Best practices for highly effective test design; Part 1 – Beginners’ guide to mapping the T&E strategy

Luis A. Cortes, PE

The goal of the STAT T&E COE is to assist in developing rigorous, defensible test strategies to more effectively quantify and characterize system performance and provide information that reduces risk. This and other COE products are available at www.AFIT.edu/STAT.
This page intentionally left blank.
Abstract

Test & Evaluation (T&E) produces knowledge about the true capability of a system by comparing the analysis of empirical observations obtained from stimulating a system to requirements and standards. The ultimate goal is to transform the knowledge gained from testing into “decision-quality-information” to inform the system engineering process and key acquisition decisions. A practical and efficient method for transforming knowledge into information is design of experiments (DOE). DOE is the systematic integration of well-defined and structured scientific strategies for gathering empirical knowledge about a system and transforming it into information by using statistical methods. Coleman & Montgomery (1993) proposed guidelines for design of experiments. This document consists of two parts. Part 1 provides a strategy for mapping the T&E strategy that will be used to drive the development of the design of experiments. Part 2 provides best practices for planning, designing, executing, and analyzing a test within the Department of Defense’s typical T&E strategy: (1) well-defined, end-to-end, mission-oriented objectives within the scope of the decision to be informed; (2) mission-oriented, traceable measures of capability and readiness; (3) complete coverage of the test space; (4) designs with good properties and test design structures; (5) disciplined test protocols; (6) statistical data analysis and assessment techniques and standardized evaluation criteria; and (7) reliable products to inform decisions. The use of these practices will result in robust and disciplined T&E strategies that can yield better interpretation of test observations and, consequently, a better understanding of the state of the system capabilities and the risks associated with the decisions to be made.

Key words: design of experiments, DOD test and evaluation, test design
This page intentionally left blank.
# Table of Contents

1.0 Introduction .................................................................................................................. 1

2.0 Overview of Test & Evaluation in DOD .................................................................. 8
   2.1 Developmental Test & Evaluation ......................................................................... 9
   2.2 Operational Test & Evaluation ............................................................................. 10
   2.3 Life Fire Test & Evaluation ................................................................................... 10
   2.4 Integrated Test & Evaluation ............................................................................... 10

3.0 Mapping the Test & Evaluation Strategy ................................................................. 11

4.0 T&E paradigm ............................................................................................................. 14

References ......................................................................................................................... 17
Best practices for highly effective test design; Part 1 - Beginners’ guide to mapping the T&E strategy

1.0 Introduction

The Department of Defense (DOD) makes significant investment decisions in the acquisition of weapons systems, defense business systems, national security systems, and joint systems. For instance, the investment for just 86 weapon system programs in 2012 was $1.6T. Because the imperative is to deploy a cost effective, safe, survivable, operationally effective, and operationally suitable system that meets the warfighter needs, every decision must be well informed. Figure 1 outlines the wealth of decision-quality-information required at key decision points of the defense acquisition system.

---

In 2013, the Government Accounting Office (GAO)\(^2\) correlated a development cost growth of $271B of 86 weapons systems programs to a "knowledge deficit" early in the program. This knowledge deficit ripples through the acquisition system feeding upon itself and leaving decision makers with less and less knowledge to make informed decisions on “how and when to proceed to the next acquisition phases that require commitment for additional funding”. The cost growth reduces the funding available for other priorities, including solutions to reduce the knowledge deficit itself. A knowledge deficit means “the program is proceeding throughout the defense acquisition system without sufficient knowledge about its technologies, design, or manufacturing processes, and faces unresolved risks that could lead to cost increases and schedule delays”. GAO also correlated the cost growth rate with the maturity of the technologies—86% for programs that start with immature technologies and about 43% for those that start with nearing maturity technologies.

GAO assessed the program’s attainment of product knowledge by scoring eight knowledge-based acquisition practices measured at three critical points in the acquisition process as shown in Figure 2: Milestone B, Milestone C, and half-way between the milestones at the transition from Integration to Demonstration. The

---

\(^2\) GAO Report to Congressional Committees *Defense Acquisitions; Assessment of Selected Weapon Programs* (GAO-13-294SP), March 2013
knowledge points are: (1) resources and requirements match; (2) design stability; and (3) manufacturing process maturity. The knowledge-based shortfalls includes demonstrating critical technologies in relevant or realistic environments and testing integrated or representative prototypes. Figure 3 is an example of the scorecard.

