Assessing Rigor in Test Plans

Authored by: Michael Harman

30 April 2016

Revised 27 August 2018

The goal of the STAT 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.

STAT Center of Excellence
2950 Hobson Way – Wright-Patterson AFB, OH 45433

Table of Contents

Revision 1, 27 Aug 2018: Formatting and minor typographical/grammatical edits.

Executive Summary

Department of Defense (DoD) Instruction 5000.02 requires the use of scientific test and analysis techniques (STAT) but there is no definition of acceptable rigor or plan for assessments.

Assessing rigor must be a formal, objective, and repeatable process with the goal of evaluating test plans to identify them as sufficiently rigorous or call out specific topics for targeted improvement. This paper proposes an assessment process based on the definition of rigor and ties it to test planning metrics. Applying a detailed rigor definition to test planning steps supports the creation of an assessment matrix. The assessment is not intended to produce a numeric score or letter grade but facilitate quality assurance to planning activities and identify risk in the design. The matrix metrics are broad enough to be applied to all tests and design types.

Keywords: scientific test and analysis techniques, rigorous testing, assessment

Introduction

There is now policy in Department of Defense (DoD) Instruction 5000.02 (USD AT&L, 2015) requiring the use of scientific test and analysis techniques (STAT) to include rationale and calculations but there is no indication of how the level of rigor is to be assessed. Furthermore, rigor is not defined in 5000.02, the defense acquisition guidebook (DAG) (DAU, April 2016), or the DoD Test and Evaluation Management Guide (DOD, 2012). Rigor is imposed to ensure sufficient data is collected for a quantitative evaluation of performance. But assessing rigor must be a formal, objective, and repeatable process with the goal of evaluating test plans to identify them as sufficiently rigorous or call out specific topics for targeted improvement. Yang (2016) provides an example for the medical research community with extensive descriptions and links to references and examples. This is just one of many sites dedicated to rigor that can be found among research communities. Without such a source, this paper proposes an assessment process based on actionable words contained in the definition of rigor and ties them to test planning metrics. This assessment requires a fairly deep technical knowledge of the system and test plan under review and is aimed at test planners and immediate supervisors. A set of questions for higher leadership is contained in the “Critical STAT Questions” Best Practice (STAT COE, March 2015).

The Assessment Process

Rigor Defined

A combined definition for rigor, derived from definitions for rigor at merriam-webster.com, oxforddictionaries.com, and thefreedictionary.com, includes the following, most relevant, terms:

- Accurate
- Exact

- Exhaustive
- Meticulous
- Precise

If we continue one step further with the relevant definitions of these first five words, we can begin to relate them to the planning process. Excerpts from the complete definitions provided at thefreedictionary.com.

- Accurate
    - Deviating only slightly or within acceptable limits from a standard
- Exact
    - Accurate measurements
    - Small margins of error
    - Strict adherence to standards or rules
- Exhaustive
    - Treating all parts or aspects without omission; thorough
- Meticulous
    - Showing or acting with extreme care and concern for details.
- Precise
    - Clearly expressed or delineated

Relating Definitions to Planning Steps

Figure 1 is a schematic of the STAT COE test planning process, the specifics of which are covered in the STAT COE Guide (STAT COE, 2015).

