412TW-TIH-22-01

# TEST PLAN AUTHOR'S GUIDE

## Office of the Technical Director

4
1
2
T
W


**NOVEMBER 2022**

**TECHNICAL INFORMATION HANDBOOK**


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.
412TW-PA-22373


**412TH TEST WING**
**EDWARDS AIR FORCE BASE, CALIFORNIA**
**AIR FORCE MATERIEL COMMAND**
**UNITED STATES AIR FORCE**

This technical information handbook (412TW-TIH-22-01, *Test Plan Author's Guide*) replaces 412TW-TIH-20-01, *Test Plan Author's Guide* (Reference 1). This handbook was submitted by the Technical Director, 412th Test Wing, Edwards AFB, California 93524-6843.

Prepared by:

CARLSON.DANIEL.C.1394878843 Digitally signed by CARLSON.DANIEL.C.1394878843 Date: 2022.11.06 05:29:31 -08'00'

DANIEL CARLSON
AFTC Detachment 1/MDTF

LIEBMANN.CHRISTOPHER.J.1387297010 Digitally signed by LIEBMANN.CHRISTOPHER.J.1387297010 Date: 2022.11.04 16:35:19 -07'00'

CHRISTOPHER J. LIEBMANN
773rd Test Squadron

MAILEN.STEPHANIE.SHARP.1386619952 Digitally signed by MAILEN.STEPHANIE.SHARP.1386619952 Date: 2022.11.07 10:13:30 -08'00'

STEPHANIE S. MAILEN
445th Test Squadron

MARTIN.JEFFREY.D.1369066478 Digitally signed by MARTIN.JEFFREY.D.1369066478 Date: 2022.11.07 05:03:29 -08'00'

JEFFREY D. MARTIN
773rd Test Squadron

PECK.KEVIN.L.1250713017 Digitally signed by PECK.KEVIN.L.1250713017 Date: 2022.11.07 12:06:34 -08'00'

KEVIN L. PECK
775th Test Squadron

This handbook has been reviewed and is approved for publication: 16 November 2022

OSBURN.DANIEL.W.1241451239 Digitally signed by OSBURN.DANIEL.W.1241451239 Date: 2022.11.16 12:09:18 -08'00'

DANIEL W. OSBURN
Technical Director
412th Test Wing

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to the Department of Defense, Executive Service Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
**PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.**

| | | |
|---|---|---|
| **1. REPORT DATE** ***(DD-MM-YYYY)***<br>16-11-2022 | **2. REPORT TYPE**<br>Technical Information Handbook | **3. DATES COVERED** ***(FROM - THROUGH)***<br>N/A |
| **4. TITLE AND SUBTITLE**<br>Test Plan Author's Guide | | **5A. CONTRACT NUMBER**<br><br>**5B. GRANT NUMBER**<br><br>**5C. PROGRAM ELEMENT NUMBER** |
| **6. AUTHOR(S)**<br>Carlson, Daniel, Project Engineer<br>Liebmann, Christopher J., Project Engineer<br>Mailen, Stephanie S., Project Engineer<br>Martin, Jeffrey D., Project Engineer<br>Peck, Kevin L., Project Engineer | | **5D. PROJECT NUMBER**<br><br>**5E. TASK NUMBER**<br><br>**5F. WORK UNIT NUMBER** |
| **7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**<br>412th Test Wing<br>Edwards AFB, CA 93524-6843 | | **8. PERFORMING ORGANIZATION REPORT NUMBER**<br>412TW-TIH-22-01 |
| **9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**<br>412th Test Wing<br>307 East Popson Ave<br>Edwards AFB CA 93524 | | **10. SPONSOR/MONITOR'S ACRONYM(S)**<br><br>**11. SPONSOR/MONITOR'S REPORT NO.(S)** |

**12. DISTRIBUTION/AVAILABILITY STATEMENT**
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited. 412TW-PA-22373.

**13. SUPPLEMENTARY NOTES**
SC: 012100 (This number comes from STINFO via EAFB Technical Library)
CA: 412th Test Wing Edwards AFB CA — Print this document in COLOR.

**14. ABSTRACT**
This handbook provides guidelines promoting consistency across 412th Test Wing (412 TW) test plans. This handbook is a companion to the 412 TW test plan template. First-time authors should read this handbook before attempting to use the template, as this handbook contains basic test plan philosophy. The template contains critical content guidance for every section and element in the test plan.

**15. SUBJECT TERMS**
Reports; test plans; technical reports; handbook; technical writing; writing instruction manual; AFTC(Air Force Test Center); 412 TW(Test Wing)

| **16. SECURITY CLASSIFICATION OF:** | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19A. NAME OF RESPONSIBLE PERSON** |
|---|---|---|---|---|---|
| **A. REPORT** | **B. ABSTRACT** | **C. THIS PAGE** | | | |
| U | U | U | Same as Report | 38 | Daniel W. Osburn<br>Senior Technical Director, 412th Test Wing<br>**19B. TELEPHONE NUMBER (INCLUDE AREA CODE)**<br>661-275-4000 |

**STANDARD FORM 298 (REV. 8-98)**
**PRESCRIBED BY ANSI STD. Z39.18**

## TEST PLAN AUTHOR'S GUIDE CHANGE LOG

| Date (DD-MM-YYYY) | Document | Description of Significant Changes |
|---|---|---|
| 16-11-2022 | 412TW-TIH-22-01 | • Updated Overall Test Objective verb guidance<br>• Updated Rating Scales section<br>• Updated regression guidance<br>• Organized Test Planning Special Considerations section and moved to 2.0 Test and Evaluation<br>• Added information regarding experimentation testing<br>• Updated example formatting for consistency<br>• Moved instructional template to a separate document |
| 05-02-2021 | 412TW-TIH-20-01 | Full rewrite of 412TW-TIH-99-01 |

# TABLE OF CONTENTS

# INTRODUCTION

This handbook provides guidelines promoting consistency across 412th Test Wing (412 TW) test plans. This handbook is a companion to the 412 TW test plan template. First-time authors should read this handbook before attempting to use the template, as this handbook contains basic test plan philosophy. The template contains critical content guidance for every section and element in the test plan. For the current MS Word version of the template, consult your Technical Editor.

This handbook is intended to provide overall guidance that allows test plan authors and test teams the flexibility to tailor test plans to specific test programs. Engineering squadron-specific guidance and consultations with technical experts and the chief engineer should provide additional detail to complement this handbook. Although this guide and template are intended to promote consistency, variations may be desired in formatting or section organization (e.g., developmental test and evaluation [DT&E]/operational test and evaluation [OT&E] combined test plans or joint service test plans). Additionally, test teams may choose from a variety of media (e.g., documents, slide presentations). In those cases, this guide should still be consulted to ensure the inclusion of necessary content.

Authors assemble and use test plans to communicate the technical details and logistics required to execute and report results of flight, ground, and laboratory tests of air vehicles, subsystems, and components. Likewise, test plans provide context to 412 TW leadership, program office (PO) and test support personnel, and other testers, giving the reader a better understanding of the test objectives and methodologies for the system under test (SUT). Test teams should also keep in mind that approved test plans serve as archival documents for the capture of test and evaluation (T&E) enterprise knowledge for future test efforts.

This handbook is not intended to cover the entire test preparation and execution process. For those details, see EdwardsAFBI 99-101, *412 TW Test Plans* (Reference 2), EdwardsAFBI 99-105, *Test Control and Conduct* (Reference 3) and AFTCI 91-202, *AFTC Test Safety Review Policy* (Reference 4).

This handbook does not include guidance for test plan classification markings, which is covered by DoDM 5200.01, Volume 2, *DoD Information Security Program: Marking of Classified Information* (Reference 5). If a portion of test plan content is expected to be classified, test teams should consider whether to classify the entire test plan or to place all classified content in an appendix published under separate cover, leaving the bulk of the test plan unclassified.

## KEY CONCEPTS FOR SUCCESSFUL TEST PLAN WRITING

1. **Test Engineer Preparation is Critical** – Authors should understand the technical details of how the system is designed to work and the associated requirements for verifying the design. This preparation is essential for a test design that balances test support requirements, program risk, and defensible data that answer the right questions. By clearly and concisely detailing program requirements and methodologies in a well-organized test plan, the test team provides enough detail for another experienced test team to pick it up and execute the test.

2. **Get Involved Early** – Authors should get involved in the system development program early to help identify potential problems as soon as possible in order to save time and trouble in the long run. The earlier that testers get involved, the more impact they can provide for how best to plan testing the system, including: appropriate documentation, DT&E and OT&E involvement, determining test strategy, and early acquisition of long-lead-time assets. Test plan writing will go faster and smoother, avoiding interruptions and requiring fewer revisions.

3. **Leverage Prior Work Wisely, Balancing Efficiency with Innovation** – Test teams are rarely doing the first-ever test of its kind and are discouraged from reinventing the wheel. Authors should consider using previously vetted language (sources include 412 TW technical documents, flight manuals, etc.) to describe concepts rather than writing original words. New authors should consult experienced personnel for in-depth content and formatting suggestions. However, authors should not blindly copy and paste content, as even minor differences in the SUT or previous test methodology may not be appropriate for the current test effort.

   There is a healthy tension between efficiency and innovation. Uniformly formatted/structured test plans (program to program) provide familiarity and are therefore faster for authors to produce and stakeholders to consume. However, a novel format, plot, plan structure, medium, test objective, measure of performance (MOP), etc., may communicate content so much better that it's worth the extra time and effort. Innovation should be encouraged and not automatically rejected on the basis of unfamiliarity.

