Scientific Test and Analysis Techniques Center of Excellence (STAT COE)

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.

What is STAT?

Scientific Test and Analysis Techniques (STAT) are deliberate, methodical processes and techniques that create traceability from requirements to analysis. STAT helps identify and quantify the risk associated with different experimental design choices and thus helps practitioners develop efficient, rigorous test strategies that will yield defensible results. The result of the repeated use of STAT is a more progressive sequential testing approach that carefully leverages past test information, along with informing the systems engineering process.

What is the STAT COE?

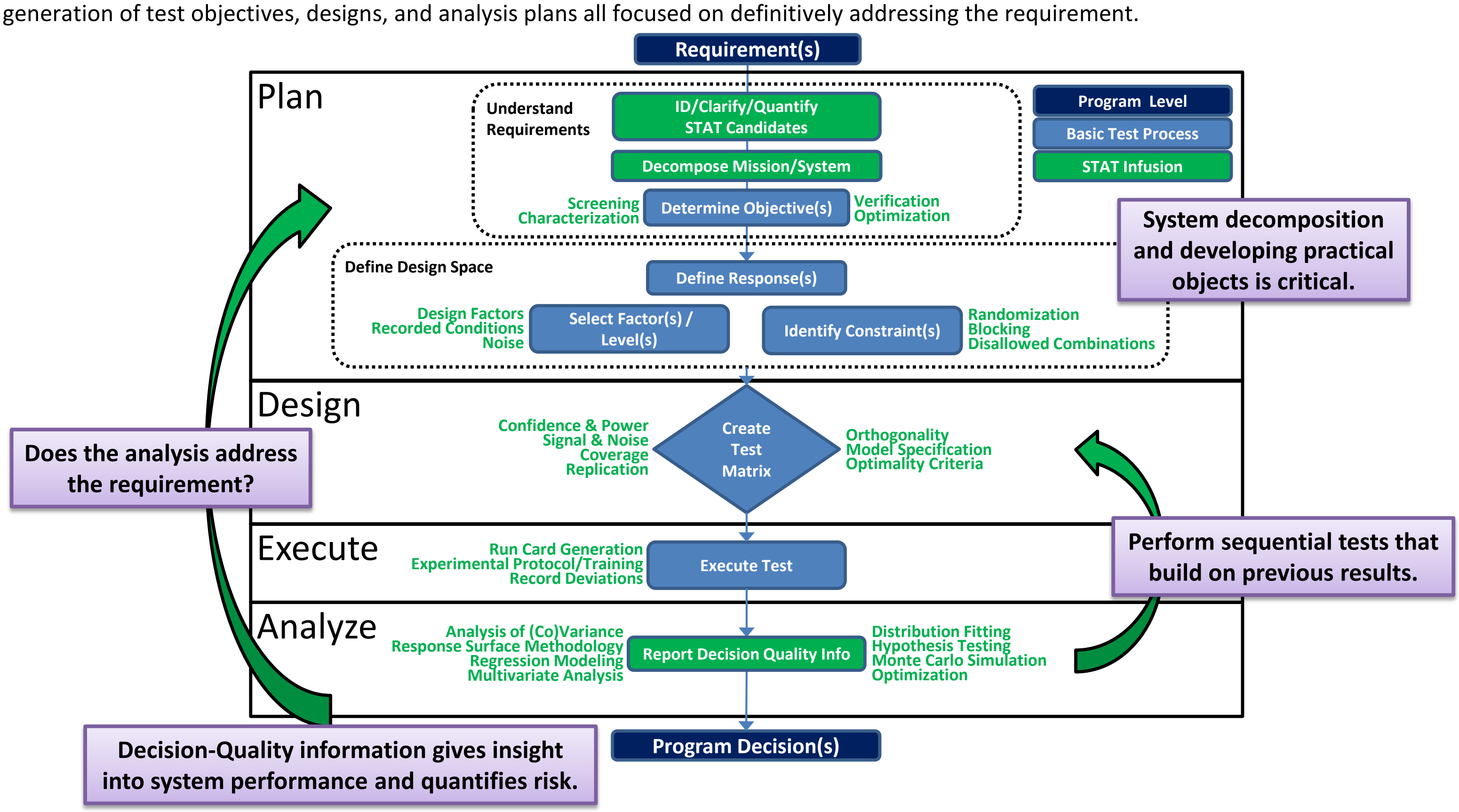
- In 2012, DASD(DT&E) collaborated with DOT&E and the service test components to create the Scientific Test and Analysis Techniques in Test and Evaluation Center of Excellence (STAT COE).
- The STAT COE directly supports program managers within major defense acquisition programs and automated information systems across the entire Department of Defense.
- The STAT COE provides recommendations and best practices regarding test planning, execution, and analysis, focused on optimizing return from test resources.
- The STAT COE is **independent** of DT&E and DOT&E involvement in support of the T&E WIPTs.

