RADC-TR-74-270
Final Report
October 1974

VOLUME I

STANDARD PROCEDURES FOR AIR FORCE OPERATIONAL TEST AND EVALUATION

Braddock, Dunn & McDonald, Incorporated*

Distribution limited to US Gov't agencies only; test and evaluation; October 1974. Other requests for this document must be referred to RADC (IRAA), GAFB, NY 13441

*with subcontractors RCA Government and Commercial Systems Missile and Surface Radar Division and the Xerox Corporation Data System Division

Rome Air Development Center
Air Force Systems Command
Griffiss Air Force Base, New York 13441

Do not return this copy. Retain or destroy.

VOLUME I

STANDARD PROCEDURES FOR AIR FORCE OPERATIONAL TEST AND EVALUATION

D.E. Simon (RCA)
et al

Braddock, Dunn & McDonald, Incorporated*

Distribution limited to US Gov't agencies only; test and evaluation; October 1974. Other requests for this document must be referred to RADC (IRAA), GAFB, NY 13441.

*with subcontractors RCA Government and Commercial Systems Missile and Surface Radar Division and the Xerox Corporation Data System Division

FOREWORD

This is Volume I (with Appendices bound separately) of a two-volume Final Report by Braddock, Dunn & McDonald, Incorporated, 5301 Central Avenue, Albuquerque, New Mexico, under Contract F30602-73-C-0375, Program Element 65804D, Job Order 31050102, for Rome Air Development Center, Griffiss Air Force Base, New York. Mr. Bernard D. Rydelek (IRAA) was the RADC Project Engineer. Technical direction was also provided by Mr. Gus St. John of RADC.

Subcontract assistance was provided by the RCA Government and Commercial Systems Missile and Surface Radar Division and the Xerox Corporation Data System Division.

This report contains standard methods and procedures for Air Force Operational Test and Evaluation. It represents the results of the Air Force Constant Improvement Task 2 effort.

Major contributors to this report include: Mr. W.H. Norris of BDM, Technical Director; Mr. D.E. Simon of RCA, Task 2 Leader; Mr. R.C.W. Blessley and Mr. D.P. Vanarsdall, of BDM and Dr. M. Heinberg and Messrs. G.R. Havermahl, J.I. Keener, J.T. Nopanen, and L.J. Smith, all of RCA, Team Members; and Mr. J.L. Dickman, Maj. Gen. Ret. and Mr. G.W. Lutz, of BDM and Mr. G.O.Godfrey and Mr. V.W. Hammond, of RCA, Consultants.

This technical report has been reviewed and is approved.

APPROVED: Bernard D. Rydelek

BERNARD D. RYDELEK
Project Engineer

APPROVED:

HOWARD DAVIS
Technical Director
Intel & Recon Division

FOR THE COMMANDER:

JAMES G. McGINNIS, Lt Col, USAF
Deputy Chief, Plans Office

ABSTRACT

This report describes the overall structure of Air Force Operational Test and Evaluation process from the appearance of the test directive to the production of the final report. It describes the major documentation requirements applicable to major and minor weapon system acquisitions and provides guidance to the Air Force OT&E community in the areas of formulation of test objectives, selection of test concepts, determination of test planning criteria, OT&E data collection and analysis requirements, formulation of OT&E conclusions and recommendations, and test reporting. It further describes procedures for the development of statistical design of an operational test. This report culminates and effort to standardize the management and analytical procedures applicable to Air Force Operational Test and Evaluation. Toward this end, standardized data elements of measures of effectiveness are developed.

Evaluation

This report is the culmination of Task 2 under the Constant Improvement Program, Project 3105. This was one of several tasks under this program to improve the management and technical aspects of Operational Test and Evaluation within the Air Force.

This report is intended to be used as a guide for all levels of personnel in the Operational Test and Evaluation (OT&E) community. This effort was started with the full realization that it was necessary to guide the OT&E Test Director and Project Officers into areas of interface with the Operations Analysis areas, Test Methodology, Test Planning and Statistical Test Design areas without limiting anyone's innovativeness or imagination, but yet maintaining a constant eye on test objectives and resource limitations reached through the major weapon system's documentation chain. We feel that this report has accomplished this role quite well.

