Evaluating Delivered Test Plans For STAT Content

Authored by: Michael Harman

30 March 2020

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

Executive Summary 2
Introduction 2
Application of the STAT Process to Test Plan Reviews 2
Critical Review Questions and Details 4
Conclusion 6
References 6

Executive Summary

Test teams must ensure they create events that generate the necessary and sufficient analysis to inform later tests and decisions. To be efficient and effective, test planning must include all relevant stakeholders so the outcome meets everyone's performance and analytical requirements and the STAT process should be an integral part of this team. Sometimes a test plan may be delivered without STAT input, or the STAT content may be unknown. In these cases, the STAT Expert needs to have a defined and repeatable review process. The questions presented in this paper are based on the STAT process and are listed in a specific order to ensure foundational planning concepts are addressed before dependent aspects (factor levels, design details, sample size, analytical methods) are considered.

Keywords: defense, rigor, developmental, contractor, deliverable

Introduction

Test events generate data to inform later tests and support decisions and the team must ensure they create the necessary and sufficient analysis to address these objectives. To be efficient and effective, the test planning team must include all relevant stakeholders so the selected test points adequately represent system performance and ensure analytical requirements are met. From the analytical perspective, the STAT process should be an integral part of planning from the beginning, ensuring rigor and minimizing lost time and re-work to modify shortfalls. In some cases, however, the STAT Expert may not be available to participate in test planning, potentially due to contractual or workload reasons, and they are forced to review the product after some level of completion. While not ideal, this situation is commonplace and the STAT Expert needs to have a defined and repeatable review process. The review process described in this paper is based on the STAT process and includes a set of questions to ask of the planning team throughout the course of the review. These questions are listed in a specific order to ensure foundational planning concepts are addressed before dependent aspects (factor levels, design details, sample size, analytical methods) are considered. The reviewer is able to employ these questions to generate recommendations for improving the test plan and to cite the level of risk with the various topics.

Application of the STAT Process to Test Plan Reviews

The STAT process is shown in Figure 1. Details covering the implementation for initial planning efforts are contained in the *Guide to Developing an Effective Test Strategy* (Burke, 2019). The reviewer should be completely familiar with this process to ensure the broadest ability to understand the proposed test design components, where they lie in the range of realistic possibilities, and their relative strength compared to other options.

