Quantitative Mission-focused Measures – Guidance

General Guidance

TEMPs should include quantitative mission-focused measures (also referred to as quantitative mission-oriented response variables and can be thought of mathematically as dependent variables) for effectiveness, suitability, and survivability. These measures are key to good test designs; poorly-chosen or poorly-defined measures, even if they are directly connected to Key Performance Parameters (KPPs) or Key System Attributes (KSAs), could result in a poorly designed test, and can lead to test results that are not relevant to the mission effectiveness, suitability, or survivability of the system.

Choosing Quantitative Mission-focused Measures

The selection of quantitative mission-focused measures is a critical part of the test design effort, and should occur as test planning begins. Step 1 is to identify the goal of the test. This should reflect evaluation of end-to-end mission effectiveness, suitability, or survivability in an operationally realistic environment. Once the test goal is known, testers should select appropriate measures of system performance and provide data for addressing the goal of the test. Ideally, the measures will be quantitative, mission oriented, relevant, informative, and not rigidly adhere to the narrowest possible interpretation of definitions in requirements documents. Measures should provide a criterion of mission accomplishment (not technical performance for a single subsystem), lend themselves to good test design (i.e. be continuous in nature), and in general comprehensively cover the reasons for procuring the system.

Although many measures can be used to characterize system performance in a given mission, it is desirable that a small number of quantitative mission-focused measures be identified to be the focus of the evaluation of operational effectiveness, suitability, or survivability and used in concert with statistical test design methodologies. Additional secondary measures are encouraged, and are necessary to characterize other aspects of system performance. For example, for test design, the hit success rate may be identified as the quantitative mission-focused measure, even though other measures are needed to characterize success in the dependent portions of the kill chain (e.g., detection, identification, time to engage, engagement range).

Exceptions to using CDD/CPD-defined Measures

The quantitative mission-focused measure identified for test design need not be the KPPs. Often KPPs are insufficient for measuring the mission effectiveness, suitability, and survivability of the system. See the Inspector General report dated May 15, 2015 for two examples. If the requirements cannot be revised to define those system characteristics most critical for providing an effective military capability, the TEMP must identify and define those characteristics. Examples of quantitative mission-focused measures that enable mission-focused test design include detection/classification range, miss distance, probability of hit, search rate, time to accomplish a successful mission, counter-detection range, and probability of successful intercept.

Quantitative Mission-focused Measures – Guidance

When testers select these quantitative mission-focused measures, the resultant test design should ensure that adequate data will be collected to accomplish several goals:

- Provide adequate data to evaluate the effective military capability of the system
- Provide a meaningful measure of system performance across the operational envelope
- Provide sufficient data for the secondary measures needed to characterize system performance.

Types of Quantitative Mission-focused Measures

Quantitative mission-focused measures can be continuous or discrete, but continuous measures are always preferable. A continuous measure will almost always require a smaller sample size and fewer test resources for the risk levels chosen (confidence and power). Additionally, continuous measures often contain more information regarding the performance of the system, whereas a corresponding discrete measure will throw away information. For example, measuring detect/not detect provides no information about how close the sensor approached. Using the range at which detection occurred in concert with the closest point of approach in cases where no detection occurred provides a better characterization of sensor performance. The probability of detection over all ranges is the only quantity that can be calculated with the discrete data, but if the continuous variable (range) is measured, one can understand the distribution of detection ranges as well as the probability of detection as a function of range. Even if the requirements document defines a probability-based metric, great effort should be expended to find a related continuous measure on which to base the test design.

Examples of continuous measures include time to detect, miss distance, human error rate, time to complete task, and range of engagement. Examples of discrete measures include hit/miss, message complete/not complete, and detect/not detect.

Definitions of Quantitative Mission-focused Measures

The measures chosen must also be well-defined and meaningful. Testers and evaluators should consider example operational scenarios to ensure that the measure can be unambiguously measured (scored) and calculated in all cases. The following principles are critical:

- Formulas for the measures should not be ambiguous – TEMPs should provide amplifying information (explicit formulas and/or scoring criteria) if the CDD requirement is unclear
- Measures should be testable and not require unsafe or unexecutable test constructs or cost-prohibitive instrumentation
- Measures should accurately represent the desired performance of the system – Good scores should correspond to desired operational performance
- Measures should not lead to non-production representative modifications to the system or unrealistic tactics

Measure Selection for Survey Data

In operationally focused testing, the use of operator surveys and subject matter expert panels are needed and useful to aid in the characterization of system effectiveness, suitability, and survivability. This is particularly true when quantitative data is scarce due to expensive field testing or low sample sizes. However, before using them, other objective measures such as time to complete a task or human error rate should be explored.

Survey data, like physical measures, should be collected systematically for each test condition to facilitate statistical comparisons of survey responses across the operational test space. Imagine, for example, that one is interested in evaluating the performance of an aircraft under different light conditions, day versus night. Administering the same survey to pilots under both test conditions allows testers to determine how pilots' experiences shift under different operational conditions and if those shifts in experience explain observed differences in aircraft performance. Additionally, many important aspects of operational suitability are best addressed by survey data (e.g., human machine interface, operator workload). Ideally, survey data and subject matter expert panels should be used in concert with objective quantitative data.

Survey use should follow best practices, such as:

- Clearly identify survey objectives: TEMP should indicate which test goals will be addressed by survey data, the goal of the survey measure, and who will provide the survey data required to address the measure, for instance, operators or maintainers.
- Surveys should be pre-tested on an appropriate group to reveal if questions are confusing or if information is missing. The TEMP should include a plan for pre-testing custom-made surveys before IOT&E. This can often be accomplished during developmental testing or operational assessments occurring before IOT&E.
- Survey questions should be clear and unbiased (e.g., no leading questions)

Surveys should use quantitative (e.g., Likert-scale) and qualitative responses (open ended questions); quantitative data should be coded, compiled and summarized using statistical methods to aid in system characterization in concert with the measures employed in field testing.

References

Inspector General Report, May 15, 2015

Reporting of Operational Test and Evaluation (OT&E) Results, DOT&E, January 6, 2010

Test and Evaluation Policy Revisions, DOT&E, December 22, 2007

Guidance on the Use of Design of Experiments (DOE), DOT&E, October 19, 2010

Guidance on the Use and Design of Surveys in Operational Test and Evaluation (OT&E), DOT&E, June 23, 2014

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