The contractor was mandated by his contract to develop standards in the following areas: test planning, test reporting, test procedures, statistical test design and measures of effectiveness. Survey teams were sent to each significant Air Force OT&E center within each major command to develop a data base. In addition, significant documentation was evaluated to form a larger data base from which the process of standardization could be initiated. Survey teams and documentation acquisition also touched upon the domain of DDR&E, IDA/WSEG, HQ USAF, Army and Navy.

There was a related task performed under the same contract. This was the task entitled "Data Management Information System for Air Force Operational Test and Evaluation". This task had as its objective the formulation of alternative approaches for automating the management and analytical capability of the Air Force OT&E community. It describes the present organizational and functional OT&E structure and possible ways of automating the entire structure. Also under different contracts, but under the same program structure, AFLC conducted Constant Improvement efforts which resulted in the development of procedures and methodology covering planning, specifying, testing, evaluating and reporting of Logistics Supportability, Reliability and Maintainability and Life Cycle Costs.

Bernard D Rydelek

BERNARD D. RYDELEK
Project Engineer

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

APPENDIX VOLUME

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF TABLES

Chapter 1

INTRODUCTION

Volume I Standard Procedures for Air Force Operational Test and Evaluation has been written for the use of (a) the Test Officer who is a well qualified operational officer but who has had no previous experience in testing, (b) the more experienced Test Director and (c) the OT&E staff officer at higher headquarters. It provides an overall skeletal structure of the process taking place in OT&E from the appearance of the Test Directive to the production of the Final Report. This permits a clear view of the various choices open at the many stages of the process. The volume also forms a framework for the addition of new information as its use indicates in areas where expansion into more detail is productive. From the structure, a step-by-step procedure is derived that will aid in organizing a total effort including enlisting the assistance of support (such as operations analysts).

Standardized formats with accompanying instructions will provide guidance in the key steps such as writing the Test Plan and producing the Final Test Report. Checklists for important actions will give detailed assistance in accomplishing the many OT&E associated tasks. In addition to presenting fundamental discussions on the important steps in the OT&E process, reference lists have been provided on the more technical subjects to aid in further study.

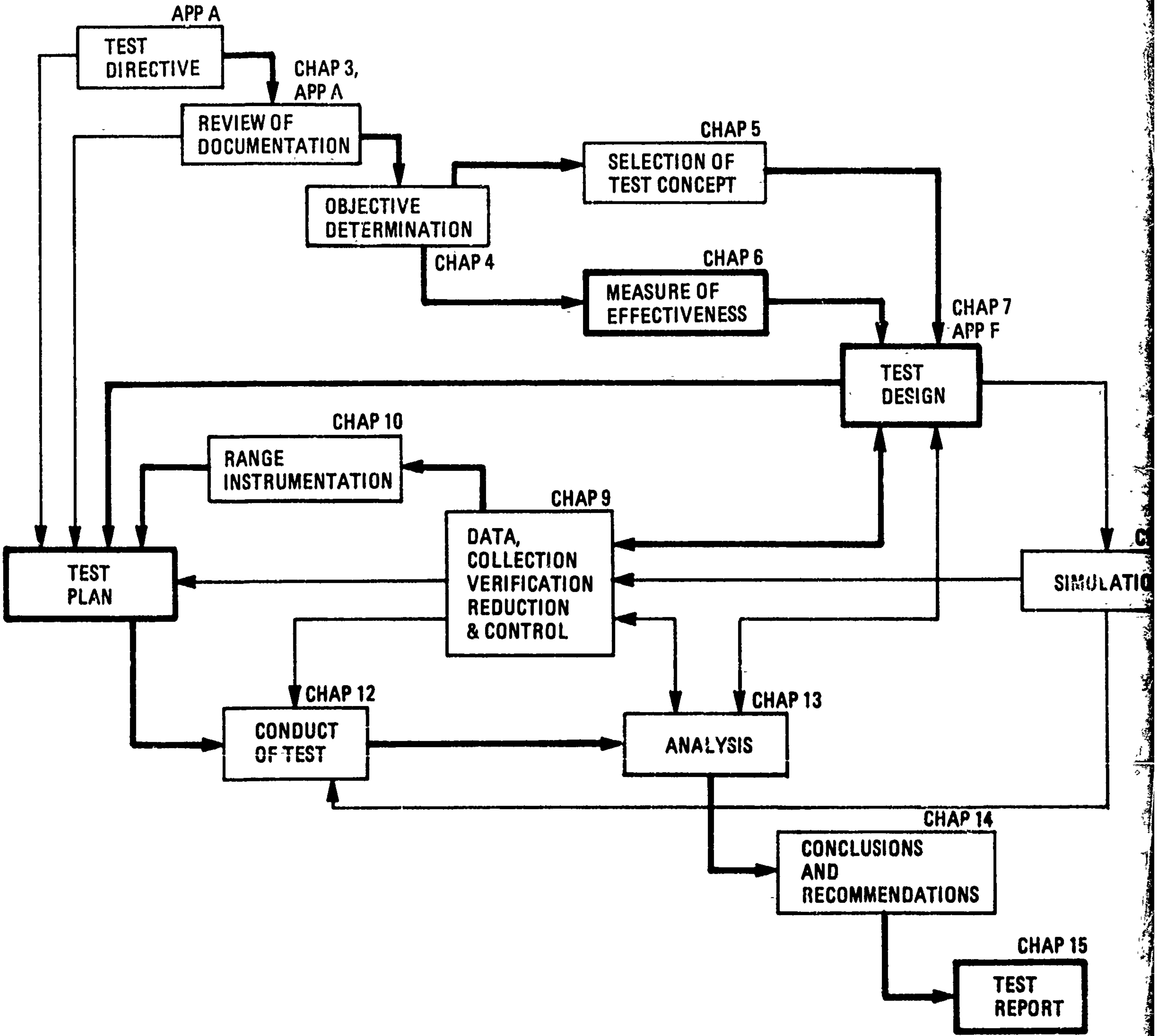
The volume is not expected to enable the Test Officer to personally perform all of the technical tasks in the total process. Rather it will enable him to know and understand the various elements needed and to integrate and manage the required support. In most cases(and especially with regard to OT&E of major systems) the conduct of OT&E is a complex evolution requiring the best efforts of a team assigned to the task. Manageable segments of the evolution are assigned to individual team members, and it is the task of the Test Director to coordinate, manage and supervise the various plans and activities of each segment.