> "The Navy team used STAT COE resources to challenge their original Design of Experiment (DOE) assumptions resulting in the ability to reduce the number of non-firing runs while maintaining efficient and rigorous test designs that yielded defensible test results."
>
> ***James Geurts ASN(RD&A), January 23, 2019***

Contact the COE if you need assistance with a test plan, overall planning, or desire long term program support; CoE@AFIT.edu

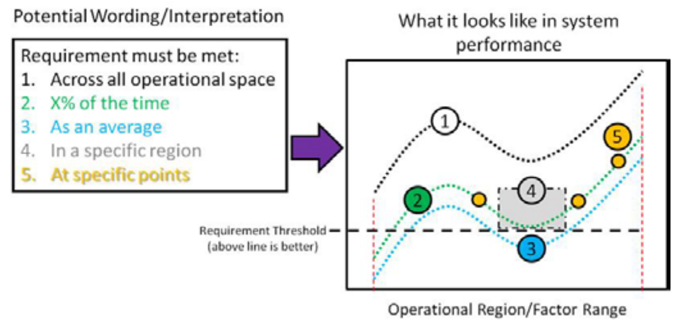
STAT Process Overview

The primary challenge in applying STAT to DOD testing is the broad scale and complexity of the systems, missions, and conditions. This challenge is best addressed by breaking down the requirement, system, and/or mission into more manageable pieces. At a lower level these pieces are more readily translated into statistical designs that provide rigor, efficiency, and definitive and quantifiable analysis. The STAT methodology is not a one-pass process. Rather, it describes an iterative process that begins with the requirement and proceeds through the generation of test objectives, designs, and analysis plans all focused on definitively addressing the requirement.



Understand the Requirement(s)

The information, purpose, and intent contained in the requirements drive the entire process. Requirements are often complex and/or ambiguous and typically not well defined for the application of STAT. The STAT working group will review all system requirements and identify all STAT applicable requirements. The working group will then subdivide the system or mission into smaller segments. Smaller segments make it easier to discern relevant test objectives, response variables, associated factors, and implied analysis.



Questions to ask:

1. To what part of the mission does this requirement pertain?
2. Are specific factors described?
3. What analysis is implied?

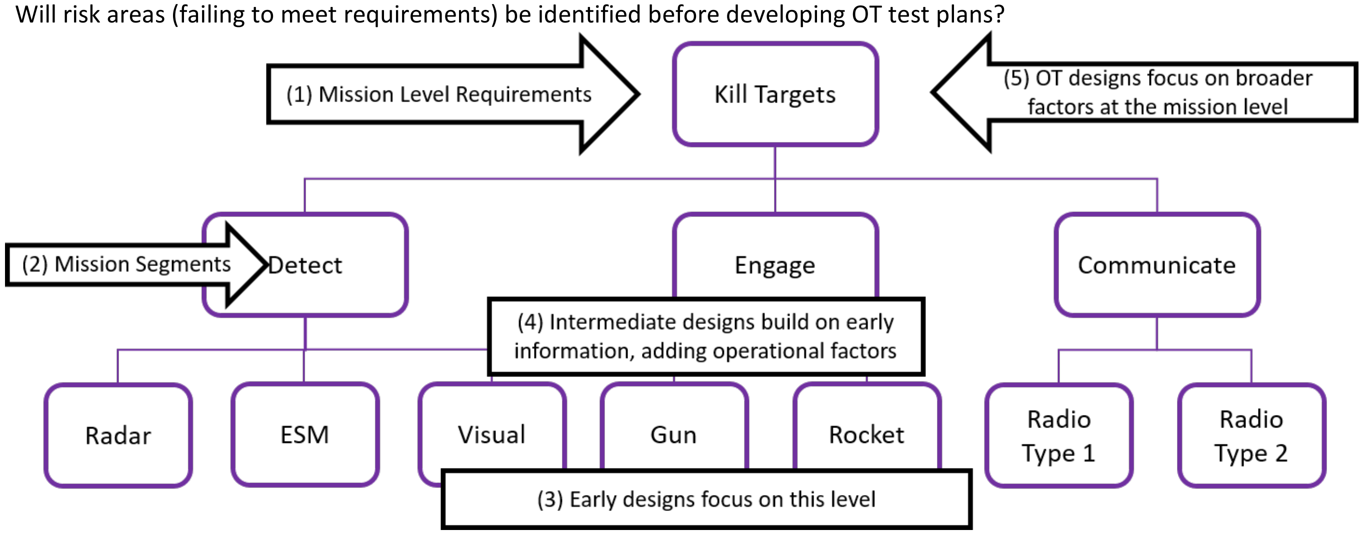
Takeaway: Question, understand, clarify, and document details in the requirements first.

