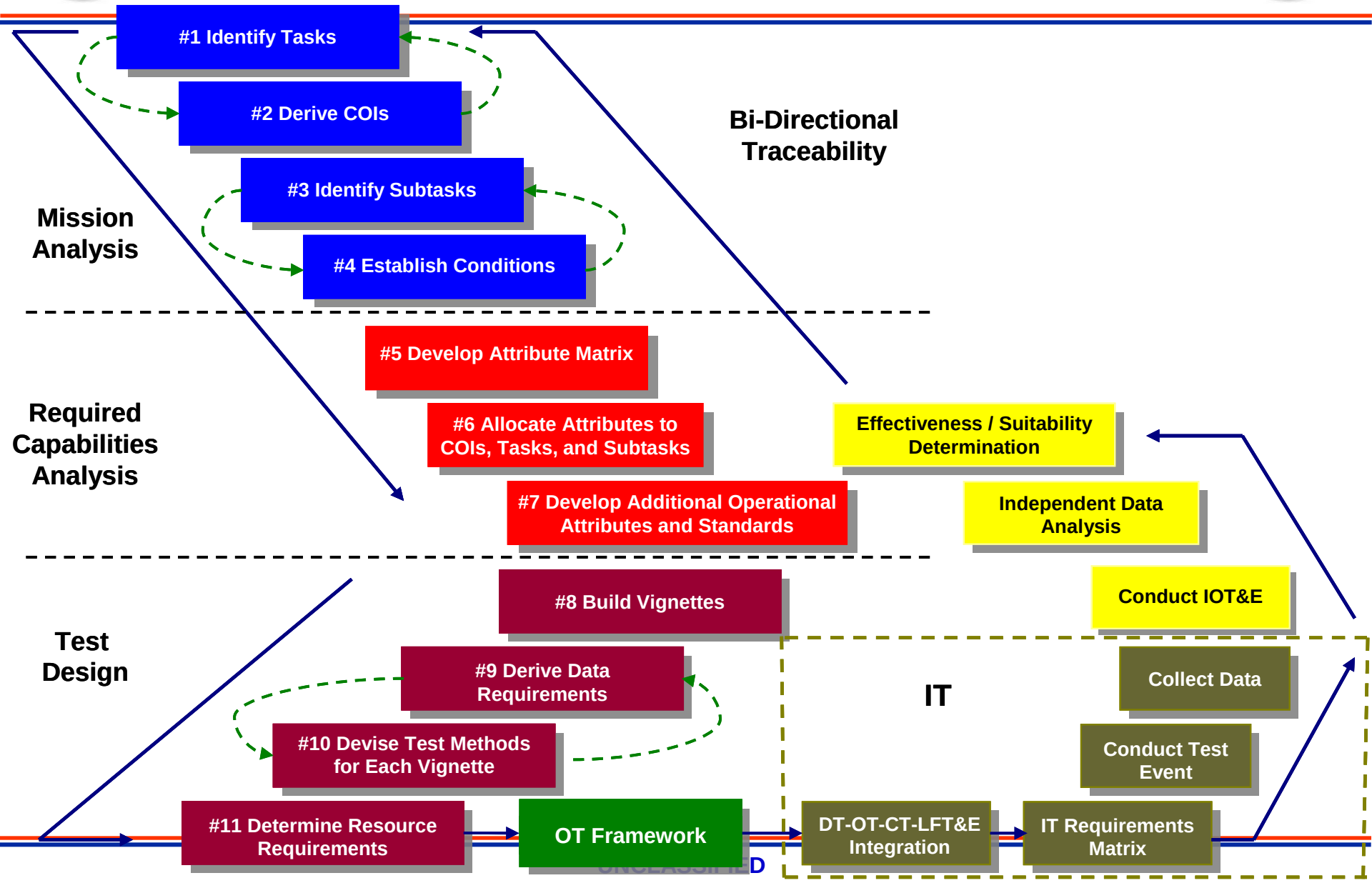
External Validity

4. **Relating Results to Military Operations**

DOE in U.S. DoD T&E—*Navy*

- **COMOPTEVFOR**
  - DOE part of Mission-based Test Design (MBTD)
  - A shift functional-based to mission-based OT.
  - OT team provides detailed OT input earlier in program schedule.
  - OT designed around factorial design
  - Sharing of T&E responsibility, resources, and data throughout system development.
  - IOT&E as mission capability confirmation.