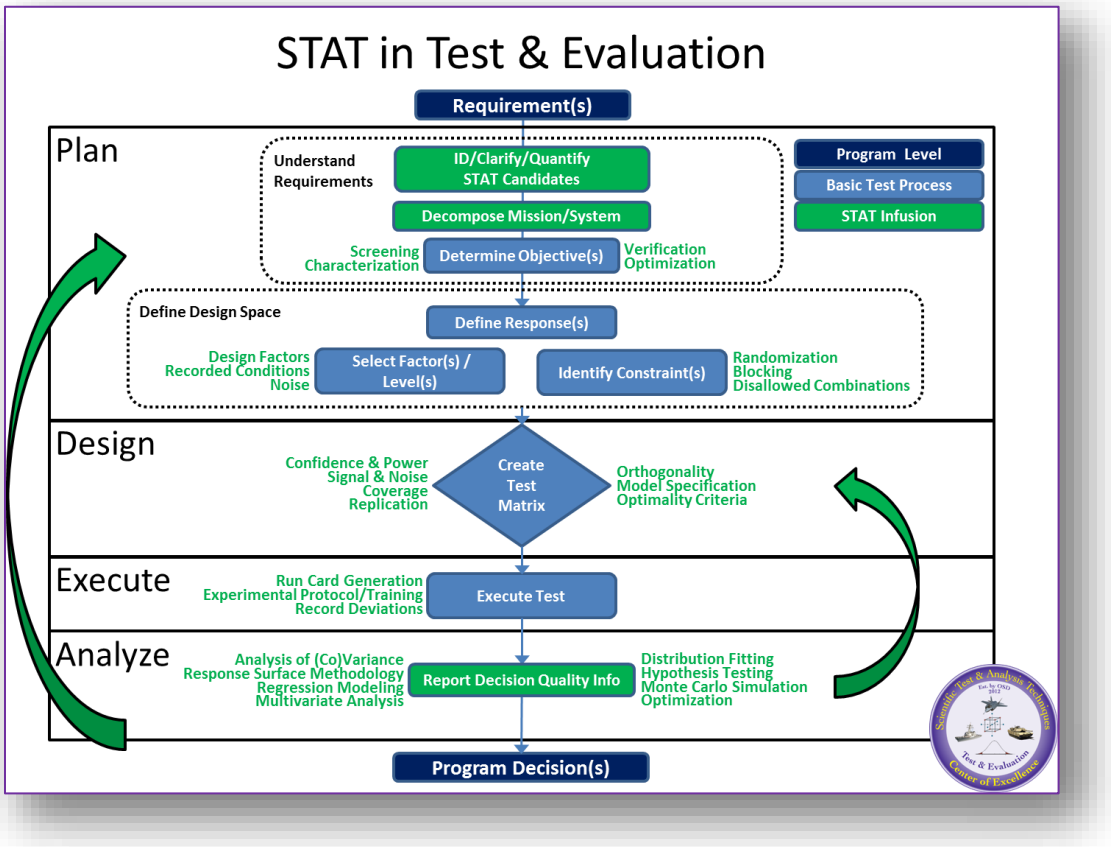

Figure 1: STAT COE test planning process

The main phase components (in temporal order) are:

- Requirements
- Objectives
- Responses
- Factors
- Constraints
- Design Type
- Analysis Goals

Applying the detailed rigor definitions to each of the planning components supports the creation of an assessment matrix.

The Assessment Matrix

The matrix (excerpted in Figure 2) contains the STAT Phase and Phase Components and maps them to the rigor terms and definitions. One can see that multiple rigor terms and definitions may apply to a single phase component (e.g. Requirements).

STAT Phase	Phase Component	Rigor Term	Expanded Term Definition
1: Planning & Strategy Development	1: Requirements	Precise	Clearly expressed or delineated
1: Planning & Strategy Development	1: Requirements	Exact	Accurate measurements
1: Planning & Strategy Development	1: Requirements	Exact	Small margins of error
1: Planning & Strategy Development	2: Strategy	Meticulous	Showing or acting with extreme care and concern for details
1: Planning & Strategy Development	2: Strategy	Precise	Clearly expressed or delineated
1: Planning & Strategy Development	2: Strategy	Exhaustive	Treating all parts or aspects without omission; thorough
1: Planning & Strategy Development	3: Objectives	Precise	Clearly expressed or delineated
1: Planning & Strategy Development	3: Objectives	Exhaustive	Treating all parts or aspects without omission; thorough
1: Planning & Strategy Development	4: Decompose System	Precise	Clearly expressed or delineated
1: Planning & Strategy Development	5: Responses	Precise	Clearly expressed or delineated
1: Planning & Strategy Development	5: Responses	Exact	Accurate measurements
1: Planning & Strategy Development	5: Responses	Exhaustive	Treating all parts or aspects without omission; thorough

Figure 2: Assessment matrix excerpt

Building on the Expended Term Definition, a Component Metric/Criterion has been created to provide the assessor with a way to evaluate the test plan under review (Figure 3).