Mission/System Decomposition

Understand what functions the system must perform to meet the objectives then the functions can be broken down into the components that are required to accomplish them.

Questions to ask:

1. Are all the steps necessary to meet the requirements included in the decomposition?
2. Will DT testing cover the entire range of operational conditions?
3. Will risk areas (failing to meet requirements) be identified before developing OT test plans?

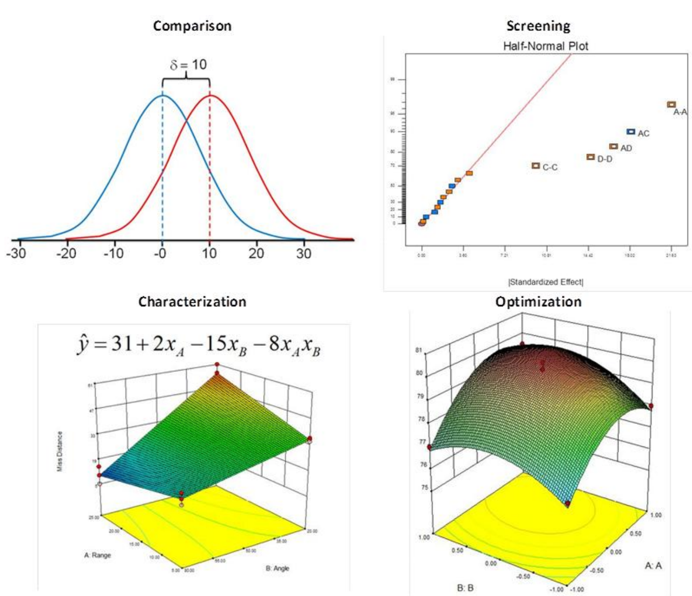


Takeaway: Decompose missions, systems, and functions until obvious and meaningful responses and factors are revealed; then develop a sequential test strategy that builds on previous testing as complexity increases.

Setting Test Objectives

The objectives are derived from the requirements and serve to focus resources and test designs toward generating sufficient information to address the requirement in a clear, quantitative, and unambiguous way. Common types of objectives:

- **Characterization**: Model performance across an operational envelope.
- **Screening**: Identify the key factors affecting performance.
- **Optimization**: Find settings of controllable factors that achieve ideal performance.
- **Comparison**: Evaluate two (or more) systems in an objective fashion across multiple conditions.



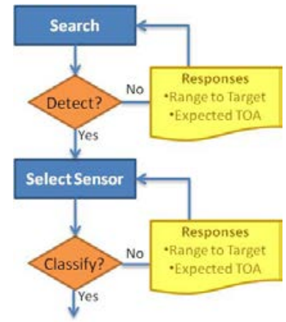
Questions to ask:

1. How does the requirement drive the focus of testing?
2. What is needed to address the requirement?
3. Can you state the objective clearly and concisely?

Takeaway: Objectives should be unbiased, specific, measurable, and of practical consequence (Coleman and Montgomery).

Responses

The key to good design is fully understanding the goals of the test and what responses should be collected. Responses are the measured output(s) (performance measures) of a test event. There may be several responses measured for a given test and/or in support of a requirement. Process maps are a good tool for determining the various performance measures (responses) that can be recorded.