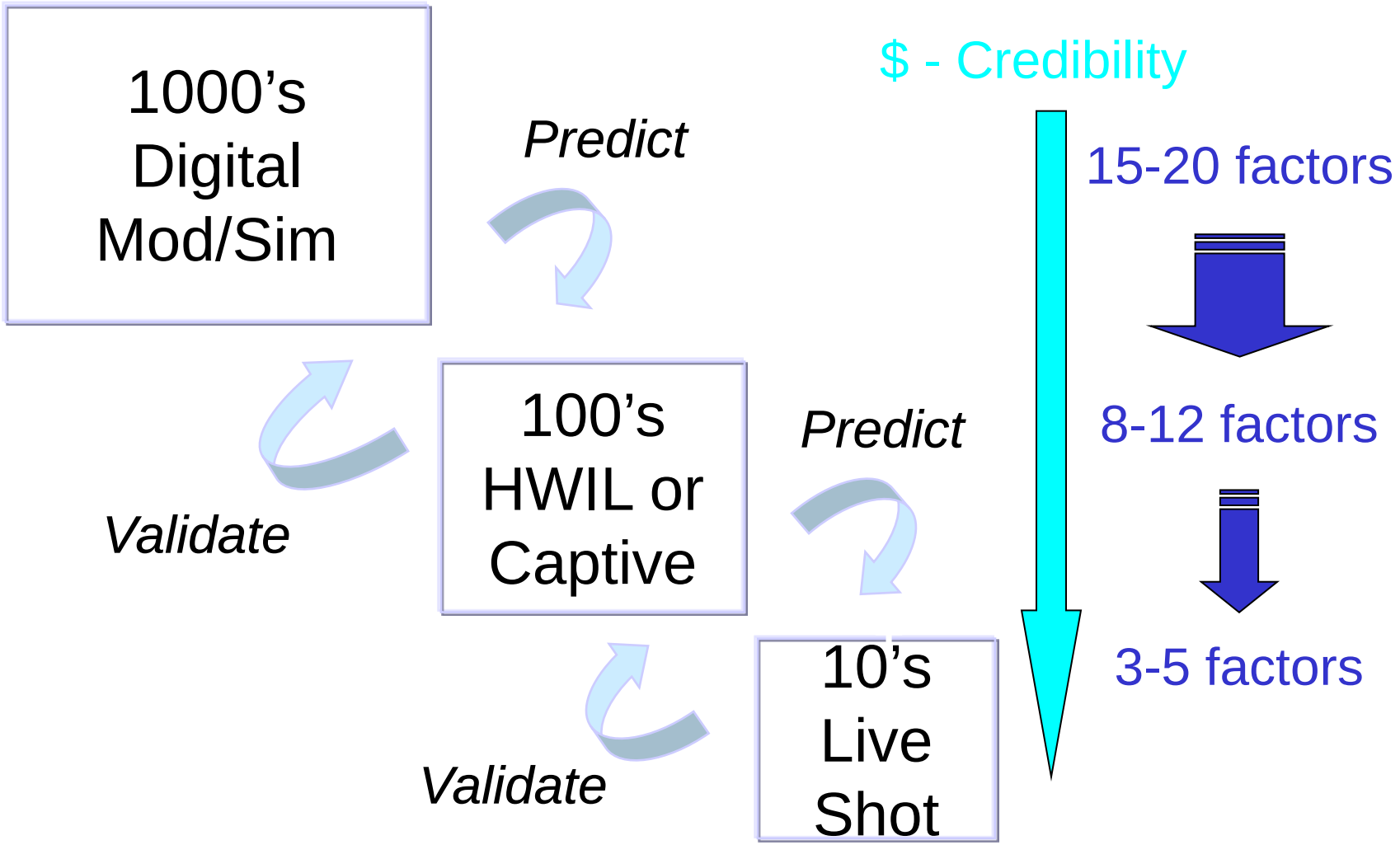
Navy Mission-based Test Design



DOE in U.S. DoD T&E—*Air Force*

- **53RD Test Wing**
  - With digital simulations, screen 15-20 variables with fractional factorials and predict performance
  - In HWIL, confirm digital prediction (validate model) and further screen 8-12 factors; predict
  - In live fly, confirm prediction (validate) & test 3-5 most vital variables
  - Prediction Discrepancies provide opportunity to improve simulations