**Figure 3. Examples of knowledge scorecards.**

GAO indicated that there is a significant need for programs to improve their knowledge-based approach when reducing gaps in technology, design, and production knowledge. GAO stated that the likelihood of a weapon system to be delivered within its estimated cost and schedule is a function of the knowledge the program has reached by each of the three key decision points:

“Positive acquisition outcomes require the use of a knowledge-based approach to product development that demonstrates high levels of knowledge before significant commitments are made.”

One of the most effective and efficient means of gaining knowledge about a system and reducing the knowledge gaps in technology, design, and production is Test & Evaluation (T&E):
DOD Directive 5000.01; The Defense Acquisition System

“Test and evaluation shall be integrated throughout the defense acquisition process. Test and evaluation shall be structured to provide essential information to decision-makers, assess attainment of technical performance parameters, and determine whether systems are operationally effective, suitable, survivable, and safe for intended use. The conduct of test and evaluation, integrated with modeling and simulation, shall facilitate learning, assess technology maturity and interoperability, facilitate integration into fielded forces, and confirm performance against documented capability needs and adversary capabilities as described in the system threat assessment. “

Test, in a native context, is the clearly defined and well formulated process of submitting a system fixed in pre-determined conditions (such as time, space, energy, operational conditions, etc.) to deliberate and planned changes in those conditions to generate empirical observations for judging changes in the attributes of the system (such as physical properties, quality, performance, reliability, etc.). Likewise, Evaluation is the rigorous and systematic interpretation of those changes through examination, analysis, assessment, and comparison of the empirical observations to standards, requirements, and specifications. Thus, the “T” in T&E refers to “establishing appropriate conditions, applying physical changes to a system, and generating empirical observations” while the “E” refers to “examining, analyzing, assessing, and comparing the observations to requirements, specifications, and standards to interpret the changes of the system in the conditions of the test”. The trade-off, illustrated in Figure 4, is to determine the true state of a system, in all possible scenarios, with a finite number of samples that yield valid conclusions while minimizing the risks of making bad decisions.

---

3 DOD Directive 5000.01, The Defense Acquisition System. (20 Nov 07)
A balance in the trade-off is achieved through an efficient and effective T&E strategy that includes:

1. well-defined, end-to-end, mission-oriented objectives within the scope of the decision to be informed;
2. mission-oriented, traceable measures of capability and readiness;
3. complete coverage of the test space;
4. designs with good properties and test design structures;
5. disciplined test protocols;
6. statistical data analysis and assessment techniques and standardized evaluation criteria; and
7. reliable products to inform decisions. However, there are instances in which the strategy falls short.

Earlier in 1998, the National Research Council concluded that T&E did not take full advantage of the benefits afforded by statistics. Statistics is the mathematical science dealing with the planning, collection, analysis, organization, explanation, and presentation of data. Statistics seeks to investigate and establish relationships between events. Scientific test and analysis techniques (STAT) exploits the tremendous power of grafting statistical methods and testing to acquire a deeper knowledge of the system’s capabilities and to maximize the utility of the information. STAT are knowledge-based scientific and statistical methods and processes used to enable the development of efficient, rigorous test strategies that yield defensible results. STAT ensures scientifically rigorous testing and better interpretation of test data, which helps inform decision makers of the true state of system capabilities across the entire technical and operational requirements space and the risks associated with the decisions to be made. STAT encompasses such techniques as design of experiments and reliability growth. DOD Instruction
5000.02, *Operation of the Defense Acquisition System*[^4], recognizes the value of STAT and a comprehensive evaluation methodology:

**DOD Instruction 5000.02; Operation of the Defense Acquisition System**

“Use scientific test and analysis techniques to design an effective and efficient test program that will produce the required data to characterize system behavior across an appropriately selected set of factors and conditions.”

“Ensure that each major developmental test phase or event in the planned test program has a well-defined description of the event, specific objectives, scope, appropriate use of modeling and simulation, and an evaluation methodology.”