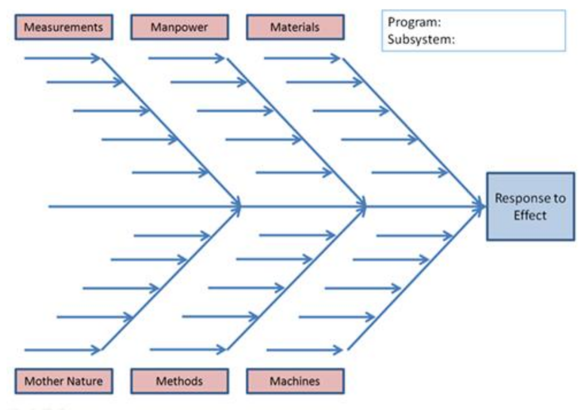
Questions to ask:

1. Does the response relate to specifications or top level requirements (e.g. through KPP, MOE)?
2. Is the response directly measureable?
3. Is the response defined in a clear and unambiguous manner?

Takeaway: The response must be clear, concise, measureable, preferably continuous, and directly related to the requirement.

Factors

Factors are inputs into or conditions for a test event and influence the variability in the response. Factors can be derived from prior testing, system knowledge, or insight into the underlying physics of the problem. Levels are the settings for each factor in a design. The test team will determine if a factor is to be controlled, held constant, or treated as noise. Each factor is also classified as easy, hard, or very hard to change. A fishbone diagram is a good tool for brainstorming potential factors of interest.



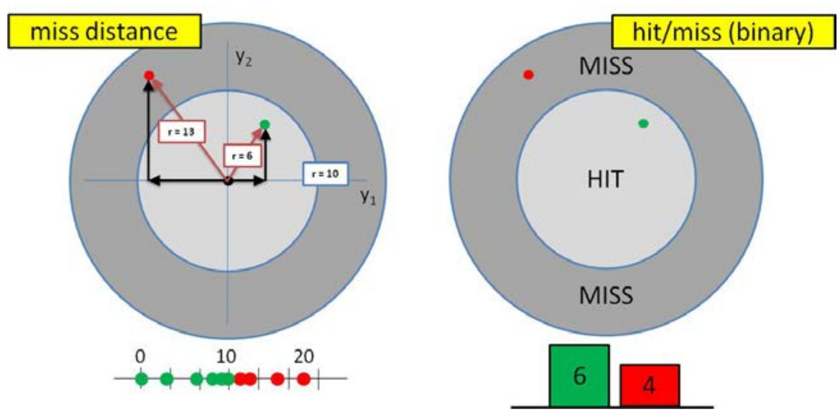
Questions to ask:

1. Have all factors been considered?
2. What are the factor priorities for testing?
3. Which factors will be controlled in testing?

Takeaway: Factors must be clear, concise, controllable, preferably continuous, and must relate directly to the response.

Data Type Consideration

The data type chosen to represent factors and responses in an experiment can have a major effect on the resources needed to conduct an experiment and the quality of its respective analysis. In the case of responses, categorical data types contain a relatively poor amount of information compared to continuous data types. Using categorical data types to describe factors may also have a detrimental effect on the overall size of the test and the quality of the analysis.



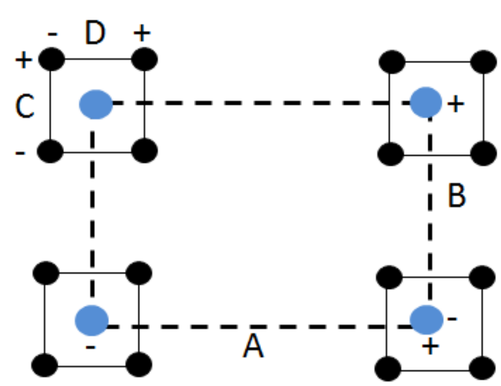
Questions to ask:

1. Is the factor or response a continuous variable?
2. What is the estimated signal-to-noise ratio for each of my responses?
3. Do I need to have an understanding of what is happening between categorical levels?

Takeaway: Continuous responses provide more information for analysis and typically require fewer runs.

Constraints

An important step in the test planning process is to identify any restrictions on the test design or execution. This identification is critical because any restrictions will influence the design choice and analysis. Common constraints include the budget, the experimental region, difficulty changing factor levels, and restrictions on randomization. If there are hard-to-change factors and the runs cannot be completely randomized, then we can create a split-plot design that accounts for this restriction on randomization.



Questions to ask:

1. What is the budget for this test?
2. Are there restrictions on the operational space or design region?
3. What factors are hard or costly to change?

Takeaway: Detail, define, and document how constraints will limit or impact the design or execution.