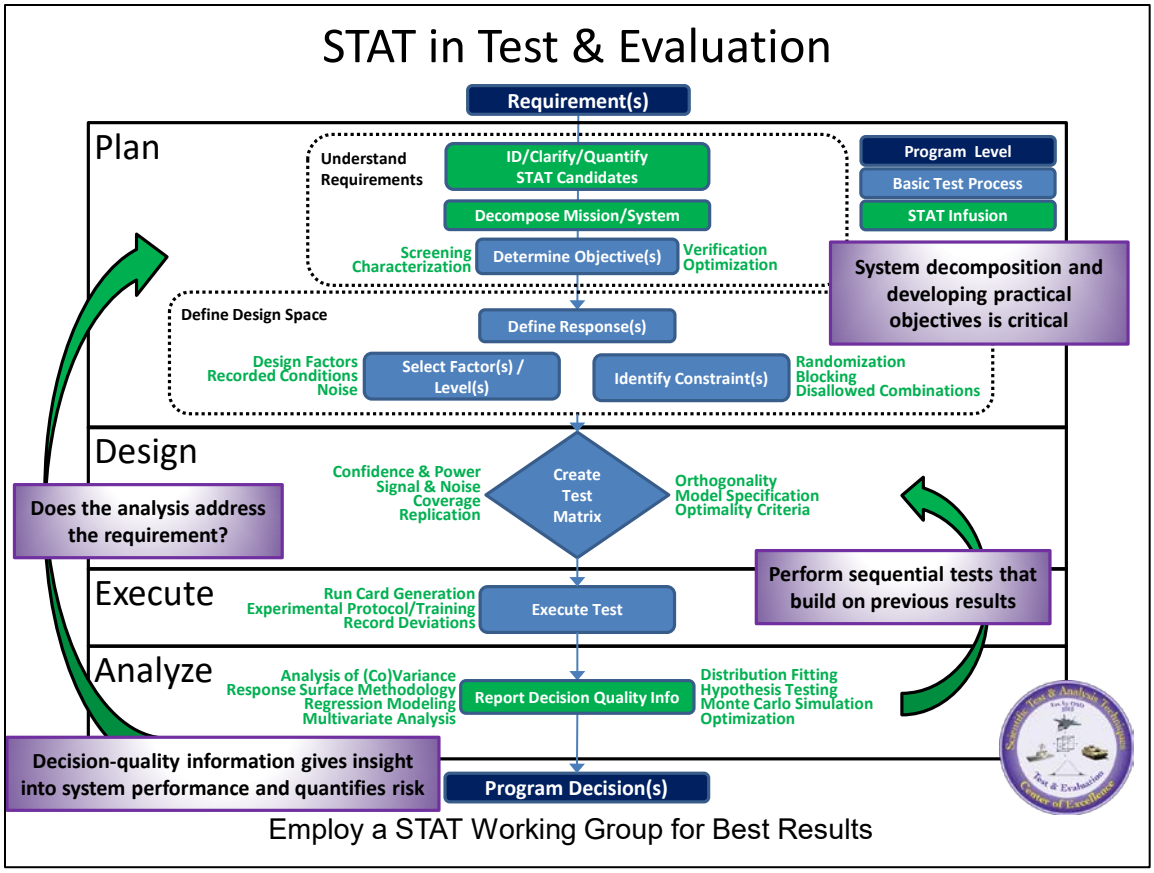

Figure 1: The STAT Process

More importantly, prior to beginning the assessment the reviewer must have some good ideas about what to expect. It is important to understand the most likely field of available choices/answers for a given question. This places the reviewer in a critical thinking mindset, ready to consider the proposed choice but able to potentially question why another selection was not made, or why a detail was not defined in another manner. Simply reading the proposal like a single source of information places the reviewer in a challenging "accept/reject" mindset, potentially lacking the perspective to ask more probing questions. For instance, the factor "Speed" might have levels defined categorically as slow, medium, and fast. Analytically speaking, this is better expressed as a continuous variable ranging from 10 to 30 mph unless a constraint or circumstance is citied that precludes that definition. In all, the reviewer must keep in mind that expectations must be tempered by context. The proposed method may not be in line with expectations but still may be sufficiently rigorous to justify its use. A well-documented proposal makes for an easier review.

To effectively review an existing test plan, a set of questions based on the STAT process can be employed for the best results. The process in this paper is a repeatable approach that can serve as a training resource to ensure consistency between reviewers. The goal of the review is to understand the described methods and ensure the design formulation is sufficient to address the stated objectives and to ensure that the objectives address the requirements. Also, risk areas can be uncovered and

appropriate recommendations made to improve the plan and decrease analytical risk. This implies the reviewer has approval authority so the deliverable will be updated and corrected based upon recommendations.

The STAT process puts the questions in an order that ensures the foundational details (decomposition, responses, factors, constraints) are in place and acceptable before reviewing the subsequent design dependencies (factor levels, design details, sample size, analytical methods). If a rigorous process was employed initially, these questions are easily answered and the perceived risk should be low.

Once the reviewer understands the proposed methods, the next step is to document in words what should be done to maintain or improve the test plan. This text contains the requisite detail but makes it difficult for a manager to understand where issues were uncovered. For this reason, a risk statement can be cited for each question: Low, Medium, and High. The risk can be broken into two components, (1) whether the approach meets "expectations" of the reviewer and (2) if the proposed approach is "suitable."

- "Expectations" include the definition of the object of the question, its perspective related to the objectives, and its placement in the range of potential options.
- "Suitability" includes the application of rigor and the analytical ability to satisfy objectives.

The combination of these two assessments results in a summary risk statement

- As per expectations and suitable approach: Low Risk
- ***Not*** as per expectations but suitable approach: Low Risk
- As per expectations but ***un***suitable approach: Medium Risk
- ***Not*** as per expectations and ***un***suitable approach: High Risk

Critical Review Questions and Details

The following questions can be used as part of a checklist to assess and report risk and recommended changes to improve rigor.

