

Mission Focused Metrics – Guidance

General Guidance

TEMPs should include quantitative mission-focused metrics (also referred to as response variables) for effectiveness and suitability. Evaluation metrics are key to good test designs; poorly-chosen or poorly-defined measures, even if they are 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 of the system.

Choosing Metrics

The selection of evaluation metrics is a critical part of test design effort, and should occur as test planning begins. Step 1 is to identify the critical operational issues (COIs): what capability is this system intended to provide? Once this is known, testers should select appropriate metrics that provide a means to measure performance and provide data for answering the COIs. Ideally, the metrics will provide a determination of mission capability, lend well to good experimental design ([DOE](#)), and encapsulate the reasons for procuring the system.

Evaluation metrics are ideally selected from KPPs, measures of effectiveness, measures of suitability, critical technical parameters, KSAs, and/or measures of performance already documented in requirements documents. Although many metrics can be used to characterize system performance in a given mission, it is desirable that one or two primary metrics be identified to be the focus the evaluation of mission effectiveness and used in concert with design of experiments methodologies. Additional secondary metrics 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 primary variable, even though other metrics to characterize success in the dependent portions of the kill chain are valuable (e.g., detection, identification, time to engage, engagement range).

Exceptions to using CDD/CPD-defined Metrics

The primary metric identified for test design need not be the KPPs. Often KPPs are insufficient for measuring the mission effectiveness 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 mission-focused metrics 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.

When testers select these primary metrics, 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

Mission Focused Metrics – Guidance

- Provide sufficient data for the secondary metrics needed to characterize system performance.

Types of Metrics

Response variables can be continuous or discrete. Examples of continuous responses include time to detect, miss distance, and range of engagement. Examples of discrete responses include hit/miss, message complete/not complete, and detect/not detect. A continuous response variable is preferred to a discrete one, since it will almost always require a smaller sample size and fewer test resources for the risk levels chosen (confidence and power).

Continuous variables also often contain more information regarding the performance of the system, whereas a corresponding discrete variable 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.

Definitions of Metrics

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

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

Metric Selection for Survey Data and Expert Panels

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 performance. This is particularly true when quantitative data is scarce due to expensive field testing or low sample sizes. 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:

Mission Focused Metrics – Guidance

- Clearly identify survey objectives: TEMP should indicate which COIs will be addressed by survey data
- Surveys should be tested on an appropriate group to reveal if questions are confusing or if information is missing
- 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 metrics 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.](#)