Design Choice

There are many considerations to balance when choosing between candidate designs:

- Test objective (characterization, screening, etc.)
- Underlying model supported
- Significant change in performance measure to be detected (δ)
- Confidence level
- Desired power
- Prediction quality
- Aliasing

Design #	1	2	3
# Factors	4	4	4
Levels	2	2	2
Model Supported	ME	ME, 2FI	ME, 2FI, Q
Signal to Noise Ratio	1.0	1.0	1.0
Alpha	0.05	0.05	0.05
# Center Points	0	4	5
# Repetitions	7	7	14
Total Runs	20	24	36
Power for ME @ S/N	0.63	0.54	0.95
Power for 2FI/Q @ S/N	0.54	0.53	0.87
FDS Pred Err @50%	0.63	0.58	0.36
FDS Pred Err @95%	0.90	0.75	0.45
VIF Avg	4.09	3.73	3.45
VIF Max	11.00	11.30	11.48
Confounding/Aliasing	med	med	low

These criteria should be used to compare multiple designs to evaluate and balance the risk and cost for a given design.

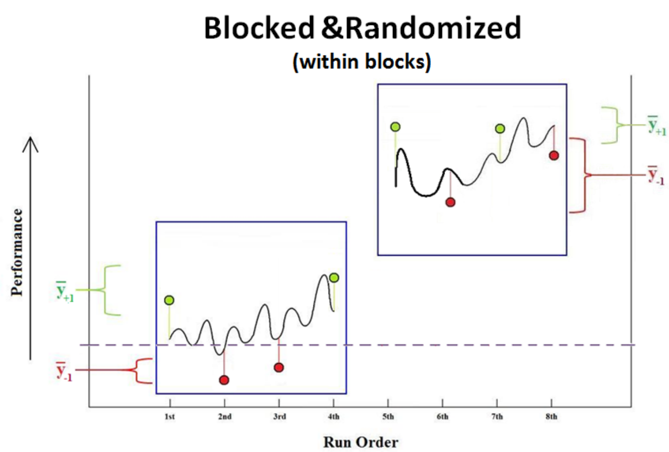
Questions to ask:

1. What type of predictive model can I build with this design?
2. Will the design clearly identify which factors are significant?
3. Which design will help me meet my test objective in the most efficient manner?

Takeaway: A design must be tailored to meet the test objectives and unique test circumstances.

Execution Planning

Some issues to consider when planning the execution of a design are randomization, replication, and local control of error (e.g. blocking). Ideally, runs in a test matrix are executed in random run order to protect against known, but uncontrollable, factors as well as unknown or "lurking" factors. If there are deviations from the planned protocol, these deviations should be recorded and appropriately dealt with in the analysis.



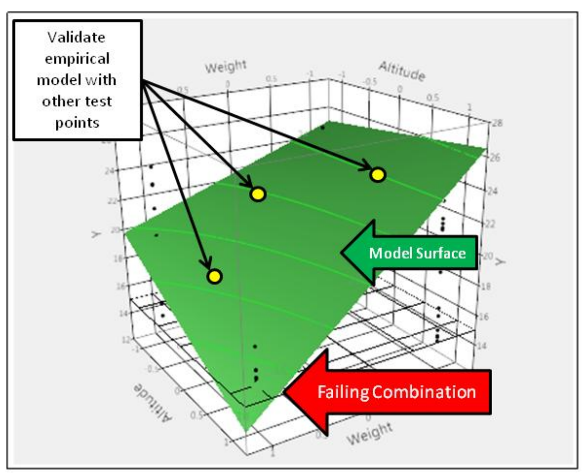
Questions to ask:

1. Will the design be executed in random order?
2. How will deviations from the test design plan be recorded and reported?
3. Are there known but controllable nuisance variables we wish to block out from the analysis?

Takeaway: Implement randomization, replication, and blocking whenever possible; record any deviations from the planned protocol.

Planning the Analysis: Will the Analysis Address the Requirement?

The requirement will guide us to the type of analysis we must perform. Additionally it is imperative to analyze the data as it was actually collected during the test, not the way it was planned to be collected. Analysis should facilitate and inform acquisition decisions.



Questions to ask:

1. What types of analysis methods are needed to provide decision-quality analysis?
2. Will the analysis answer the questions posed by the requirement?
3. Is the analysis appropriate for the data collection method?

Takeaway: Check that you are answering the right question with the analysis.

"To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of."

-Sir Ronald Fisher, Address to Indian Statistical Congress, 1938.