4. **Clearly Understand Customer Requirements** – Test requirements are derived from warfighter needs. Test scope, test objectives, and final deliverables must capture customer requirements with an agreed upon level of technical rigor, requiring coordination with all stakeholders (i.e., any organization including contractors/vendors that produce, consume, analyze, and/or report the test data). Often, initial customer requirements and expectations must be clarified to achieve an effective document. Example customer requirement sources include: Capabilities Development Document (CDD), Capability Productions Document, and Operations Requirement Document (ORD). Additionally, requirements may be generated by the test team and coordinated with the customer, such as system regression objectives and military utility assessments.

5. **Consider Both New and Legacy Capabilities** – New capabilities tend to grab the attention of the stakeholders and require in-depth scrutiny. However, sometimes the system changes that add new capabilities can impact previously existing (also known as legacy) capabilities. Test teams should understand the potential interactions among systems, and test the critical legacy capabilities (regression testing) that are most likely to be affected.

6. **Include and Maintain Traceability** – The test plan is critical to defining the test methodology and data requirements that support the conclusions and recommendations addressing customer requirements. System requirements drive test objectives from which MOPs can be developed. Testing generates data that are analyzed to derive conclusions and recommendations, which then answer test objectives and inform system requirements (Figure 1).

7. **Anticipate Final Deliverable(s)** – Well-written test plans serve as the foundation for successful test reports and are key to streamlining the test reporting process. Conceptualizing how test results will be presented can sharpen the test plan by identifying inconsistencies and revealing areas requiring more (or less) emphasis. The 412 TW standardized set of reporting options (per EdwardsAFBI 99-103, Reference 6) is summarized in Appendix A of 412TW-TIH-22-02, *Test Report Author's Guide* (Reference 7).

8. **Distinguish between Technical and Safety Requirements** – Test teams should distinguish between technical requirements and safety requirements to maintain a clear understanding of the reasons behind those requirements. Separating requirements ensures test teams prioritize safety over data during execution. These distinctions allow teams to properly provision a specific mission in the event individual items become unavailable. For example, if a safety-of-flight (SOF) parameter became unavailable during a mission, the aircraft would return to base, but if a required for data (RFD) parameter applicable to only one planned test point became unavailable, the mission

might continue to execute other lines of test. A combined test force (CTF) unit test safety officer (UTSO) can assist in determining which information is appropriate for the test plan vs. the safety plan.

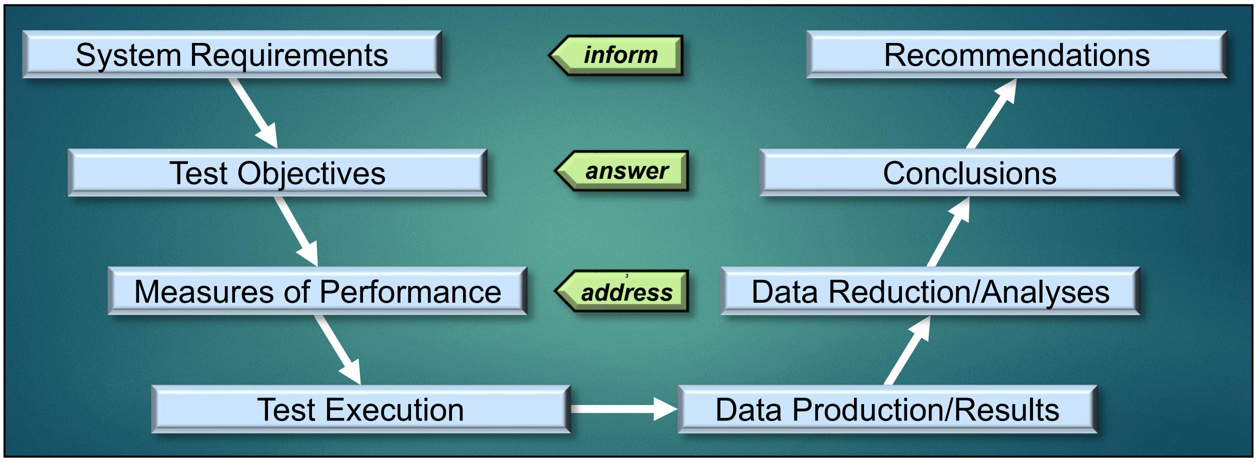



Figure 1 Requirements Traceability[1]

[1] Abbreviations, acronyms, and symbols in all figures and tables are defined in Appendix A.

# TEST PLAN CONTENT

Test plans consist of three major components: Front Matter, Main Body, and Appendices.

## FRONT MATTER

### Outside Front Cover:

As an official U.S. Government publication, the test plan and particularly its outside front cover should reflect the professionalism of 412 TW and the USAF. The cover is readers' first impression of the contents, and a photograph or professional illustration of the aircraft or SUT is encouraged. Cartoons or personalized logos are inappropriate. The outside front cover contains standardized information, including:

- **Document Number** – 412TW-TP-##-##; the number is assigned by 412 TW via the technical editor and the Technical Publications Office.
- **Title –** The title should be brief and descriptive of the test project. Including specifics in the title may be helpful in managing reader expectations, as well as quickly differentiating multiple test projects under a single program. For example, *SR-71 Block 40 Radar Performance Baseline Flight Test Plan* is a more informative title than *SR-71 Radar Test Plan*. Additionally, descriptive titles make documents easier to locate. If at all possible, keep the test plan title unclassified, regardless of the document's classification.
- **Authors –** Test plans typically include the project engineer and project pilot/operator/aircrew. If the aircrew is not a major contributor, they are not required to be on the cover. Personnel appearing on the cover may be government or support contractor (not prime).
- **Distribution Statement –** All test plans must have a distribution statement that has been selected by the test team and customer. The distribution statement will be IAW DoDD 5230.24, *Distribution Statements on Technical Documents* (Reference 8), as implemented by AFI 61-201, *Management of Scientific and Technical Information (STINFO)* (Reference 9). The determination date (month and year) of the distribution statement can refer to the statement of capability (SOC) date for the relevant test program, or another program date determined by the controlling authority. Technical editors may be consulted regarding squadron/CTF preferences. The controlling authority may vary, but is usually the PO. Further guidance may be found by searching 'distribution statements' at the Defense Technical Information Center (DTIC) website (Reference 10).
- **Document Control Marking** – All documents requiring control markings, such as Controlled Unclassified Information (CUI), must follow current markup guidelines. The test plan template reflects the most recent guidance, and the technical editor will help ensure the markup is complete and correctly formatted.

### Signature Page (Inside Front Cover):

The purpose of the signature page is to document who wrote the test plan and approved its publication. Major contributors must be able to support the key points of the document and will acknowledge their concurrence by signing the signature page.

The primary author and major contributors should be listed on the left-hand side, with any support contractors identified as such. The approval authority signatures on the right side are listed in EdwardsAFBI 99-101 (Reference 2).

## Standard Form (SF) 298:

The SF 298 is required (per DoDM 3200.14, Volume 1, *Principles and Operational Parameters of the DoD Scientific and Technical Information Program (STIP): General Processes* [Reference 11]) for all documents delivered to DTIC, except those classified higher than collateral secret. The test plan template will indicate which fields are required and their proper format.

## Qualified Requestors and Export Control Statements:

The qualified requestors and export control statements (standard and/or program specific) are required for documents not cleared for public release. The format and wording for these statements are governed by AFI 61-201 (Reference 9).

## Table of Contents:

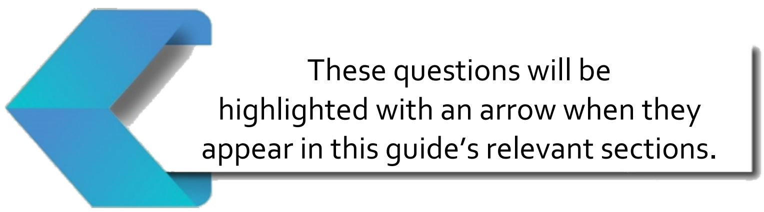
The table of contents is included in the test plan template.

# MAIN BODY

The main body of the test plan contains all of the elements that logically answer the following questions:

- **Section 1: Introduction**

**WHO** are the customers?
**WHO** will conduct the tests?
**WHEN** will the tests be conducted?
**WHAT** is the overall test objective?
**WHY** are the tests being conducted?
**WHAT** will be tested?
**WHERE** will the tests be conducted?




- **Section 2: T&E**

**WHAT** are the objectives of the tests?
**WHAT** could interfere with meeting test objectives?
**WHAT** is being measured and how? MOPs?
**WHEN** is testing finished?
**WHAT** are the evaluation criteria for each MOP?
**WHAT** is the test approach?
**WHAT** are the most realistic test outcomes?
**HOW** will the data be analyzed?
**WHAT** data products will be reported?

- **Section 3: Test Conduct**

**WHEN** is the team ready to test?
**HOW** exactly will the tests be conducted?

- **Section 4: Test Reporting**

**HOW** will test results be reported?

**Section Numbering:**

Test plan sections may be indicated with numbered or non-numbered headers. Consistency should be maintained either way. If numbering is used, the format used in this guide is recommended; if non-numbered headers are used, the overall structure and formatting should align with planned reports of results.

## 1.0 INTRODUCTION

The Introduction section provides an overview of the test project and includes test scope, background, resources, and test objectives.

