Scientific Test and Analysis Techniques – Guidance

Background

The authors of the TEMP should employ scientific test and analysis techniques to develop a defensible analytical basis for the size and scope of the T&E program. Recent guidance in this regard has focused on Design of Experiments (DOE) for sizing test events, but there are other tools available in the scientific test planning toolbox. This guidance provides an overview of the scientific test and analysis techniques (STAT) that may be used in planning, conducting, and analyzing a test or experiment. Additionally, this guidance summarizes key content that should be summarized in the TEMP.

Guidance

Any program that applies scientific test principles should commence doing so early in the test planning process. The program should assemble a test and evaluation working integrated product team (T&E WIPT) of subject matter experts who can identify the primary quantitative mission-focused measures (in DOE terminology: response variables) of interest that will characterize the performance of the system in the context of a mission-focused evaluation. The T&E WIPT 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).

With these measures and factors in mind, the T&E WIPT should populate the Developmental and Operational Evaluation Frameworks and decide which of the following analysis techniques will best ensure adequate coverage of all important factors while evaluating the quantitative mission-focused measures through planned testing. The testing strategy should be iterative in nature throughout integrated test and evaluation 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.

Elements of Scientific Test Design for the TEMP

A brief overview of the test design philosophy should be outlined in Section 3 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. Additionally, if previous test data will be used to augment operational testing, the methodology for using that data should be discussed. See the Bayesian Example for an example of how previous test data can be used to scope future testing. The details of each of the test designs should be provided in a supporting STAT appendix to the TEMP. Elements of any test design (regardless of statistical methodology) should include the following:

- The goal of the test (experiment).
- Quantitative mission-focused measures (also known as quantitative mission-oriented response variables) for effectiveness, suitability, and survivability. See Quantitative Mission-Focused Measures 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) and corresponding effect sizes on the relevant quantitative mission-focused measures (response variables) (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.

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: Test Design Information Content for the TEMP

Milestone Supported	Information Content
A	Responsibilities of T&E WIPT for test design purposes The goal(s) to be addressed at each stage of testing Quantitative mission-focused measures for each goal/question An initial listing of factors for each quantitative mission-focused measure Language for the overall testing strategy, including: • Screening experiments to ensure important factors are considered in operational testing • Sequential experimentation Test designs to support resourcing for limited user tests (LUT), operational assessments (OA), and IOT&Es • While test designs may be very preliminary at MS A, it is essential that scientific test designs be used as a basis to estimate resources for a Request for Proposal near MS A. Therefore, special attention should be paid towards making sure adequate resources are allocated for long lead items (e.g., number of targets, weapons, specialize range capabilities, etc.).

B	Identify responsibilities of T&E WIPT for test design purposes The goal(s) to be addressed at each stage of testing Quantitative mission-focused measures for each goal/question Refined listing of factors and levels for each quantitative mission-focused measure Test designs to support resourcing for limited user tests (LUT), operational assessments (OA), and IOT&E • Test Designs should be updated from MS A to account for any new information. Language for the overall testing strategy, including: • Screening experiments to ensure important factors are considered in operational testing • Sequential experimentation
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 Quantitative mission-focused measure for each goal/question Refined listing of factors and levels, based on prior testing and the operational mission, for each quantitative mission-focused measure 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

Scientific Test and Analysis Tools

There are many different scientific and statistical design and analysis tools that are appropriate to use in T&E. The most common tools and methods for T&E include:

- Design of Experiments (see DOE Guidance)
  a. DOE TEMP Body Example
  b. DOE Appendix D Artillery Example
  c. DOE Appendix D Precision Guided Weapon Example
  d. DOE Appendix D Software Intensive System Example
- Observational Analyses (see Observational Analysis Example)
- Survey Design and Analysis
- Reliability Test Planning
- Hypothesis Testing
- Statistical Methods for Validation of Modeling and Simulation
- Bayesian Analysis Methodologies (see Bayesian Guidance and Bayesian Example)

The STAT Appendix is an appropriate location to discuss any of these tools if they were

used to determine the required testing outlined in the TEMP.

Design of Experiments (DOE) requires that the tester has control over at least the important factors when executing a test. There are many types of experimental designs that allow for different test constraints. A list of common test designs and their applicability to operational testing is here. In addition to the content of Table 1, the specific test design strategy should be discussed if DOE is being used to plan the operational test. However, it is not always possible for a tester to control all test conditions. In these cases we can still analyze the adequacy of the test using observational analyses. It is still essential that the TEMP outline the key information in Table 1 as well as a minimum acceptable test size. If historical data is available, it should be used to approximate the test scope.

Surveys of operators and maintainers are essential aspects of operational testing. See Survey Design and Administration Guidance to read about what survey content belongs in the TEMP. The detailed surveys, administration plan, and other related information should be included in the Test Plan.

Simple hypothesis tests are rarely appropriate for scoping the operational test; however they do provide a methodology for assessing if the reliability testing is adequate. See the reliability test guidance for additional guidance.

Finally, Bayesian methodologies may be appropriate in cases where there is a lot of additional test data (operationally realistic developmental testing, operational assessments, etc.) that will be incorporated into the operational evaluation. In these cases, the TEMP should also discuss what information will be carried forward, the analysis methodology for doing so, and what testing must still be conducted in OT.

Guidance

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

Best Practices for Assessing the Statistical Adequacy of Experimental Designs Used in Operational Test and Evaluation, July 23, 2013

Guidance on the Use and Design of Surveys in OT&E, June 23, 2014

Discussion on the Use and Design of Surveys, 24 February, 2015

Discussion on Including Neutral Responses on Survey Questions, April 2, 2015

Survey Pre-Testing and Administration in Operational Test and Evaluation, 6 January 2017

Guidance on the Validation of Models and Simulation used in Operational Test and Live Fire Assessments, March 14, 2016

Clarifications on Guidance on the Validation of Models and Simulations used in Operational Test and Live Fire Assessments, 17 January 2017

References

Scientific Test and Analysis Techniques (STAT) in Test and Evaluation (T&E) Implementation Plan, DASD(DT&E), January 2012.

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.