Design of Experiments – Guidance

General

Design of Experiments (DOE) is a statistical methodology for planning, conducting, and analyzing a test. Any program that applies DOE principles should begin early in the test planning process. The test planners should assemble a group of subject matter experts who can identify the primary evaluation metrics (in DOE parlance: response variables) of interest that will characterize the performance of the system in the context of a mission-oriented evaluation. The test planners should identify environmental and operational factors that are expected to drive the performance of the system, as well as the levels of these factors (i.e., the various conditions or settings that the factors can take). A master test strategy should include the resources needed, the concept for early tests (including component tests), and the use of the results of early tests to plan further testing. One goal of the test strategy should be to ensure adequate coverage of all important factors while demonstrating the evaluation metrics (response variables) through planned testing. The testing strategy should be iterative in nature to ensure an adequate Initial Operational Test and Evaluation (IOT&E). The testing strategy should accumulate evidence that the system performs across its operational envelope before and during IOT&E. The test planners should apply DOE at each test iteration.

Elements of DOE for the TEMP

A brief overview of the design philosophy should be outlined in Section 3.2 of the TEMP. The information content may vary depending on the Milestone that the TEMP is supporting. Table 1 outlines information content that is appropriate for each milestone. Systems with legacy data will be expected to include more detail and have more robust test designs. The details of each of the test designs should be provided in a supporting appendix to the TEMP. Elements of experimental design should include the following:

- The goal of the test (experiment). See Mission Focused Evaluation Guidance.
- Quantitative mission-oriented response variables (evaluation metrics) for effectiveness, suitability, and survivability. See Mission Focused Metrics Guidance.
- Factors that affect those measures of effectiveness, suitability, and survivability. See Integrated Survivability Evaluation Guidance.
- A method for strategically varying factors across developmental, operational, and live fire testing with respect to responses of interest. See Integrated Testing Guidance.

Statistical measures of merit (power and confidence) on the relevant response variables (evaluation metrics) (i.e., those for which doing so makes sense). These statistical measures are important to understand "how much testing is enough," and can be evaluated by decision makers on a quantitative basis so they can trade off test resources for desired confidence in results.

These elements include all of the planning steps for designing an experiment, with the exception of execution order. Standard statistical designs assume the test point execution order can be randomized. This is often not the case in T&E, since many factors cannot be easily

controlled or changed (e.g., weather, test range location). Therefore, designs including blocking and/or split-plot techniques should be considered. The execution of the test, including run plans/order, should be discussed in the Test Plan.

Commonly, the system under test (SUT) is a complex system with multiple missions and functionalities. The test design should reflect the complexity of the system. Often, multiple test designs will be necessary to fully characterize SUT mission performance. This might also require multiple experimental designs to capture all stages or aspects of mission execution.

Table 1: DOE Information Content for the TEMP

	Information Content
Milestone A	Identify responsibilities of T&E WIPT for test design purposes The goal(s) to be addressed at each stage of testing Metrics for each goal/question Initial listing of factors Language for the overall testing strategy, including: Screening experiments to ensure important factors are considered in operational testing Sequential experimentation
Milestone B	Identify responsibilities of T&E WIPT for test design purposes The goal(s) to be addressed at each stage of testing Metrics for each goal/question Refined listing of factors and levels Test designs to support resourcing for limited user tests (LUT) and operational assessments (OA) • While test designs for the IOT&E are not required, the TEMP should identify key resources for the IOT&E including test assets that require long lead times to acquire. Language for the overall testing strategy, including: • Screening experiments to ensure important factors are considered in operational testing • Sequential experimentation

Design of Experiments – Guidance

Milestone C	Identify responsibilities of T&E WIPT for test design purposes The goal(s) to be addressed at each stage of testing, focusing on IOT&E Metrics for each goal/question Refined listing of factors and levels, based on prior testing and the operational mission. Details on how the factors and levels will be varied and controlled during each stage of testing Complete test designs to support resourcing for IOT&E Language for the overall testing strategy, including: How previous knowledge is being used to inform IOT&E test planning. Analysis plans to support power calculations

References

Guidance on the use of Design of Experiments (DOE) in Operational Test and Evaluation, DOT&E, October 19, 2010

Montgomery, D. C. (2009), *Design and Analysis of Experiments*, John Wiley and Sons

Myers, R. H., and Montgomery, D. C. (2002), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, John Wiley and Sons.

TEMP Body Examples

Precision Guided Weapon Example Appendix

Artillery Example Appendix

Software Example Body and Appendix