### 1.1 Overview

**WHO** are the customers?
**WHO** will conduct the tests?
**WHEN** will the tests be conducted?

This section contains standardized content for easy readability. Generally, a single paragraph is used to convey the basic programmatic context of the test plan. The following elements should be included:

- Standardized introduction sentence
- The overall test objective
- The customer(s)
- Test organization(s), as appropriate. More information on these organizational definitions is found in DoDI 5000.89_DAFI 99-103, *Capabilities-Based Test and Evaluation* (Reference 12):
  - Lead developmental test and evaluation organization
  - Executing test organization
  - Participating test organization(s)
  - Operational test organization(s)
- Test stakeholders
- Test location(s)
- Approximate test date(s) (test teams should avoid using concrete dates, given the dynamics of program scheduling)
- Test scope (number of planned ground/flight test hours, test points, or other applicable metric[s]).

**Overall Test Objective Guidance:**

**WHAT** is the overall test objective?

Test projects contain a single overall test objective, which may be further divided into two or more general test objectives (GTOs), such as by mission areas (by discipline) or by scope. For small-scope or single-discipline test programs, GTOs may be omitted. Specific test objectives (STOs) are detailed objectives that describe the focus areas of the overall test objective or a GTO. Test objective hierarchy examples are shown in Figure 2.

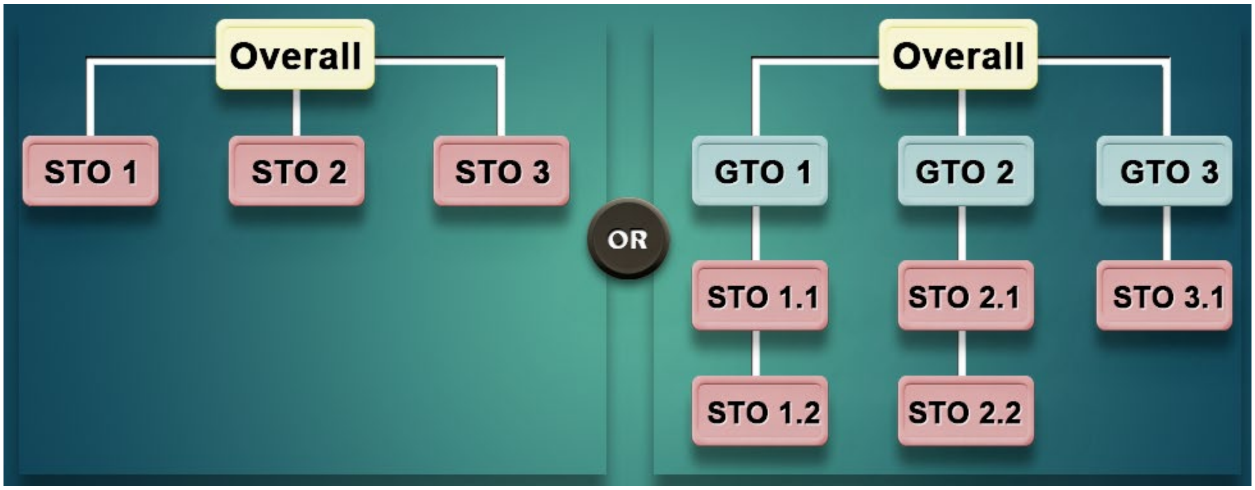



Figure 2  Test Objective Hierarchy Examples

The overall test objective should communicate the purpose of the test clearly and concisely using a uniquely defined test objective verb (Example 1). The overall verb should most clearly capture the scope of the test effort (Table 1). For example, if a test plan contained several specific test objectives to *evaluate* aspects of the SUT, an overall test objective to *collect* data on the SUT does not accurately reflect the scope of the evaluation and would not be appropriate. The following three statements provide a framework for the overall test objective.

1. Evaluate/Demonstrate/Determine/Characterize the SUT to provide ratings and/or a recommendation.
2. Evaluate/Demonstrate/Determine/Characterize the SUT to show characteristics, performance, or functionality.
3. Collect data on the SUT in support of analysis efforts by external organization(s).

**EXAMPLE 1**

Evaluate the integration of the map radar mode with the F-16 Block 99 rehosted software to provide ratings and a recommendation as to whether to proceed to operational testing.

★★★

Characterize the F-16 Block 99 rehosted map radar mode performance in support of modeling and simulation development.

★★★

Collect F-16 Block 99 rehosted map radar resolution data for the contractor in support of system development.

These statements maintain flexibility in addressing the variety of tests conducted by the 412 TW, including providing recommendations without ratings. Test plan authors should adjust the wording to provide the appropriate test project details. If a major portion of testing will include regression, the overall test objective should reflect that (see the Regression Testing section).

Table 1  412 TW Test Objective Verbs

| Qualities | Verbs: Collect | Demonstrate | Determine or Characterize | Evaluate* |
|---|---|---|---|---|
| Definition | To gather data for an external organization. | To show system functionality or performance. | To measure or ascertain system attributes. | To assess system effectiveness or capability. |
| Scope of Test and Evaluation | No 412 TW analysis or evaluation is performed. | Implies limited testing. May involve technology demonstrators, verification of fixes, or regression. | Involves establishing a system performance baseline. Comparison against a specification/standard may be appropriate. | Implies robust testing. System performance is usually tied to effectiveness and military utility. |
| Probable Test Outcomes | N/A | Generally well understood. System functionality or performance is typically binary; works as expected or does not. | Generally less well understood; test results are typically not binary. | Generally well understood; test results are typically not binary. |
| Use of Rating Scales | May use the 412 TW Test Completion Scale to assess the sufficiency of collected data. | Rating scales may or may not be used, based on system maturity and extent of testing. Qualitative assessment may be used to describe test results. | May use the 412 TW Rating Scale or 412 TW Test Completion Scale to assess test objectives and provide an overall rating. May use discipline-specific rating scales to establish system attributes†. | Uses the 412 TW Rating Scale and/or 412 TW Regression Scale to assess test objectives and provide an overall rating. May use discipline-specific rating to support the overall rating. |
| Emphasis on Military Utility | N/A | Results may include some emphasis on military utility, but general focus is on system functionality or performance. | Results may include some emphasis on military utility, but general focus is on system attributes. | Results should include a strong emphasis on military utility when appropriate. |

* The nature of experimentation is centered primarily on collecting data and determining/characterizing unknowns; therefore, "Evaluate" is rarely appropriate for experimentation testing.

† Discipline-specific rating scales may include the Cooper-Harper Rating Scale, 412 TW Revised Bedford Workload Scale, and general purpose scales.

## 1.2 Background

**WHY** are the tests being conducted?

The Background section should provide a summary of relevant program history leading up to the test, to include: why this test is being accomplished, any previous related test efforts and significant results, and problems found during operational use. These details should explain how this particular test fits into the broader scope of the platform/test program/enterprise. The technical maturity of the SUT, to include any test entrance criteria (e.g., modeling and simulation [M&S]) performed to prepare for this test, may be explained. However, any fine details of the SUT should be reserved for the Test Item Description section.

The Background section may also introduce technical concepts important in understanding aspects or methodologies used later in the test plan. Discussion of these concepts should be kept at a high level in this section, with more detailed descriptions placed in an appendix as necessary.

## 1.3 Test Item Description

**WHAT** will be tested?

The test item description should provide enough details to understand the SUT, including any relevant information that impacts test design. If the SUT is part of a larger system on the aircraft, this section should focus on the parts of the system that are new or test-unique, and should differentiate the SUT from all of the supporting equipment. Functional control diagrams (with an outline around the SUT) help make the distinction clear. If the SUT is intangible, such as an algorithm, the description should focus on the algorithm rather than the hardware supporting that algorithm. Identification of the host system or aircraft (not test unique) should be captured in section 1.4.3 Test System/Aircraft. If M&S resources are used as the system under test, they should be described in this section. Assumptions and/or system modifications made to facilitate the simulation, along with why/how that simulator is an appropriate test venue, should also be included. Generally, the Test Item Description section should not exceed two pages; lengthier descriptions should be provided in an appendix and summarized in this section.

## 1.4 Test Resource Requirements

The resource requirements for the test program should be identified in this section. Teams should list the resources, elaborating on any whose purpose is not obvious. This can include a wide variety of needs, ranging from special test equipment and analysis tools to outside range or technical support, such as system integration laboratories (SILs), hardware-in-the-loop (HITL) facilities, installed system test facility (ISTF), etc. Common resource requirements sections include but are not limited to the following elements.

### 1.4.1 Modeling and Simulation Resources

This section states which M&S resources will be used during the test, including their maturity or known accuracy. The resources can be contractor or government owned, and are often used to predict test results, establish system maturity, provide understanding of system behavior, augment or serve as the primary venue for test results, and train personnel.

### 1.4.2 Test Facilities, Ranges, and Resources

**WHERE** will the tests be conducted?

This section outlines the contractor- or government-owned test ranges, airspaces, airfields, facilities and their associated resources. Table 2 provides examples of test facilities, ranges, and resources, and their types and descriptions. Engineering squadrons may provide additional information.