“Describe an evaluation methodology in the TEMP starting at Milestone A that will provide essential information on programmatic and technical risks as well as information for major programmatic decisions. Starting at Milestone B, the evaluation methodology will include an evaluation framework to identify key data that will contribute to assessing progress toward achieving: key performance parameters, critical technical parameters, key system attributes, interoperability requirements, cybersecurity requirements, reliability growth, maintainability attributes, developmental test objectives, and others as needed. In addition, the evaluation framework will show the correlation/mapping between test events, key resources, and the decision supported. The evaluation methodology will support a Milestone B assessment of planning, schedule, and resources and a Milestone C assessment of performance, reliability, interoperability, and cybersecurity.”

DOE is the cornerstone statistical method embedded in the STAT portfolio.

---

strategies for gathering empirical knowledge about a system or process using statistical methods for planning, designing, executing, and analyzing an experiment. The core principle is to make purposeful changes to the input variables of a process or system to observe and exploit the changes in the output response. DOE adds rigor and discipline to T&E and facilitates a comprehensive understanding of the tradeoffs in the techno-programmatic domains: risks, cost, and utility of information. The use DOE in T&E allows for:

- Statistically identifying the performance drivers and their interactions.
- Characterizing system performance over the entire battle space.
- Illuminating the tradeoff between number of tests, risks, and amount and quality of the information obtained.
- Developing empirical models that could be useful for tactical decision making and performance assessment.
- Providing means for optimizing system performance.

Testing using DOE is faster, more effective, more economical, and more informative than other traditional methods of test. DoE provides a direct relationship between the level of knowledge desired and the cost required to achieve that level of knowledge. DOE provides decision makers with a unique management tool to help them determine how much knowledge they can afford.

### Guidelines

- Develop and implement, as early as possible, a knowledge-based process to inform the systems engineering process and key acquisition decisions.
- Use scientific test and analysis techniques (STAT) to characterize system performance across the entire technical and operational space.
- Develop and implement evaluation frameworks that correlate the key data used to assess capability maturity to test events, resources, and decisions.
This document provides basic guidelines and principles to help T&E professionals enhance their techniques for planning, designing, executing, and analyzing effective and efficient tests. The document presumes the reader has a basic knowledge of T&E and the operation of the defense acquisition system.

The document consists of two parts. Part 1 provides guidelines for mapping the T&E strategy, and consists of two additional sections. Section 2 is an overview of T&E within DOD. Section 3 discusses the mapping of the strategy. Part 2 consists of one section: best practices for highly effective test design.

2.0 Overview of Test & Evaluation in DOD

The Defense Acquisition Guidebook (DAG)\(^5\) describes in detail the overall DOD acquisition process. The DOD Test and Evaluation Management Guide\(^6\) is a generic desk reference to assist program management and T&E personnel in executing their responsibilities. The DOD Test and Evaluation Management Guide provides the following definitions:

- **Test** denotes any procedure designed to obtain, verify, or provide data for the evaluation of any of the following: (1) progress in accomplishing developmental objectives; (2) the performance, operational capability, and suitability of systems, subsystems, components, and equipment items; and (3) the vulnerability and lethality of systems, subsystems, components, and equipment items.

- **Evaluation** denotes the process whereby data are logically assembled, analyzed, and compared to expected performance to aid in systematic decision making. It may involve review and analysis of qualitative or quantitative data obtained from design reviews, hardware inspections,

---


modeling and simulation (M&S), hardware and software testing, metrics review, and operational usage of equipment.

- **Test & Evaluation** is a process by which a system or components are tested and results analyzed to provide performance related information. This information has many uses, including risk identification and mitigation as well as providing empirical data to validate models and simulations. T&E enables an assessment of the attainment of technical performance, specifications, and system maturity to determine whether systems are operationally effective, suitable, and survivable for their intended use.

DOD exercises three formal and statutory types of tests administered by the Office of the Secretary of Defense (OSD): developmental test & evaluation (DT&E), operational test & evaluation (OT&E), and live fire test & evaluation (LFT&E). The simultaneous execution and independent assessment of developmental and operational testing is called integrated test & evaluation (IT&E). The Deputy Assistant Secretary of Defense for Developmental Test and Evaluation (DASD(DT&E)) provides oversight and policy for DT&E of certain acquisition programs within OSD in accordance with Title 10, United States Code (U.S.C.). Likewise, the Director of Operational Test and Evaluation (DOT&E) provides oversight and policy for OT&E and LFT&E. Service components have their own directives, guidance, organizations, resources, ranges, and facilities for their needs.