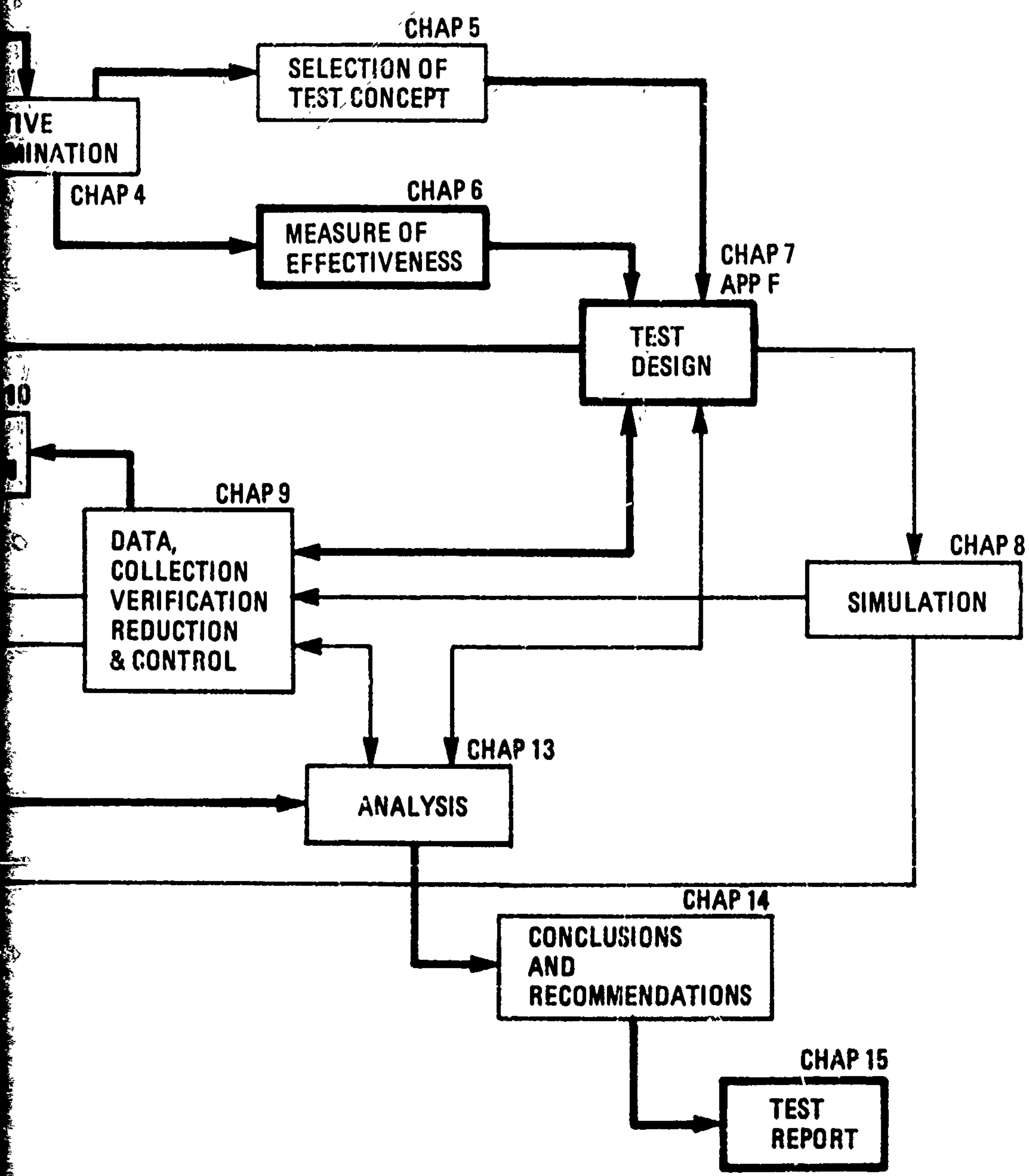
The structure of this document is heavily influenced by the process (Figure 1-1) the new Test Officer experiences from the appearance of the Test Directive to the production of the Test Report. With the appearance of the Test Directive the collection of information (Chapter 3) accelerates. The various official OT&E associated documents (Appendix A) must be searched out and studied, as well as plans and reports on similar programs. From this information the objectives can be set down and the sub-objectives formulated if necessary (Chapter 4). If the determination of effectiveness is one of the objectives, the standardized Measure of Effectiveness (Chapter 6) can be employed to meet this objective. Also at this time the latitude available to the Test Officer in the choice of test concept can be ascertained. If appropriate, a tradeoff study is conducted to select the most applicable test concept (Chapter 5), and test design (Chapter 7) can begin.