1. How has the system (or functions) been decomposed to facilitate lower level designs (to the lowest level required)?
    a. Does the complexity impact the ability to define test objectives or responses?
2. Is there traceability between the test activities, objectives, and requirements?
3. How does the strategy ensure early testing informs later testing?
    a. Is modeling and simulation (M&S) set up to inform hardware-in-the-loop (HWIL) or ground testing (GT)?
    b. Can live testing adequately validate M&S, HWIL, or GT results?
4. What is the source of rigor required?
    a. Examination, Demonstration, Analysis, or Testing?
    b. Will an inspection suffice?

   c. Is there an applicable military standard (MIL STD) or industry standard?
      i. Is the test method prescribed?
   d. Will a custom test design be required to address the uniqueness of this system?
      i. Potential examples: acceptance sampling, simple random sample, design of experiments (DOE), reliability study, observational study?
5. Are test objectives clearly articulated?
   a. Are they specific, unbiased, measurable, and of practical consequence?
   b. Is it stated as a testable question or with action verbs?
6. Response variables:
   a. How are they defined; pass fail, continuous measure, categorically?
   b. Will these address the test objective?
   c. Are they directly measurable?
7. Factors (if applicable):
   a. How are they defined; continuous, categorically?
   b. Do these support the test objective?
   c. Are they relevant to the stated response variable?
8. Factor control:
   a. Are factors identified as hard to change, held constant, recorded, or nuisance variables?
9. How do factor levels (number, value) support the test objective?
   a. Are there sufficient levels to model the desired effects?
   b. Does factor level spacing meet analytical leverage requirements?
10. Are all testing constraints and restrictions defined and documented?
   a. Examples: budget, range, safety, disallowed combinations
   b. Do these negatively impact assumptions made about the response or factors?
   c. Is there an alternate plan (such as M&S) to gather the information?
11. What is the rationale for the type of test design/strategy selected?
   a. Was the design selected for analytical, risk, acquisition, or other reasons?
   b. Does the design meet analytical output requirements?
12. What is the rationale for sample size/test time?
   a. Is there a statistical basis for the selected sample size?
   b. If not, is the rationale clearly stated and does it align with objectives?
   c. Could a more rigorous approach be employed?
   d. For a reliability test, what method was used to arrive at the test time and number of acceptable failures?
13. How do the location/organization of the test points support the test objective (if applicable)?
   a. Is the factor space well covered?
   b. Are considerations for orthogonality and correlation apparent in the design?
   c. If a constrained space design is employed, is information about aliasing included?
14. Are assumptions that impact the choice of test design documented?
15. Are assumptions in the analysis capability documented?
   a. For example, is previous/historical data required and available?

16. What is the rationale for the level of statistical merit associated with the selected design?
    a. What is the stated confidence level?
    b. Is power cited for the desired model terms and is it sufficient?
    c. For a design of experiments (DOE), is the orthogonality (DOE) sufficient for estimation of the effects?
        i. What signal to noise ratio (SNR) is cited and how was it determined?
17. How will the designs adequately explore the expected region of operability?
    a. What regions will not be testable and why were these left out?
18. What specific resources are required?
19. What is the planned analytical method?
    a. Analysis of variance (ANOVA)?
    b. Effects modeling (DOE)?
    c. Simple statistical distribution fitting?
    d. Graphical/exploratory data analysis?
    e. Decision trees?
20. Are modeling and simulation being used?
    a. If so, has M&S been accredited for use?
    b. If not accredited, how does this affect the analysis capability?
21. How will the analysis method help predict performance for points not actually tested?
    a. Will later testing be required to replicate earlier conditions for validation?
22. Is the analytical software defined or is custom coding going to be developed?
23. What is the format for data/products?
24. Can the test labs/ranges facilitate the fundamentals of STAT/DOE (randomization, replication, and blocking)?
25. How is test risk quantified?
    a. How sure will we be that observed/calculated results match true system performance?
26. If applicable, how would the design need to change to reduce the perceived risk?

Conclusion

Knowledge of the STAT process is a critical ingredient for any test plan reviewer. When STAT has not been integrated into the planning process, a post-delivery review can be an effective method to assess rigor and risk and provide positive feedback to improve the test plan. Employing questions based on the STAT process ensures foundational planning concepts are addressed before dependent aspects are considered. The reviewer is able to employ these questions to generate recommendations for improving the test plan and to cite the level of risk with the various topics.

References

Burke et al. “Guide to Developing an Effective Test Strategy V7.0.” Scientific Test and Analysis Techniques Center of Excellence (STAT COE), 31 Dec. 2019.