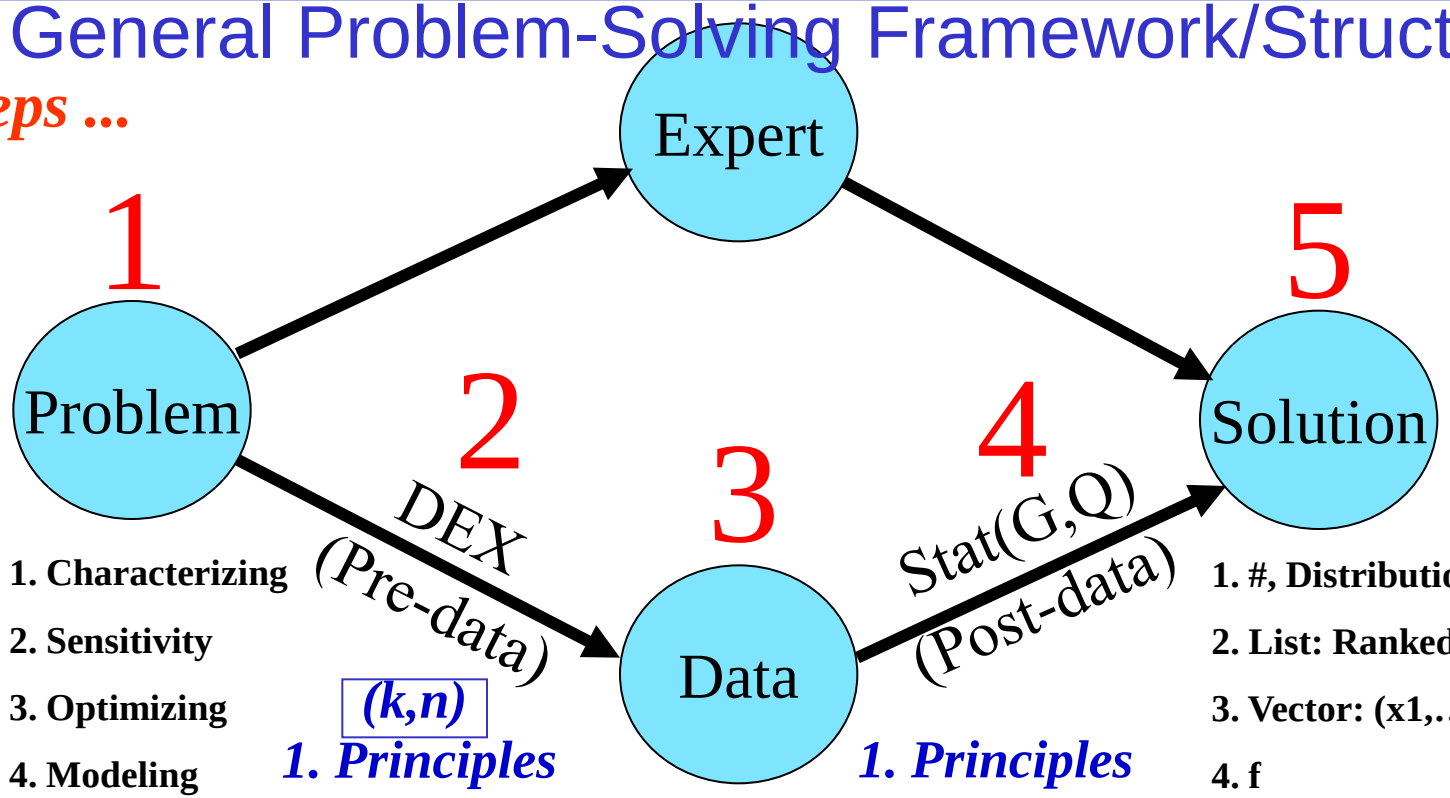
Air Force DOE Model



Other DOE in U.S.—*NIST*

General Problem-Solving Framework/Structure

5 Steps ...



1. Characterizing
2. Sensitivity
3. Optimizing
4. Modeling
5. Comparing
6. Predicting
7. Uncertainty
8. Verifying
9. Validating

(k,n)

1. Principles

2. Techniques

- 1_FAT
- Monte Carlo
- Latin HC
- Orthogonal
- Fractional
- Resp Surface

1. Principles

2. Techniques

1. Estimation

2. Testing

1. Graphical

2. Quantitative

1. #, Distribution
2. List: Ranked Factors
3. Vector: (x1,…,xk)
4. f
5. Y/N
6 #
7. SD(#)
8. Y/N, Vector: (x1, …,xk)
9. Y/N, Vector: (x1, …,xk)

Problem Classification

Comparative

Focus: 1 primary factor

Q1. Does that factor have an effect (Y/N)?

Q2. If yes, then best setting for that that factor = ? (vector)

Constraint: Want conclusions to be robust over all other factors

Designs: CRD, RBD, LSqD,TPD

Screening/Sensitivity

Focus: all factors

Q1. Most important factors (ranked list)

Q2. Best settings (vector)

Q3. Good model (function)

Designs: 2^kD, 2^{k-p}D, TD

Regression

Focus: all factors

Q1. Good model (function)

Continuous factors

Designs: BBD,XOD

Optimization

Focus: all factors

Q1. Best setting

s (vector)

Continuous factors

Designs: RSD, CD, BBD

Acceptance

Focus: all population points

=> all t-tuples of settings

Q1. Accept the product/system as safe?

Q2. Points ➔ failure

Q3. t-tuples of settings ➔ failure

Q4. Factors affecting safety?

Designs: 2^{k-p}D,CD

Many real-world problems should be done in 2 stages:

1. exploratory (= sensitivity analysis)

2. ultimate objective