In the past, test design has been a difficult subject for the new Test Officer because of the absence of good clear simple literature on the subject. Both Service and civilian texts suffer from the archaic "agricultural" vocabulary used in the original development of the science. This handicap, plus the absence of a straightforward step-by-step process describing test design, has placed the Test Officer at a great disadvantage. Chapter 7 on test design uses an innovative non-technical approach to remove the mystery from the subject. A more technical treatment is in Appendix F.

The collection, verification, reduction, and control of data (Chapter 9) plays a central role in several of the key steps in the overall sequence. It affects test design, the Test Plan, the conduct of the test, and the analysis. Numerous examples are used to illustrate the process. Chapter 10 on Range Instrumentation has been included to acquaint the Test Officer with range equipment and procedures. These steps plus all of the preceding ones have a strong impact on the Test Plan.

The standardized Test Plan format (Chapter 11) was derived from a review of high level requirements for OT&E information, from an analysis of the shortcomings of previous Test Plans, and from the use of existing formats in order to minimize the disruptive effects of a complete change of format. The format is presented with instructions for its use and a checklist for follow up. A standardized format for test procedures is also supplied as a part of Chapter 11.


APP A
TEST DIRECTIVE
CHAP 3, APP A
REVIEW OF DOCUMENTATION
CHAP 5
SELECTION OF TEST CONCEPT
OBJECTIVE DETERMINATION
CHAP 4
CHAP 6
MEASURE OF EFFECTIVENESS
CHAP 7 APP F
TEST DESIGN
CHAP 10
RANGE INSTRUMENTATION
CHAP 9
DATA, COLLECTION VERIFICATION REDUCTION & CONTROL
TEST PLAN
SIMULATIO
CHAP 12
CONDUCT OF TEST
CHAP 13
ANALYSIS
CHAP 14
CONCLUSIONS AND RECOMMENDATIONS
CHAP 15
TEST REPORT



NOTE: HEAVY LINES SHOW MAJOR FLOW AND MAJOR STEPS

Figure 1-1. Flow diagram of the OT&E process.