**2.1 Developmental Test & Evaluation.**

DT&E verifies a system is built correctly and meets the technical requirements specified in the contract. DT&E is conducted throughout the lifecycle of a system to inform the systems engineering process and acquisition decisions, to help manage design and programmatic risks, and to evaluate the combat capability of the system and its ability to provide information to the users.
2.2 Operational Test & Evaluation.

OT&E validates a system executes its mission in a realistic operational environment. OT&E determines the operational effectiveness, operational suitability, and the survivability of a system. Operational effectiveness is the overall degree of mission accomplishment of a system when used by representative personnel in the environment planned or expected for operational employment considering organization, training, doctrine, tactics, survivability, vulnerability, and threat. Operational suitability is the degree to which a system can be placed in field use considering reliability, availability, compatibility, transportability, interoperability, wartime usage rates, maintainability, safety, human factors, manpower supportability, logistics supportability, documentation, environmental effects, and training requirements. The evaluation of operational suitability informs about the ability of sustaining operations tempo over an extended period while conducting realistic missions.

2.3 Life Fire Test & Evaluation.

LFT&E provides an assessment of the vulnerability or lethality of a system as it progresses through design and development that informs on potential user casualties considering the susceptibility to attack and combat performance of the system. A sound LFT&E strategy incorporates design changes resulting from testing and analysis before proceeding beyond Low Rate Initial Production (LRIP).

2.4 Integrated Test & Evaluation.

IT&E is “the collaborative planning and collaborative execution of test phases and events to provide shared data in support of independent analysis, evaluation, and reporting by all stakeholders, particularly the development (both contractor and government) and operational test and evaluation communities.” Integrated testing focuses the test strategy on designing, developing, and implementing

---

7 OSD Memorandum Definition of Integrated Testing dated 25 April 2008
comprehensive and economical test events that produce information that can be used collaboratively and unobtrusively by the various organizations participating in the test event to provide evaluation results from different points of views to the decision makers.

3.0 Mapping the Test & Evaluation Strategy

Defense acquisition programs are designated by category and type. The acquisition category (ACAT) reflects the estimated statutorily defined cost threshold or a designation by the Defense Acquisition Authority (DAE)—the Under Secretary of Defense for Acquisition, Technology, and Logistics (USD(AT&L)). The acquisition types include Major Defense Acquisition Program (MDAP), Major Automated Information System (MAIS) Program, or Defense Business System (DBS) Program. Depending on the category, a test strategy is developed to describe how T&E supports the acquisition strategy. While test strategies can be different, they all adhere to sound and proven T&E processes and practices.

The Test and Evaluation Master Plan (TEMP) is the main test planning and management tool for the program. It describes T&E throughout the program life cycle, the fundamental T&E actions, goals, processes, and practices that support the acquisition strategy, and the overall approach to meet the programmatic, technical, and operational needs. The TEMP is a regulatory requirement stipulated by DOD Instruction 5000.02. The initial TEMP submission is due at the Milestone A decision review. Updates are required for the Development Request for Proposal (RFP) Release decision point, for approval at the Milestone B decision review, for the Milestone C decision review, and for the Full Rate Production (FRP) or Full Deployment Decision (FDD) decision point. A specific format for the TEMP is required for ACAT I, IA, and designated oversight programs. TEMP outline guidance can be found at http://www.dote.osd.mil/docs/dote-temp-guidebook/20130712_TEMP_Guide_2.1.pdf. The TEMP documents the overall
structure, strategy, schedule, resources, objectives, and evaluation frameworks of the T&E program.

Figure 5 illustrates the concept behind the T&E strategy and the place of design of experiments within it. The figure reflects the symmetry between DT&E and OT&E. T&E starts with the decomposition of the capability requirements, which are specified by the warfighter and documented in the Capability Requirements Document (CRD), into technical and operational requirements. Those requirements are expressed in the form of measurable and testable technical performance measures (TPM) that make-up the backbone of the T&E strategy—the Developmental Evaluation Framework and Operational Evaluation Framework. The evaluation frameworks describe how a system will be evaluated against its technical and operational requirements to inform programmatic, technical, and operational decisions. They link the capability requirements, the TPM, test issues for examination, and test objectives to the decisions that will be informed.