Table 2 Examples of Test Facilities, Ranges, and Resources

| Type | Description | Example Facility/Range(s) | Example Resources |
|---|---|---|---|
| Virtual Environment | Facilities with computer models of the system under test, friendly/non-friendly players, scenarios, combat environment, and threat systems used to replace or supplement on-aircraft test | • Digital Integrated Air Defense System (DIADS), Edwards AFB<br>• Integrated Facility for Avionics Systems Testing (IFAST), Edwards AFB<br>• Joint Simulation Environment (JSE), Edwards AFB | • Threat laydowns<br>• Scenario laydowns |
| Measurement Facilities | Facilities with capabilities to establish known quantities of the SUT (e.g., mass properties). | • Metrology Facilities<br>• Stores Weight and Inertial System Facility | • Scales<br>• Other measurement equipment (especially if brought from off-site) |
| SILs/ HITL Facilities | Facilities designed to integrate aggregations of hardware and software in a laboratory environment. | • Integrated Defense Avionics Lab (IDAL)<br>• Handling Qualities Simulator (HQS) | • Special Test Equipment (STE)<br>• Line-Replaceable Units (LRUs) |
| ISTFs | Facilities designed to evaluate integrated systems in installed configurations to test specific functions of complete, full-scale weapons systems. | • Benefield Anechoic Facility (BAF), Edwards AFB<br>• McKinley Climatic Laboratory, Eglin AFB, Florida | • Surrogate signal sources<br>• Threat simulators<br>• Avionic simulators |
| Open-Air | Facilities that provide the ability to evaluate the systems under natural environment operating conditions. | • Precision Impact Range Area (PIRA), Edwards AFB<br>• Point Mugu Sea Range (PMSR), California<br>• Naval Base Ventura County (NBVC), California<br>• White Sands Test Center (WSTC), New Mexico | • Inter-range links<br>• Threats/threat simulators<br>• Airspace (R-2508, etc.) |
| Other Resources | Additional facilities and equipment required for the test. | • Ridley Mission Control Center (RMCC), Edwards AFB<br>• Hangar 1600, Edwards AFB | • Control room<br>• Telemetry<br>• Tracking radars<br>• Hangars or ramps<br>• Drop pit<br>• Trucks and cranes |

### 1.4.3 Test System/Aircraft

If the aircraft is the SUT, this section may be deleted, as the information is already in the Test Item Description section.

This section states the test aircraft and associated test configuration requirements. Aircraft configurations may include software, hardware, and/or stores. Any flight certification requirements that allow the SUT to be installed and operated on the host system should be included (e.g., Temporary 2 [T-2] aircraft and any applicable modifications, military flight releases, contractor aircraft/engine operating limitations, contractor-owned/contractor-operated contracts, or PO configuration control boards).

### 1.4.4 Instrumentation and Parameter Requirements

This section states the instrumentation required on the SUT, onboard the test/support aircraft, and on the range(s). It also defines the recorded data parameters (also known as test measurands) produced by those instrumentation systems. Some data parameters are available via a data bus (often in MIL-STD-1553B format) and some are available via special instrumentation (often known as orange wire). The parameter list should address both data bus and orange wire parameters. Lengthy parameter lists should be provided in an appendix.

The RFD parameters will typically be verified as operable before each test mission. Usually, RFD parameters need not be telemetered and monitored in real time unless they are also safety-of-test (SOT) or SOF parameters. Generally, the failure of any non-SOT/SOF RFD parameter would cause a pause until the responsible engineer or test team can determine whether testing may proceed without the parameter.

The SOT/SOF parameters are those essential for ensuring the safety of a test or flight. The SOT parameters must be monitored in real time against established limits during the execution of test points. The SOF parameters must be monitored during the entire flight including between test points. Generally, the test/flight will not proceed if any SOT/SOF parameters are unavailable. If the SOT/SOF parameters are listed here or in an appendix, they may be referenced by the safety plan, rather than repeated.

### 1.4.5 Support Vehicles/Aircraft

This section briefly states support vehicle and/or aircraft requirements. If appropriate, documents containing detailed support vehicle/aircraft descriptions may be listed as references. The focus should be on the technical requirements needed to accomplish the test objectives, such as aircraft with a desired radar cross section. Safety requirements such as safety chase should be identified in the safety plan.

## 1.5 Safety Considerations

All of the safety requirements should be clearly stated in the safety plan. Safety considerations that also affect the technical approach, such as resource requirements or specific process/execution considerations, may be noted here.

## 1.6 Security Requirements

This section should inform the reader which security measures (general, operations, communications, and competition sensitivity) will be required before, during, and after the test. Guidance is included in the test plan template.

### 1.7 Key Stakeholder Contact Info

The purpose of this section is to provide the contact information of the personnel and/or offices with responsibilities essential to test execution. This section is not meant to be a comprehensive list of all test team participants.

### 1.8 Test Environment Requirements

This section should describe requirements pertaining to locations, times of day, weather, etc., required for testing, including any technical limits.

### 1.9 Environmental Impact Assessment

Federal and state environmental laws regarding air pollution, noise pollution, waste disposal, disturbing the ground in drop zones, fuel spills, wildlife, etc., must be followed when planning a test. For tests conducted at or by the 412 TW, the test team will coordinate with the 412 TW Environmental Management Office at the beginning of the test project to produce an approved environmental checklist, usually as an appendix. This checklist and a statement about the assessment must be included in the test plan, regardless of whether or not environmental impacts are expected. The 412 TW Environmental Management Office may be contacted at: (661) 277-1401 or 412TW.CEV.EIAP@us.af.mil.

## 2.0 Test and Evaluation

The T&E section is the technical core of the test plan. Authors may add paragraphs or subsections at the beginning of Section 2 to provide readers with contextual information regarding definition of test terms, test phases, technical build-up approach, etc.

**Test Planning Special Considerations:**

Statistical methods, previous test approaches, regression, M&S, military utility, and experimentation are special considerations that should be understood before writing the T&E section.

**Statistical Methods.**

Test teams should consider the use of statistical analyses when developing the overall test approach in Section 2. Such methods are of greatest value when outcomes are uncertain and resources are severely constrained. The important components of a statistical test include: the estimated noise (uncertainty) expected in the data, the desired signal (effect) the test is designed to detect, and the desired level of precision (e.g., confidence and power). These statistical components serve to guard against system mischaracterization and/or inaccurate system ratings, and are typically displayed in a table and/or graph to inform decision makers. Authors should utilize statisticians local to the CTF and/or from the Statistics Home Office to help identify whether statistics are appropriate, and, if so, to help formulate correct tests and prepare required deliverables. A statement regarding the use of statistics should be included in an external document if the use of statistics is not referenced in the test plan (see Section 3.2 of EdwardsAFBI 99-101 [Reference 2]).

**Previous Test Approaches.**

If the test is assessing the next iteration in a series of system developments (such as subsequent software blocks), authors should consider using previous test approaches to help design the current test. Consistency is important to assess system performance over time. If the team is not confident in previous test approaches, comparing the SUT with past performance may not be appropriate. Authors should take precautions to review the technical report (TR) and lessons learned, especially if the test did not gather the intended data or if the previous results were less than satisfactory.

### Regression Testing.

As developments or additions are made to previously tested or fielded systems such as hardware upgrades or subsequent software blocks, both new and legacy capabilities and/or performance will generally be tested. Regression testing assesses the impacts of a new capability integration on legacy functionality by comparing the current system operation with a baseline, or previous system functionality. Regression testing should focus on identifying impacts to legacy capabilities, rather than repeating a complete system evaluation as though it were a new capability. The scope of regression testing is generally based on design considerations (such as safety and operational relevance), previous deficiencies, customer requirements, and engineering judgment, and is prioritized according to funding, scheduling, and test program priorities.

When planning regression testing, test teams should consider how the changed system integrates with legacy unchanged systems. Systems that interact with or may be affected by the changed system should be considered adjacent systems, and should be identified and included in regression testing. Adjacent systems that have safety implications or are necessary for mission completion are generally considered to be higher priority for regression testing. For instance, if a change is made to the high angle-of-attack logic in a flight control system, the main out-of-control-recovery-logic paths in the flight control software would typically be regression tested, as those paths are adjacent to the change and have safety implications. In another example, a change to an aircraft's GPS/inertial navigation system might warrant regression tests of weapons cueing because software errors could result in incorrect cueing angles being passed to a weapon.

Regression testing assumes no measurable/detected change in performance exists until proven otherwise. The absence of measurable changes can be expressed as the null hypothesis in a statistical inference test. The alternative hypothesis would then state that a measurable change in performance exists, and the point estimate of the relevant metric would be bounded with a statistical interval.

Regression test approach should be specifically coordinated with relevant 412 TW technical experts and system operators (usually aircrew), and should consider the following:

- Extent of the intended changes (e.g., adding a new symbol to a display vs. rehosting an entire operational flight program in a new coding language).
- Likelihood that unintended effects might occur elsewhere in the larger system (e.g., changing a datalink system resulted in degradation in the air collision avoidance cues).
- Criticality of the modified capability (e.g., a seldom-used radio mode vs. flight control software).
- System complexity.
- Amount of overall system change since a capability was last tested.
- Programmatics (e.g., time/funding available, tolerance of technical risk).

When planning regression testing, test teams should carefully consider how to structure regression test objectives and/or MOPs to make test reporting easier. Separating regression test objectives and/or MOPs from new capability test objectives and/or MOPs generally allows for a cleaner presentation of test results in the test report as well as an easier roll up of test results to higher-level ratings.