Phase Component	Rigor Term	Expanded Term Definition	Component Metric/Criterion
1: Requirements	Precise	Clearly expressed or delineated	Clear and unambiguous
1: Requirements	Exact	Accurate measurements	Measurable, testable
1: Requirements	Exact	Small margins of error	Quantifiable
2: Strategy	Meticulous	Showing or acting with extreme care and concern for details	Decomposes the system/mission into clearly definable and manageable segments
2: Strategy	Precise	Clearly expressed or delineated	Defines a sequential test process by which knowledge is gained
2: Strategy	Exhaustive	Treating all parts or aspects without omission; thorough	Identifies all systems/components required for testing
3: Objectives	Precise	Clearly expressed or delineated	Unbiased, specific, measurable, practical
3: Objectives	Exhaustive	Treating all parts or aspects without omission; thorough	Cover all facets of the system requirements
4: Decompose System	Precise	Clearly expressed or delineated	Small, specific, quantifiable mission segments that drive design scope
5: Responses	Precise	Clearly expressed or delineated	Clear, unambiguous, and specifically related to the stated requirements
5: Responses	Exact	Accurate measurements	Measurable, testable
5: Responses	Exhaustive	Treating all parts or aspects without omission; thorough	If not a discreet number (e.g. time series), a sampling method is needed

Figure 3: Assessment matrix metrics and criterion

The metrics/criterion prompts the assessor to consider the depth and quality of the test plan details. The actual test plan content should

- Be Objective
- Be Actionable
- Support the Basis for a Decision
- Define Passing and/or Failing.

Assessment

The assessment is not intended to produce a numeric score or letter grade. Instead, this process applies quality assurance to test planning and identifies risk in the design which may not be identified in the Test and Evaluation Master Plan (TEMP). Any questions that cannot be answered specifically or quantitatively require further action to the satisfaction of the reviewer. The full matrix is contained in the appendix.

Obtaining the Assessment Matrix

The matrix can be obtained by emailing COE@AFIT.edu and is provided in Microsoft Excel to be responsive and flexible for the reviewer. Non-applicable lines can be hidden and the information can be sorted in various ways.

Applicability to Various Test and Design Types

Test and design types vary based on the system, goals, phase of acquisition and other reasons. The matrix metrics and criterion were developed from a statistical design perspective. However, the rigor terms and expanded definitions are broad enough to be applied to all tests and design types, with the metrics used for guidance.

For instance, component qualification testing may be dictated by policy or industry best practices and not developed from scratch by the tester. However, there are still choices to be made in the tailoring of the test plan and questions regarding accuracy, exactness, exhaustiveness, meticulousness and precision still apply. More strategically, one could assess the policy itself to determine if it satisfies desired levels of rigor or could be improved.

Acceptance testing or sampling plans similarly have some relevant statistical metrics that should be considered during the planning and design phase which are worth questioning during the review process. This includes reliability test time determination and the level of risk assumed in the duration or sample size calculation.

Conclusion

Rigor must be designed into test plans and a methodical approach will maximize that effort. This assessment matrix assists test plan reviewers in their quality assurance role and is broadly applicable across many test types and designs. The outcome of the process is assurance that the test contains sufficient rigor as-is, or will once identified risk areas are addressed.

References

Burke, Sarah, et al. “Guide to Developing an Effective STAT Test Strategy V5.0.” Scientific Test and Analysis Techniques Center of Excellence (STAT COE), 31 Dec. 2017.

Kensler, Jennifer, et al. “Critical STAT Questions DoD Leadership Should Ask: Injecting Rigor into the Test and Evaluation Process through Effective Inquiry.” Scientific Test and Analysis Techniques Center of Excellence (STAT COE), 18 March 2015.

Under Secretary of Defense for Acquisition Technology and Logistics. “Operation of the Defense Acquisition System DOD INST 5000.02.” 7 Jan. 2015.