The requirements are typically associated with two types of test issues: the critical technical issues (CTI), which are associated with the technical requirements, and the critical operational issues (COI), which are associated with the operational requirements. These issues determine how test results will be judged, and are examined in technical or operational evaluations throughout the acquisition process to refine performance requirements and the system design or to enrich milestone decision
reviews. They are normally stated as questions and often require a sufficient number of measures to adequately determine an answer.

CTI are issues that must be examined in DT&E to evaluate technical parameters, characteristics, and engineering specifications. CTI, also referred to as developmental test issues (DTI), are analogous to the COI but in the context of technical evaluation. CTI are normally resolved based on the demonstration or evaluation of key performance parameters (KPP), key systems attributes (KSA), and critical technical parameters (CTP). KPP, KSA, and CTP are a small number of TPM derived from the CRD and the Systems Engineering Management Plan (SEMP). They provide quantitative and qualitative information on how well a system, when performing the mission essential tasks specified in the CRD, is designed and manufactured. KPP and KSA reflect a significant capability or characteristic that failure to meet the threshold can be cause for the concept or system selection to be reevaluated, or the program to be reassessed or terminated. CTP are parameters usually at a higher level of risk during design.

COI are operational effectiveness and operational suitability issues that must be examined in OT&E to evaluate the capability of a system to perform its mission. The resolution of the COI is based on the evaluation of measures of effectiveness (MOE) or measures of suitability (MOS) using a standard criteria. MOE and MOS are the subset of TPM that reflect the operational requirements derived from the KPP and the CRD. MOE inform how well a system or performs its mission under a given set of conditions while MOS inform how ready, supportable, survivable, and safe a system is to sustain effective performance in combat and peacetime operations. COI address the overall system’s operational capability when examined in realistic operational mission environments and operated by warfighters.

A well-designed test should produce the maximum amount of information with the minimum amount of test time or test events to inform its objective. Often, the ultimate objective of a test is to obtain information to enable a complete
assessment of the system capabilities and its performance. The use of STAT can simultaneously help generate test efficiencies, improve the fidelity of test results, illuminate risks, enable better-informed decisions, and ultimately, enable fielding a more effective, suitable, and survivable system.

DOE is the cornerstone statistical method embedded in the STAT portfolio. DOE is the systematic integration of well-defined and structured scientific strategies for gathering empirical knowledge about a system or process using statistical methods for planning, designing, executing, and analyzing an experiment. The core principle is to make purposeful changes to the input variables of a process or system to observe and exploit the changes in the output response. DOE adds rigor and discipline to T&E and facilitates a comprehensive understanding of the tradeoffs in the techno-programmatic domains: risks, cost, and utility of information. The use DOE in T&E allows for:

- Statistically identifying the performance drivers and their interactions.
- Characterizing system performance over the entire battle space.
- Illuminating the tradeoff between number of tests, risks, and amount and quality of the information obtained.
- Developing empirical models that could be useful for tactical decision making and performance assessment.
- Providing means for optimizing system performance.

4.0 T&E paradigm

T&E are means to provide essential information to decision makers regarding the capabilities and operational conditions of a system. The paradigm of a well-defined test typically involves:

- Clear, testable, and measurable test objectives
  - Systems engineering requirements
  - Operational requirements
- Previously identified deficiencies
- Added capabilities
- Continuing assessment
- Verification and validation (V&V)
- Tactics, techniques, and procedures (TTP)

➤ Suitable measures to assess or evaluate the objectives
  - Key performance parameters (KPP)
  - Key system attributes (KSA)
  - Critical technical parameters (CTP)
  - Measures of effectiveness (MOE) and measures of suitability (MOS)
  - Measures of performance (MOP)

➤ Appropriate analysis and assessment methods and evaluation criteria
  - Data analysis plan
  - Evaluation criteria

➤ Data needed
  - Data collection plan (DCP)
  - Data management plan (DMAP)

➤ Adequate test venues and test resources
  - Hardware in the loop (HWIL)
  - Modeling and simulation (M&S)
  - Live events
  - Real world events

➤ Test plans
  - Scenario plans and certification
  - Instrumentation plans
  - Test execution plans

➤ Data analysis and assessment
  - Functionality
  - Performance
  - Suitability
Results and outcomes

- Modify requirements
- Correct training deficiencies
- Correct system deficiencies (design, implementation, suitability)
- Improve capability
- Determine performance limiters
- Improve models
References