Regression test results will be assessed as improved, unchanged, impacted, or degraded from the baseline IAW the 412 TW Regression Scale (see the 412 TW Rating Scales section). Because regression test results are a comparison to baseline performance or functionality, it is important for the test team to understand the original ratings of the baseline performance or functionality in order to accurately status the regression test results. For example, when test results are unchanged, but the original performance was less than satisfactory, the original rating should be referenced in the text describing the regression results to clarify that the performance is unchanged, but still unsatisfactory (or marginal).

### Modeling and Simulation.

The M&S may be used in lieu of full ground and/or flight testing in certain situations; for example, if a real-world asset is not available to be tested against, or a very large set of test points is required to fully vet a system, a simulation may be appropriate. If testing uses M&S assets as either a significant portion of the system under test or as a significant portion of the testing environment (e.g., Joint Simulation Environment [JSE] or Digital Integrated Air Defense System [DIADS]), the test objectives and MOPs should include the M&S information.

Test teams should ensure that tests have real-world applicability. During test planning, the use of M&S assets should be examined with the relevant discipline and platform technical experts. When planning how to test a system using M&S, the following should be considered:

- Which test points are best suited to simulation, and why
- Model validation, verification, and accreditation
- Whether similar testing has previously been done using simulation

### Military Utility Assessment.

Test teams should consider the system value to the warfighter, commonly called a military utility assessment, which can provide a critical early look to subsequent OT&E and decision makers regarding how the system will perform in operationally representative scenarios. The specific approach to military utility testing should be based on the method of system use by the operational end user, as well as the priority of the system to eventual operational usage. The usability, workload, and functionality of the system are all potential considerations, as are mission planning and ground support systems. The military utility assessment often requires coordination with system operators (usually aircrew). Any discoveries that affect the military utility of the system will be included in any final report(s) of results, and noting this in the test plan can help maintain focus on warfighter impact throughout testing. If a major portion of testing will include a military utility assessment, the overall test objective should be written accordingly.

### Experimentation.

An experiment is an activity that is pursued to explore the potential of newly available technologies coupled with alternative warfighting concepts to inform follow-on acquisition and employment decisions. Experimentation puts more emphasis on military utility than traditional research. Unlike T&E, experimentation does not set out to confirm achievement of a specification, performance level, or operational effectiveness and suitability. In experimentation, the answer to the question posed is not well predicted either through theoretical hypothesis development or design-based modeling and simulation. Experiments typically tolerate greater risks in their conduct except in the area of personnel safety.

Experimentation tests may not be rigorously based in requirements the way formal acquisition programs are. High fidelity requirements traceability may not be possible. As such, the GTOs, STOs, and MOPs should be generalized and high level enough to accommodate exploration.

### 2.1 General and Specific Test Objectives

**WHAT** are the objectives of the tests?

The GTOs and STOs should be short definitive statements beginning with an action verb followed by the object or qualifying phrases (Example 2). The action verbs are intended to be single-word summaries of the scope and intent of the test; consistent use of verbs across the 412 TW helps guide discussion and common understanding among test teams. Table 1 is intended as a guide for most cases, but is not prescriptive; alternate phrasing may be appropriate, provided the team agrees on the scope and intent of the test.

The STOs should have traceability to requirements (or previous results for regression testing) to the maximum extent possible. Although military utility testing typically is not traceable to a specific requirement, DoDI 5200.02T, Enclosure 4, *Operation of the Defense Acquisition System* (Reference 13), directs that DT&E include stressing the system in an operationally relevant environment, as well as identifying capabilities, limitations, and deficiencies, meaning that assessing military utility is a function of DT&E, even when not directly tied to a contractor- or program-provided requirement. Depending on the planned scope of regression and military utility assessment, teams may write dedicated GTOs/STOs or MOPs, or the information may be captured in already existing GTOs/STOs/MOPs.

**EXAMPLE 2**

**GTO 1** – Evaluate aircraft aero-performance with AIM-9X installed.
- **STO 1.1:** Evaluate up-and-away aero-performance.
- **STO 1.2:** Evaluate powered-approach aero-performance.

★★★

**GTO 2** – Demonstrate communication link functionality.
- **STO 2.1:** Demonstrate link command and control.
- **STO 2.2:** Demonstrate link switching.
- **STO 2.3:** Demonstrate link availability.
- **STO 2.4:** Demonstrate link latency.

### 2.2 Potential Impacts to Completion Criteria

**WHAT** could interfere with meeting test objectives?

This section describes any test unique factors that could realistically interfere with meeting test objectives and become a limitation or constraint in the technical report. Common potential impacts are shown in Example 3. Teams should scope tests within the programmatic bounds provided by the customer, rather than designing an inflated test scope which will inevitably not meet the completion criteria for easily anticipated cost or schedule reasons. However, if the technical review authority determines the limited programmatic bounds and determines the test, as described in the test plan, is too limited (e.g., insufficient sample size, or too few missions) and therefore not technically adequate, that finding should be noted in the technical review memorandum.

## EXAMPLE 3

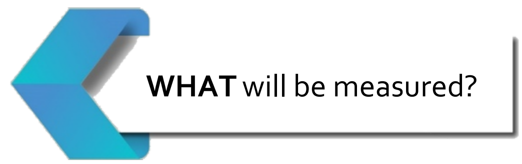
**System Under Test** – The F-16 T/N 123 requires a small amount of trim to fly straight, but it is the only F-16 available for the test. Although a technical risk, the limited trim is unlikely to prevent test objectives from being met.

**Test Instrumentation** – The instrumentation system is capable of recording only 10 samples per second, which is less than the 12 samples requested for Parameter EC9C0F. This instrumentation capability may be a factor in the evaluation, but should be sufficient to meet the test objective.

**Test Environment** – The weather radar performance characterization requires weather cells of specific intensity during the test program execution window, but this cannot be guaranteed. Therefore, weather cell detection accuracy and wet turbulence detection assessments may not be achievable.

**Test Facilities** – A representation of the next-generation threat system is currently in development. The associated test objective (STO 9) will only be completed if the threat system model development program remains on schedule.

### 2.3 Measures of Performance




This section defines and explains the measurable system-specific design and/or performance characteristics (Example 4). The MOPs may be either quantitative or qualitative, but they must be measurable and should not be confused with test objective statements or methodology. Each MOP name is a noun or noun phrase. A concise one- or two-sentence definition of the MOP immediately follows.

When determining whether to include a specific military utility or regression MOP, test teams should consider the duration of the test program; creating military utility or regression MOPs ensure continuity from planning through reporting, given personnel turnover and/or PO pressures.

Statistical analysis should be considered independently for each MOP. When variation in observed performance measures is possible (i.e., probable MOP outcomes are not binary), statistical methods should be used to ensure test conclusions are defensible. If the MOP is stochastic, there should be an accompanying statistical evaluation (e.g., statistical intervals) to quantify the uncertainty in the results. Teams should use the lower and upper bounds of the interval to formulate appropriate evaluation criteria and/or system characterization depending on test objectives.

Some MOPs may share common aspects (such as evaluation criteria or methodology). In those cases, authors may choose to create an introductory section prior to the MOPs that details those shared aspects such that they are not repeated for each MOP, or address multiple MOPs in a table (e.g., one MOP per column, with merged cells as appropriate). Authors should avoid repeating the same statement multiple times throughout the T&E section.

## EXAMPLE 4

**2.1.1.1.1 MOP 1.1.1 – Specific Range**

A performance metric used to measure normalized fuel efficiency, typically expressed as the distance traveled (nautical air miles) per unit (pound) of fuel.

★★★

**2.1.1.1.2 MOP 1.1.2 – Link Availability**

Percentage of time that the configured command and control link is in Ready/Ready status.

### 2.3.1 Test Methodology

**WHAT** is the test approach?

The test methodology section is a MOP-specific description of the test approach outlining how the team will use the SUT and support resources to gather the required data (Example 5).

This section should list any test maneuver, condition, or state used to execute the test. If methodologies are brief (less than a page), the test point matrix and test procedures/maneuvers may be discussed here; longer lists should be placed in appendices. Often two appendices will be used: a Test Point Matrix and a Test Procedure/Maneuver Description appendix. Authors should maintain traceability between MOP test points and conditions. Authors should avoid including test cards in the test plan; however, there should be sufficient information in this section to develop flight cards and help establish the general order of the test cards. Table 3 provides distinguishing elements between content appropriate to the Test Methodology section versus the test cards.

Table 3 Test Methodology vs. Test Cards

| Element | Test Methodology | Test Cards |
|---|---|---|
| Typical Location | Test Plan Section 2 | Separate from Test Plan |
| Level of Detail | Summary and Rationale | Specific Actions |
| Focus | Test approach; translates the test strategy outlined by the test objectives into test techniques and procedures.<br><br>Users of this product are technical reviewers and the test team. Provides enough information to start writing test cards and help establish the general order of the test cards.<br><br>Examples: the number and type(s) of maneuver(s). | Executable steps; combines information from multiple sources (test methodology, test plan appendices, technical orders, regulatory guidance, and system descriptions).<br><br>Users of this product are engineers and operators during execution.<br><br>Examples: Stick movements, button pushes, tasking imagery collection, etc. |

**EXAMPLE 5**

Three speed-power test points will be flown at each test condition (see Appendix X for detailed methodology and Appendix Y for list of test conditions).

★★★

For each flight test point:

1. Establish a Ready/Ready command and control link from the ground segment.
2. Monitor Ready/Ready status for a minimum of 30 minutes. The link need not be in control and the 30 minutes need not be continuous.