Simulations are useful at several steps in the overall process from the substitution of a simulation for physical testing to the use of a simulation as a diagnostic tool in the conduct of the test. The advantages and disadvantages of the use of simulations are included in Chapter 8.

The conduct of a test is sufficiently similar to the conduct of military operations previously experienced by the Test Officer as to pose no extraordinary problems. The need to firm up responsibilities, coordinate support, and to continually supervise is a familiar one. The subjects that warrant special attention are called out in the narrative and the checklist in Chapter 12.

The processes of analysis (Chapter 13) and drawing conclusions and recommendations from the test (Chapter 14) are set down in a step-by-step sequence for the Test Officer and his supporting operations analyst to follow. Although the general process of drawing conclusions and making recommendations is an intuitive one, the particular application to OT&E might cause some difficulties without emphasis being given to the areas mentioned in Chapter 14.

The standardized Test Report format, like the Test Plan format, were both derived from high level requirements, existing formats, and analyses of previous reports. If the instructions and format are followed, sufficient information should be available to the decision makers and OT&E can effectively support the acquisition process.

Chapter 16 treats Joint OT&E. The general principles discussed in the earlier chapters are all applicable to JOT&E. Thus, simply the special interest items unique to JOT&E are covered. Additional information on the interfacing departments and agencies are covered in Appendices B, C, D, and E.

Special items of mathematics and statistics are defined and discussed in Appendix G. Other more general terms are defined in the Glossary (Appendix H) which has been intentionally limited to terms used in the preceding chapters and appendices.

This document is not directive in nature, but serves as a guide for systems Test Officers. It should not be viewed as the last word on OT&E but rather as a foundation on which to build. As the role of OT&E evolves and as use of the standardized procedures increases, modifications certainly will be in order.

Chapter 2

AN OVERVIEW OF OT&E IN THE WEAPON SYSTEM LIFE CYCLE

1. INTRODUCTION

DOD Directive 5000.3 "Test and Evaluation" defines and explains Operational Test and Evaluation as follows:

> "Operational Test and Evaluation (OT&E) is that test and evaluation conducted to estimate the prospective system's military utility, operational effectiveness, and operational suitability (including compatibility, interoperability, reliability, maintainability, and logistics and training requirements), and need for any modifications. In addition, OT&E provides information on organization, personnel requirements, doctrine and tactics. Also, it may provide data to support or verify material in operating instructions, publications, and handbooks. OT&E will be accomplished by operational and support personnel of the type and qualifications of those expected to use and maintain the system when deployed and will be conducted in as realistic an operational environment as possible. OT&E will normally be conducted in phases, each keyed to an appropriate decision point . . ."
> "In each DOD Component there will be one major field agency ... separate and distinct from the developing/procuring command ... which will be responsible for OT&E."

Operational Test and Evaluation (OT&E) planning activities commence when a new system or item of equipment is still in the conceptual stage and can be expected to continue until it is being phased out of the inventory. During this life cycle which may extend for more than 15 years, OT&E progressively changes in objective and scope. Initially, Operating Command representatives may be involved in observing demonstrations of promising technical solutions for a Required Operational Capability (ROC). Later, they will participate in and conduct tests and evaluations on new systems or equipment to provide the basis for recommending, from an Operator's point of view, whether production should be initiated. These tests are known as Initial Operational Tests and Evaluations (IOT&E). As production models become available, Follow-on OT&Es (FOT&E) are conducted to identify any remaining operational deficiencies, to determine how to operationally employ the system effectively, and to establish baselines for force structure studies and comparison against future replacements or new threats. Ideally, FOT&E is planned for completion by the time the first operational unit is equipped and ready. After the system or equipment is in inventory, the need for FOT&E continues. Changes may occur which necessitate measurements of capabilities against a new threat. While in the inventory, the need for improved operational capabilities is often met by modifying systems. This necessitates planning for additional OT&E.

The progression of these events for major programs is demonstrated by Figure 2-1. OT&E is presented against the background of the system life cycle with emphasis on the acquisition process. It is during this period that the results of testing are critical in influencing the evaluation of new systems being considered to satisfy operational requirements. A clear understanding of the progression of these events will prove helpful in identifying responsibilities and actions that must be taken to make OT&E of value to the Air Force mission. Many new items being acquired which must be operationally tested do not pass through the full range of events depicted on the chart because they do not involve the level of funding ($50M R&D/$200M procurement) or special interest requiring Defense Systems Acquisition Review Council (DSARC) procedures. These minor programs are normally approved for procurement at HQ USAF level and produce a substantial portion of the OT&E workload even though they do not have the dollar impact of major systems.