United States. Department of Defense. "Defense Acquisition Guidebook." Defense Acquisition University (DAU). https://www.dau.mil/tools/dag.

United States. Department of Defense. "Test and Evaluation Management Guide." 6th ed., Defense Acquisition University (DAU), Dec. 2012. https://www.dau.mil/guidebooks/Shared%20Documents/Test_and_Evaluation_Mgmt_Guidebook.pdf.

Yang, Lynda J-S, et al. "Methodology Rigor in Clinical Research." *Plastic and Reconstructive Surgery,* vol. 129, no. 6, 2012, pp. 979e-988e.

Appendix: Assessment Matrix

STAT Phase	Phase Component	Rigor Term	Expanded Term Definition	Component Metric/Criterion
1: Plan & Strategy Dev	1: Requirements	Precise	Clearly expressed or delineated	Clear and unambiguous
1: Plan & Strategy Dev	1: Requirements	Exact	Accurate measurements	Measurable, testable
1: Plan & Strategy Dev	1: Requirements	Exact	Small margins of error	Quantifiable and states pass/fail criteria
1: Plan & Strategy Dev	2: Strategy	Meticulous	Showing or acting with extreme care and concern for details	Decomposes the system/mission into clearly definable and manageable segments
1: Plan & Strategy Dev	2: Strategy	Precise	Clearly expressed or delineated	Defines a sequential test process by which knowledge is gained
1: Plan & Strategy Dev	2: Strategy	Exhaustive	Treating all parts or aspects without omission; thorough	Identifies all systems/components required for testing
1: Plan & Strategy Dev	3: Objectives	Precise	Clearly expressed or delineated	Unbiased, specific, measurable, of practical consequence
1: Plan & Strategy Dev	3: Objectives	Exhaustive	Treating all parts or aspects without omission; thorough	Cover all facets of the system requirements
1: Plan & Strategy Dev	4: Decompose System	Precise	Clearly expressed or delineated	Small, specific, quantifiable mission segments that drive design scope
1: Plan & Strategy Dev	4: Decompose System	Exhaustive	Treating all parts or aspects without omission; thorough	Sequential, progressive testing so learned information supports later investigations
1: Plan & Strategy Dev	5: Responses	Precise	Clearly expressed or delineated	Clear, unambiguous, and specifically related to the stated requirements, objective
1: Plan & Strategy Dev	5: Responses	Exact	Accurate measurements	Measurable, testable
1: Plan & Strategy Dev	5: Responses	Exhaustive	Treating all parts or aspects without omission; thorough	If not a discreet number (e.g. time series), a sampling method is needed
2: Test Design	1: Factors	Exhaustive	Treating all parts or aspects without omission; thorough	All relevant factors are considered for the given response (all are not required to be in the final design)
2: Test Design	1: Factors	Precise	Clearly expressed or delineated	Directly applicable to the given response
2: Test Design	1: Factors	Precise	Clearly expressed or delineated	Continuous where possible
2: Test Design	1: Factors	Exhaustive	Treating all parts or aspects without omission; thorough	Uncontrollable factors left out of the design space are assessed for noise impact to the response and documented