### 2.3.2 Test Completion Criteria

**WHEN** is testing finished?

Test completion criteria specify how much data (quantity and quality) are required to complete the analysis for each MOP (Example 6). Test completion criteria can be as simple as a finite number of test points executed, the number of hours of operation, or achieving some minimum level of operation. If a particular level of statistical rigor is required, the appropriate number of data samples should be discussed. Explain in this section why the criteria were selected (e.g., statistics, safety, etc.). Usually, deficient system performance does not affect test completion criteria; the test is still considered complete if identified completion criteria are met.

**EXAMPLE 6**

Testing will be complete when test points 99.001 through 99.102, as listed in Appendix Z, are correctly executed and required data have been collected.

★★★

Testing will be complete when the link under test is established and observed for at least 30 minutes with required data collected. Past testing has shown that link instability is likely to manifest within 30 minutes, as documented in 412TW-TR-99-99 (Reference 77).

### 2.3.3 Expected Test Results

**WHAT** are the most realistic test outcomes?

The section is intended to encourage the test team to explain possible test outcomes from a technical standpoint (Example 7). Test teams should explain when the probable outcomes are not well understood, the results are not binary, or the system is expected to be borderline or worse. When possible, test teams should identify the source of expected results, including enough specificity to understand their applicability and fidelity (such as the version number of the model).

Understanding the expected system behavior not only allows test teams to identify when the system is not operating correctly, but also allows the team to understand the impacts of unexpected test results and whether it is appropriate in a technical sense to continue testing.

The safety plan has a similar section that is reserved for test results with safety planning implications; the safety plan is the basis for determining whether an unexpected test event (UTE) has occurred. Generally, when test results are different from what is written in the test plan, the test team should investigate but does not necessarily have to declare a UTE. Refer to AFTCI 91-202 (Reference 4) for more information.

**EXAMPLE 7**

Based on results from aero-performance modeling and simulation with AIM-9X installed, specific range is expected to decrease by less than 10 percent at all subsonic flight conditions, with decreases up to 14 percent at supersonic flight conditions.

★★★

Based on preliminary lab testing, link availability is expected to be good as defined by the evaluation criteria.

### 2.3.4 Data Requirements

This section is intended to identify the MOP-specific test data required to conduct data analysis (Example 8). Data may include surveys, video, or other products as well as traditional telemetered or recorded data (expected file type[s] should be included). Traceability between data requirements and specific MOPs can reduce confusion during test execution; should parameters become inoperative, test teams need to know which test points remain executable. The data parameter list supports both instrumentation and control room personnel in obtaining the correct recorded and telemetered data (see Appendices section).

**EXAMPLE 8**

Data parameters EG001, EG003, and EG007 (as defined in Appendix X) are required for this MOP.

★★★

The following are required:

- Pilot, engineer, and test team notes
- Ground segment logs
- Pilot flight display video recording*

*Required once during test program; desirable otherwise.

### 2.3.5 Data Analysis and Final Data Products

**WHAT** data products will be reported?

This section explains how test data will be processed, analyzed, and presented (Example 9). If the processing and analysis are expected to be simple, provide analysis methods (e.g., equations, algorithms, etc.) in this section. Otherwise, data analysis tools and methods should be summarized and the final data products listed. In consultation with technical experts, detailed information should be included in a data analysis plan (DAP), either as a test plan appendix or a standalone document. If a new data analysis technique is planned, it should be mentioned here but described in a DAP.

The final data products are the tables, charts, plots, or other figures that will be used to support conclusions and recommendations in the technical report. Although specific examples of data products are not required in the test plan, the types of final data products produced should be determined by the test team and coordinated with the technical expert(s).

**EXAMPLE 9**

Aero-performance data from the test points will be provided to the contractor in order to produce updated flight manual charts. A description of test results will be provided in a final technical report. Plots and/or tables summarizing a specific range will be provided in a technical report data package; specification limits will be depicted where appropriate. Maneuver time history plots will be provided in the data package.

★★★

Link availability will be calculated as a percentage of time a link is in Ready/Ready status over time a link connection is attempted. The link is available when command and control is established. Intentional drops of the link are not scored against the link. Recorded drops of less than 1.5 seconds will be treated as data anomalies and will not be scored against the command and control link.

### 2.3.6 Evaluation Criteria

**WHAT** are the evaluation criteria of each MOP?

Evaluation criteria are used to assess each MOP against standards of system performance and/or functionality (Example 10). The sources of any specifications to be used for comparison or reference (e.g., the Interface Control

Document [ICD], MIL-STD, or CDD) should be cited. If there is no specification or baseline for comparing data, the rationale to be used for determining the evaluation criteria should be explained. Evaluation criteria are not required for data collection test objectives, and may not be appropriate for other test objectives, particularly in the case of technology demonstrations (see Table 1).

Evaluation criteria are usually stated as a single sentence or in a table listing performance attributes under certain conditions. Each MOP must be addressed; however, if multiple MOPs share evaluation criteria, a cross-reference is provided. As shown in the 412 TW Rating Scale (see the Appendices section), descriptors (e.g., good, borderline, or deficient) apply to MOPs, whereas ratings (satisfactory, marginal, and unsatisfactory) are applied to test objectives. When more than one observation is in a sample, point estimates are used to summarize system performance and should be accompanied by statistical bounds to account for uncertainty in the data. Evaluation criteria are compared to the lower and upper bounds to determine the appropriate descriptor.

For regression testing, test teams should compare test results with legacy performance/functionality; if systems are unchanged, they are expected to continue to function as in the previous iteration (not necessarily as originally designed). Thus, regression test results should be evaluated as improved, unchanged, or degraded from the baseline.

Although the evaluation criteria pertaining to military utility assessments may be qualitative (i.e., operator comments and aircrew experience are critical data to this assessment), descriptors may still be assigned IAW the 412 TW Rating Scale (e.g., good, borderline, or deficient).

**EXAMPLE 10**

Test results will be considered good if the specific range is less than 10 percent less than clean aircraft values. Test results will be considered borderline if the specific range is at least 10 and less than 15 percent. Degradation in specific range of 15 percent or greater will be considered deficient.

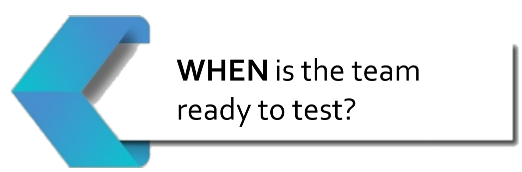
★★★

Test results will be considered good if availability of the flight critical links is 95 percent or greater and non-flight critical links are 90 percent or greater; deficient otherwise.

## 3.0 TEST CONDUCT

This section describes test-unique aspects beyond the requirements in USAF instructions, such as EdwardsAFBI 99-105 (Reference 3). This section should be closely coordinated with aircrew and test operations personnel. The Test Conduct section may include, but is not limited to, the following sections:

### 3.1 Readiness Reviews




Each organization and/or program may choose to hold reviews prior to testing to ensure test preparation is complete. The purpose of a test readiness review (TRR), or similar meetings such as a flight TRR (FTRR), launch readiness review (LRR), etc., is to gather readiness-to-test status on all the aspects of the test program, answer any interorganizational questions, and outline the final action items that must be completed. The TRRs may be held locally at the CTF level or at the PO level, but they share a common purpose. The TRRs conducted at the CTF level are led by the test team after the test package (i.e., both the test and safety plan) has been approved and ideally several weeks before testing begins.

This optional section describes elements applicable to the TRR, to include:

- A timeline of the TRR relative to test events
- Required attendees
- Readiness to test
- Program-unique aspects
- Stakeholder concurrence to proceed if required

### 3.2 Pretest Briefing(s)

This section is intended to describe the test-unique aspects of pretest briefings, to include personnel required to attend and topics to be discussed. Some organizations or programs require a day-prior brief (T-1), whereas some only require a day-of preflight brief (T-0). These briefings are narrower in scope compared to readiness reviews and focus on the individual missions. Additional briefing items may be required by the safety plan. Further guidance may be found in the test plan template and 412 OG O.I. 11-5, *Briefing/Debriefing and Flight Briefing Room Requirements* (Reference 14).

### 3.3 Test Execution

**HOW** exactly will the tests be conducted?

This section may include test procedures and setup unique to the test program. For simple tests, this section can include detailed test setup and step-by-step execution procedures. For complex tests, test setup and execution procedures should be captured in an appendix to the test plan with references in this section.

### 3.4 Posttest Briefing

This section is intended to describe the test-unique aspects of posttest briefings, to include personnel required to attend and topics to be discussed. Additional guidance may be found in the test plan template and 412 OG O.I. 11-5 (Reference 14).

### 3.5 Posttest Data Procedures

This section is intended to describe how data acquired during the test will be managed, requested, and distributed. This includes a short explanation of the system on which data will be stored, the process of how posttest data will be requested by team members, and the any test-unique methods of how data will be transferred to contractors or outside customers. Although this information may be standardized for select CTFs, identification and understanding of these processes in the planning phase is critical for efficient data processing and delivery following test.

## 4.0 TEST REPORTING

**HOW** will test results be reported?

This section describes the types of technical reporting products that may be authored following test execution, and should include expected delivery timelines. Test teams should coordinate with the customer to determine the required reporting product(s) and their delivery timeline(s). Specific guidance on 412 TW reports, including intended uses, size, delivery timelines, and expected practices, are provided in EdwardsAFBI 99-103 (Reference 6).

Test results can be communicated in a variety of formats, such as those listed in Sections 4.1 to 4.8.

## 4.1 Watch Items and Deficiency Reports

A potential deficiency may be considered a watch item (WIT) until the team determines it to be a true deficiency. Deficiency reports (DRs) document system deficiencies identified during test. If the deficiency remains and it satisfies the criteria of either a Category I or Category II DR, it will be submitted as a DR. A WIT will be closed if it does not meet the criteria of a DR. Deficiency and WIT reporting should be done IAW T.O. 00-35D-54, *USAF Deficiency Reporting, Investigation, and Resolution (DRI&R)* (Reference 15), EdwardsAFBI 99-224, *Deficiency Reporting* (Reference 16), and applicable CTF guidance.

## 4.2 Quick Look Reports

A quick look report is a high-level test summary developed by the test team after each test event and is provided to stakeholders according to an agreed upon data distribution plan, if appropriate. Quick look report information will include aircraft test configuration, test points planned, test points attempted, and a brief discussion of preliminary results with aircrew observations. These reports are usually generated after each test mission.

## 4.3 Preliminary Report of Results

The preliminary report of results (PRR) is a quick-reaction report to transmit principal T&E findings to the customer in management terms from a management perspective, and is generally not used to support major program milestone decisions.

## 4.4 Capability Report

The capability report (CR) provides overall DT&E results to support timely programmatic decisions. It is intended to address the overall results in the context of combat capability, with the respective consequences of the results on the required capabilities.

## 4.5 Technical Information Memorandum/Handbook

Technical information memorandums (TIMs) and technical information handbooks (TIHs) primarily document processes, provide instruction, or archive important technical information for engineering reference. Additionally, TIMs and TIHs may document the analysis used to substantiate recommendations regarding system models or flight manual charts.

## 4.6 Technical Report

The formal TR is a detailed report that presents the analyses, evaluation, results, and the conclusions and recommendations of the test program. The TRs and their related data packages are the most common 412 TW test reporting products.

## 4.7 Data Package

Data packages (DPs) contain supplemental test data and/or results (not ratings). The DP formatting can vary and the test team should select the best way to communicate the data. Consult with the technical expert(s) and the chief engineer, who approve the DP.

### 4.8 Test Complete Letter

If test is intended only to collect data (only collect-type objectives are planned), a test complete letter (TCL) may be issued to inform the customer that the data collection is complete and to indicate that all applicable data are transmitted. The TCL contains no analyses, subjective assessments, ratings, conclusions and/or recommendations.

## 5.0 REFERENCES

References provide the information necessary for a reader to locate and retrieve any source cited in the body and appendices of the document. References should be listed in the order they appear in the test plan. Reference information generally includes:

- Who – Author (when known; omit for test plan references)
- What – Document number, then title in italics
- Publisher – Name of publishing organization, then location (city, state)
- When – Date of publication (use a consistent format; if you have the day, month, and year for some, but only the month and year for others, use month and year in all cases)
- Classification level (if applicable)

## APPENDICES

Test plan appendices contain supplemental information that clarifies or supports the body of the test plan. Table 4 includes appendices that are often included in test plans.

Table 4 Common Test Plan Appendices

| Appendix | Inclusion |
|---|---|
| Rating Scales | Required if Ratings are Used |
| Detailed Test Item Description | Optional |
| Test Point Matrix | Optional |
| Test Procedure/Maneuver Description | Optional |
| Requirements Traceability | Optional |
| Parameter List | Required if Not in Main Body |
| Data Analysis Plan | Required if Not Addressed Elsewhere |
| Environmental Checklist | Required |
| Abbreviations, Acronyms, and Symbols | Required |
| Distribution List | Required |

Other than the Distribution List and Abbreviations, Acronyms, and Symbols appendices (which are always, respectively, the last and second-to-last appendices in the test plan), appendix order should be determined by the RIPT. The following sections describe the most common appendices.

**<u>Rating Scales:</u>**

If ratings will be used, then a Rating Scales appendix is required. The intent of rating scales is to provide consistency in the individual and overall ratings of SUTs. Generally, discipline-specific ratings are used to support the overall 412 TW rating.

If these scales do not seem appropriate to your testing, consult with appropriate technical experts before constructing a questionnaire or rating scale.

## 412 TW Rating Scales.

In general, the 412 TW Rating Scale (Table 5), 412 TW Regression Scale (Table 6), and/or the 412 TW Test Completion Scale (Table 7) are used to assess the overall SUT. If these scales do not seem appropriate to testing, the relevant technical experts should be consulted about the creation and content of a new questionnaire or rating scale. The overall rating of an SUT is based on its effectiveness or capability. Test results are captured using descriptors for individual MOPs. These descriptors are consolidated into a rating of each specific test objective, which are further consolidated into ratings for general test objectives and ultimately, the overall system rating. Further 412 TW Rating Scale guidance is available in 412TW-TIH-22-02, *Test Report Author's Guide* (Reference 7).

Table 5  412 TW Rating Scale

| How Well Does the System Meet Mission and/or Task Requirements? | Changes Recommended for Mission/Task Improvement | MOP Descriptor | Test Objective Rating |
|---|---|---|---|
| Exceeds requirements. | None. | Excellent | Satisfactory |
| Meets all or a majority of the requirements. | Negligible changes needed to enhance or improve operational test or field use. | Good | Satisfactory |
| Some requirements met; can do the job, but not as well as it could or should. | Minor changes needed to improve operational test or field use. | Adequate | |
| Minimum level of acceptable capability and/or some non-critical requirements not met. | Moderate changes needed to reduce risk in operational test or field use. | Borderline | Marginal |
| One or some of the critical functional requirements were not met. | Substantial changes needed to achieve satisfactory functionality. | Deficient | Unsatisfactory |
| A majority or all of the functional requirements were not met. | Major changes required to achieve system functionality. | Unusable | |
| Mission not safe. | Critical changes mandatory. | Unsafe | Unsatisfactory |

Table 6  412 TW Regression Scale

| How Does System Performance/Functionality Compare with Previous Test Results and Was Overall Capability Affected? | Regression Status |
|---|---|
| System performance/functionality improved, and overall capability was unaffected or improved. | Improved |
| No change to system performance/functionality, and overall capability was unaffected. | Unchanged |
| Minor changes to system performance/functionality, but overall capability was unaffected. | Impacted |
| System performance/functionality was degraded, and overall capability was affected. | Degraded |

Note: The colors in the regression scale are optional and may be omitted in cases where they do not add value.

Table 7  412 TW Test Completion Scale

| Were Collected Data Sufficient to Meet the Test Objective? | MOP Status | Test Objective Status |
|---|---|---|
| Sufficient data collected; for characterization/determination efforts, collected data were sufficient to establish a performance baseline or attribute of the system. | Complete | Met |
| Limited data collected (criticality of data not collected is relatively low); for characterization/determination efforts, collected data were sufficient to establish a limited performance baseline or attribute of the system. | Partially Complete | Partially Met |
| Insufficient data collected (criticality of data not collected is relatively high); for characterization/determination efforts, collected data were not sufficient to establish a performance baseline or attribute of the system. | Insufficient | Not Met |
| No data collected; for characterization/determination efforts, establishment of a performance baseline or system attribute were not attempted. | Not Tested | |

A detailed RGB/Hexadecimal listing of the standard rating colors is provided in Table 8.

Table 8  RGB and Hexadecimal Codes for Standard Rating Colors

| Color | Red Value | Green Value | Blue Value | Hexadecimal Code |
|---|---|---|---|---|
| **Blue** | 46 | 116 | 181 | 2E74B5 |
| **Green** | 30 | 175 | 70 | 1EAF46 |
| **Yellow** | 255 | 255 | 0 | FFFF00 |
| **Red** | 255 | 0 | 0 | FF0000 |

Note:  Teams may deviate from these nominal values for purposes of readability.

### Discipline-Specific Rating Scales.

Various organizations across the 412 TW use discipline- or test-specific rating scales. Typically, these scales include both descriptors and numbers. The numbers used in these scales are used to roughly convert subjective data into a numerical database for statistical analysis or graphical presentation and often come from a questionnaire used to solicit aircrew or maintainer opinions. The most common discipline-specific rating scales include the Handling Qualities Rating Scale, commonly called the Cooper-Harper Rating Scale; the 412 TW Revised Bedford Workload Scale; and the General Purpose Scale. Contact the appropriate discipline technical expert for guidance on the use of these and other scales.

Certain tests may include tailored or program-unique rating scales. The generation of these custom scales need to be closely coordinated with the appropriate technical experts and customer before use.

## Detailed Test Item Description:

This optional appendix is for expanded details of the test item which are too cumbersome for the main body. For readability, it may make sense to repeat portions of the test item description from the main body in this appendix. Examples of information best documented in the Detailed Test Item Descripion appendix include:

- An expanded description of the SUT
- Subsystem-level schematics
- Algorithm details
- Close-up depictions of key components
- Mass properties tables

- Instrumentation systems
- System modifications that make the SUT non-production representative, but are not expected to affect the overall test result (potentially limiting the application of test results to a broader population)

**<u>Test Point Matrix:</u>**

This optional appendix lists the test points required to meet the test completion criteria. The test point matrix contains altitudes, airspeeds, test maneuvers, any additional information required to execute the test point. The test point matrix should be one of the early planning tools that outlines the scope of the test program and ensures that no gaps are left in the planning process.

It should be reviewed at early test plan working groups and may be a driver for putting together a statement of capability for a test program. The information is often presented in a tabular format with columnar headings of information pertinent to the test, and may include figures.