2: Test Design	2: Factor Levels	Meticulous	Showing or acting with extreme care and concern for details	Sufficient number to support analysis goals (e.g. modeling) without overdesigning
2: Test Design	3: Factor Space Coverage	Exhaustive	Treating all parts or aspects without omission; thorough	Operating region is covered by sufficient factor level range (lows and highs defined for every factor considered in design)
2: Test Design	3: Factor Space Coverage	Exhaustive	Treating all parts or aspects without omission; thorough	Combinatorial design space demonstrates at least 80% coverage of all desired pairings
2: Test Design	4: Statistical Confidence	Exact	Small margins of error	Highest confidence is used that can be afforded
2: Test Design	5: Signal and Noise	Precise	Clearly expressed or delineated	Operationally relevant delta is clearly stated by user and agreed upon
2: Test Design	5: Signal and Noise	Precise	Clearly expressed or delineated	Signal and Noise should be defined separately, not just used as an abstract ratio for planning
2: Test Design	5: Signal and Noise	Exact	Small margins of error	Inherent noise is understood from previous test data or estimated using clear analytical methods
2: Test Design	6: Design Type	Meticulous	Showing or acting with extreme care and concern for details	Design is capable of producing the analysis required
2: Test Design	6: Design Type	Exhaustive	Treating all parts or aspects without omission; thorough	Design includes all relevant factors for the given response
2: Test Design	6: Design Type	Exact	Strict adherence to standards or rules	Statistical design assumptions, limitations, and analysis methods are followed
2: Test Design	6: Design Type	Exact	Strict adherence to standards or rules	Qualification/acceptance tests comply with published standards, policies, or known best practices
2: Test Design	7: Reliability Test Time	Exact	Strict adherence to standards or rules	Sampling plan/test time accounts for both Type I (alpha) and Type II (beta) error rates (no more than 20% each)
2: Test Design	7: Reliability Test Time	Precise	Clearly expressed or delineated	reliability proportion or mean time is clearly stated, measureable, and relevant
2: Test Design	8: Prediction Variance	Exact	Small margins of error	Smallest affordable value to support analysis goals. Should be less than planned signal to noise ratio in at least 80% of factor space
2: Test Design	9: Confounding	Meticulous	Showing or acting with extreme care and concern for details	Design minimizes or avoids purposeful correlation of results
2: Test Design	10: Statistical Power	Exact	Small margins of error	Minimum of 80% for main effects and unknown effects under investigation
3: Execution Planning	1: Assumptions	Precise	Clearly expressed or delineated	All assumptions are listed in a common planning document

3: Execution Planning	2: Constraints	Precise	Clearly expressed or delineated	All constraints are listed in a common planning document
3: Execution Planning	2: Constraints	Exhaustive	Treating all parts or aspects without omission; thorough	All constraints are accommodated in the designs or documented in the test plan for consideration during analysis
3: Execution Planning	3: Randomization	Exact	Strict adherence to standards or rules	Limitations to randomization are designed in as split plots or clearly annotated in the test plan for analytical accuracy
3: Execution Planning	4: Blocking	Exact	Strict adherence to standards or rules	Designs account for tests that will cover multiple days, crews, locations, etc. for analytical accuracy
3: Execution Planning	5: Replication	Exact	Strict adherence to standards or rules	Test points are replicated sufficiently to support statistical power goals
4: Analysis Planning	1: Analysis Goals	Meticulous	Showing or acting with extreme care and concern for details	Specific methods and outputs required to address objectives and requirements are clearly stated in the test plan
4: Analysis Planning	1: Analysis Goals	Exact	Accurate measurements	Estimate accuracy, confidence interval requirements, and margins are clearly stated for use with analysis
4: Analysis Planning	1: Analysis Goals	Exact	Strict adherence to standards or rules	Planned data analysis techniques do not violate inherent employment rules
4: Analysis Planning	2: Response Modeling	Exhaustive	Treating all parts or aspects without omission; thorough	Aliasing is minimized for modeling terms of interest
4: Analysis Planning	2: Response Modeling	Exact	Strict adherence to standards or rules	Design supports required number of terms for desired modeling fidelity (e.g. main effects, interactions, quadratics, etc.)
4: Analysis Planning	2: Response Modeling	Exhaustive	Treating all parts or aspects without omission; thorough	Design is sufficiently sized to investigate and uncover all terms of interest (e.g. 2 factor interactions)
4: Analysis Planning	2: Response Modeling	Exact	Accurate measurements	Use of validation points is included to demonstrate utility of modeled response
4: Analysis Planning	2: Response Modeling	Accurate	Deviating only slightly or within acceptable limits from a standard	Validation of model falls within stated confidence goals
4: Analysis Planning	2: Response Modeling	Exact	Strict adherence to standards or rules	Model validation methods follow accepted practices (e.g. numerical, visual, etc.) and are stated in the test plan