**<u>Test Procedure/Maneuver Description:</u>**

This optional appendix lists the procedure(s) required to accomplish a given maneuver or test run. The Test Procedure/Maneuver Description appendix should be one of the early planning tools and should be closely coordinated with the aircrew or system operator. The information is best presented in a step-by-step format, and will often be consistent with similar test programs.

The information in this section should be detailed enough to build test cards, but test teams are cautioned not to include final test card levels of detail in the test plan. Safety planning and other considerations must also be incorporated into the final test cards, and are not typically available when the test plan is finalized.

**<u>Requirements Traceability:</u>**

This optional appendix should cross-reference the test objectives, MOPs, and/or test points to the requirements document (e.g., CDD, ORD, specifications, etc.). If the requirement traceability is simple, it could be included in the test point matrix table instead of in a standalone appendix.

**<u>Parameter List:</u>**

This optional appendix states the minimum data parameter (also known as test measurand) requirements; if this information is not captured by any other test document, the appendix is required. The parameter list should address both data available via a data bus (often in MIL-STD-1553B format) and via special instrumentation (often known as orange wire). In cases where the entire data bus is required, teams may want to reference the ICD, rather than listing every parameter. Typical details included in the parameter list include:

- Name
- Description
- Telemetry rate
- Data rate
- Units
- Designate: SOF, SOT, or RFD

**Data Analysis Plan:**

In consultation with technical experts, detailed information should be included in a data analysis plan (either as a DAP appendix or a standalone document). The DAP should carry the reader from raw collected data to the final data product in the report. A DAP appendix or standalone document should capture the algorithms for data reduction and analysis and plans for final data products prior to testing. Each Engineering Squadron/Flight should have detailed requirements for appropriate DAP content. The DAP should be coordinated with technical experts prior to the Technical Review Board.

Although the format of the DAP is ultimately the test team's decision, a standalone DAP may provide the team greater flexibility than a DAP appendix. Having the DAP as a standalone document allows the DAP to be updated with evolving analysis methods throughout the test program without having to formally amend the test package. Additionally, DAPs may be lengthy, and keeping the DAP separate can improve the readability of the test plan.

**Environmental Checklist:**

This required appendix contains the approved environmental checklist referred to in Section 1.9 Environmental Impact Assessment. The 412 TW Environmental Management Office may be contacted at: (661) 277-1401 or 412TW.CEV.EIAP@us.af.mil.

**Abbreviations, Acronyms, and Symbols:**

Generally, the technical editor will compile and update this appendix (required except for test reports of 24 or fewer pages cover-to-cover without this section) and will ensure that all abbreviations, acronyms, and symbols appearing in text (outside of figures and tables) are defined on first use. This appendix will include all abbreviations, acronyms, and symbols in figures, tables, and text. This appendix is always the next-to-last section of the test plan, and should be referred to in a footnote to the first table or figure title in the main body (whichever appears first).

**Distribution List:**

Generally, the technical editor will compile and update this required appendix in coordination with the test team. The distribution list is always the last appendix of the test plan and contains the 412 TW / PO-approved list of recipients of the final test plan; the list is kept updated in the current test plan template. Changes to the list should be approved by the CTF's Commander or Engineering Director.

# AUTHOR'S GUIDE REFERENCES


1. 412 TW-TIH-20-01, *Test Plan Author's Guide*, 412th Test Wing, Edwards AFB, February 2021.

2. EdwardsAFBI 99-101, *412 TW Test Plans*, 412th Test Wing, Edwards AFB, California, September 2016.

3. EdwardsAFBI 99-105, *Test Control and Conduct*, 412th Test Wing, Edwards AFB, California, September 2019.

4. AFTCI 91-202, *AFTC Test Safety Review Policy*, 41th Test Wing, Edwards AFB, California, November 2022.

5. DoDM 5200.01, Volume 2, *DoD Information Security Program: Marking of Classified Information*, DoD, Washington D.C., February 2012, updated March 2013.

6. EdwardsAFBI 99-103, *412 TW Technical Report Program*, 412th Test Wing, Edwards AFB, California, August 2013. (Certified current August 2021.)

7. 412TW-TIH-22-02, *Test Report Author's Guide*, 412th Test Wing, Edwards AFB, California, November 2022.

8. DoDI 5230.24, *Distribution Statements on Technical Documents*, DoD, Washington D.C., October 2018.

9. AFI 61-201, *Management of Scientific and Technical Information (STINFO)*, USAF, Washington D.C., November 2020.

10. Defense Technical Information Center website: https://discover.dtic.mil, last accessed November 2022.

11. DoDM 3200.14, Volume 1, *Principles and Operational Parameters of the DoD Scientific and Technical Information Program (STIP): General Processes*, DoD, Washington D.C., March 2014.

12. DoDI 5000.89_DAFI99-103, *Capabilities-Based Test and Evaluation*, USAF, Washington, D.C., December 2021. [Corrective Actions applied March 2022.]

13. DoDI 5200.02T, Enclosure 4, *Operation of the Defense Acquisition System*, DoD, Washington D.C., updated April 2020.

14. 412 OG O.I. 11-5, *Briefing/Debriefing and Flight Briefing Room Requirements*, 412th Test Wing, Edwards AFB, California, January 2010.

15. T.O. 00-35D-54, *USAF Deficiency Reporting, Investigation, and Resolution (DRI&R)*, 558 CBSS/GBHA, Tinker AFB, Oklahoma, April 2021.

16. EdwardsAFBI 99-224, *Deficiency Reporting*, 412th Test Wing, Edwards AFB, California, March 2013. (Certified current April 2017).

# APPENDIX A – ABBREVIATIONS, ACRONYMS, AND SYMBOLS

| Abbreviation | Definition |
|---|---|
| 412 OG | 412th Operations Group |
| 412 TW | 412th Test Wing |
| AFB | Air Force Base |
| AFI | Air Force Instruction |
| AFTC | Air Force Test Center |
| AFTCI | Air Force Test Center Instruction |
| BAF | Benefield Anechoic Facility |
| CDD | Capabilities Development Document |
| CEP90 | circular error probable of 90 percent |
| CR | capability report |
| CTF | Combined Test Force |
| CUI | Controlled Unclassified Information |
| DAP | data analysis plan |
| DIADS | Digital Integrated Air Defense System |
| DoD | Department of Defense |
| DoDD | Department of Defense Directive |
| DoDI | Department of Defense Instruction |
| DoDM | Department of Defense Manual |
| DP | data package |
| DR | deficiency report |
| DT&E | developmental test and evaluation |
| DTIC | Defense Technical Information Center |
| EAFB | Edwards Air Force Base |
| EdwardsAFBI | Edwards Air Force Base Instruction |
| FTRR | flight test readiness review |
| GPS | global positioning system |
| GTO | general test objective |
| HITL | hardware-in-the-loop |
| HQS | handling qualities simulator |
| Hz | hertz |
| IAW | in accordance with |
| ICD | initial capabilities document |
| IDAL | Integrated Defense Avionics Lab |
| IFAST | Integrated Facility for Avionics Systems Testing |
| ISTF | integrated system test facility |
| JSE | Joint Simulation Environment |
| LRR | launch readiness review |
| LRU | line-replaceable unit |

| Abbreviation | Definition |
|---|---|
| MIL-STD | military standard |
| MOP | measure of performance |
| M&S | modeling and simulation |
| MS | Microsoft |
| N/A | not applicable |
| NBVC | Naval Base Ventura County |
| No. | number |
| O.I | Operating Instruction |
| ORD | operations requirement document |
| OT&E | operational test and evaluation |
| PIRA | Precision Impact Range Area |
| PMSR | Point Mugu Sea Range |
| PO | program office |
| PRR | preliminary report of results |
| RFD | required for data |
| RMCC | Ridley Mission Control Center |
| SF | Standard Form |
| SIL | system integration laboratory |
| SOC | statement of capability |
| SOF | safety of flight |
| SOT | safety of test |
| STE | Special Test Equipment |
| STINFO | scientific and technical information |
| STIP | Scientific and Technical Information Program |
| STO | specific test objective |
| SUT | system under test |
| T-0 | test day |
| T-1 | test day minus 1 |
| T-2 | Temporary 2 |
| TCL | test complete letter |
| T&E | test and evaluation |
| TIH | technical information handbook |
| TIM | technical information memorandum |
| T/N | tail number |
| T.O. | Technical Order |
| TP | test plan |
| TR | technical report |
| TRR | test readiness review |
| TW | Test Wing |
| U.S. | United States |

| Abbreviation | Definition |
|---|---|
| USAF | United States Air Force |
| UTE | unexpected test event |
| UTSO | Unit Test Safety Officer |
| vs, vs. | versus |
| WIT | watch item |
| WSTC | White Sands Test Center |

# APPENDIX B – DISTRIBUTION LIST

| Onsite Distribution | Number of Copies Digital | Number of Copies Paper |
|---|---|---|
| Edwards AFB Technical Research Library<br>Attn: Alison Vasquez<br>307 E Popson Ave<br>Edwards AFB CA 93524-6630 | 0 | 2 |
| AFTC/HO<br>Email: james.tucker.35@us.af.mil | 1 | 0 |
| **Offsite Distribution** | | |
| Defense Technical Information Center<br>Email: jackie.l.rike.civ@mail.mil | 1 | 0 |
| Totals | 2 | 2 |