The Air Force Test and Evaluation Command (AFTEC) provides the Air Force with a separate independent test and evaluation agency. The AFTEC mission is:

a. To manage the AF OT&E program.

b. To design, plan, direct, analyze, evaluate and report on OT&E of major and designated non-major AF systems.

c. To develop policy recommendations for Air Staff approval.

	——— SYSTEM/EQUIPMENT ACQUISITION ———				OPERATIONAL EMPLOYMENT
AGENCY	CONCEPTUAL (DSARC I)	VALIDATION (DSARC II)	FULL SCALE DEVELOPMENT (DSARC III)	PRODUCTION	DEPLOYMENT IN OPERATIONAL INVENTORY
OFFICE OF THE SECRETARY OF DEFENSE	For Major Programs: 1. DCP review/approval 2. DSARC I 3. SECDEF appv'l 4. SECDEF decision memo	For Major Programs: 1. Source Selection Review 2. Review of updated DCP 3. DSARC II 4. SECDEF apprv'l/decision memo 5. Prep/Congressional Reports	For Major Programs: 1. Source Selection Review 2. Review of updated DCP 3. SECDEF appv'l for production decision memo 4. Prep/Congressional Rpt. 5. DSARC III	For Major Programs: 1. Source Selection Review 2. Review and/or redirection in case of changes in cost, performance, schedule. 3. Monitor program 4. Congressional Reports.	May fund/direct Joint Tests Comparison Tests
HQ USAF	1. Validates ROCs based on evaluation/review of Staff & Command comments 2. Issues directive to AFSC to initiate Conceptual Phase study effort 3. Prepares draft DCP	1. Issues PMD & PA/BA to conduct Validation Phase 2. DCP (updated) includes section on T&E effort. 3. Participates in Source Selection. 4. Prepares updated DCP based on AFSC, Air Staff, and Maj. CMD inputs. 5. Recommends DSARC for full-scale development decision. 6. Prepares inputs for Congressional report.	1. Issues PMD & PA/BA for full scale development, T&E responsibilities defined in PMD. 2. Participates in Source Selection Review of contractor(s) involved in full-scale development. 3. Reviews updated system documentation, DT&E & IOT&E Reports, plus evaluations by support agencies. 4. Updates DCP and requests production deployment decision (DSARC III) 5. Prepares inputs for Congressional Report.	1. Issues PMD & PA/BA for production 2. Participates in Source Selection Review of contractor(s) involved in production 3. Monitors program 4. Review and/or redirection in case of changes in cost, performance, schedule. 5. As appropriate, directs corrective actions identified by test and evaluation. 6. Prepares inputs for Congressional Reports.	1. Approve/fund Class V MOD Proposals. Direct tests in connection with MOD programs. 2. Issue tasking directives to Air Force commands selected to participate in or conduct joint or comparison tests.
AFTEC	1. Identify critical operational issues and questions and test objectives. In coordination with the Implementating Command, develop and provide a general test concept for OT&E, a test schedule, and the estimate of test resource requirements to HQ USAF as the OT&E input to the Draft DCP, the PMD and the Initial Test Directive. 2. Participate in JOTR at Implementing Command. 3. Establish liaison with SPO cadre. 4. Provide OT&E information developed in 1 above to Implementing Command for preliminary test plans. 5. Participate in any prototyping demonstrations and in internal design reviews.	1. Prepare draft IOT&E Test Design. 2. Form AFTEC test team. 3. For combined IOT&E and DT&E Programs (Acronym: CIDT&E): AFTEC Commander assigns Deputy Test Director who exercises operational control over the IOT&E portion of the CIDT&E. 4. Prepare Test Plan using DT&E as baseline. Obtain Air Staff approval for Test Plan. 5. Prepare draft HQ USAF Test Directive. Submit draft to HQ USAF/X00W for approval and promulgation. 6. Prepare, issue AFTEC/CC Estimate to HQ USAF and commands involved in IOT&E. 7. Provide IOT&E Test Plan to PM for inclusion in PMP. 8. Participate in prototype demonstration of hardware and other DT&E events. 9. Participate in JOTR. 10. Provide IOT&E inputs for Congressional reports.	1. Conduct IOT&E. Analyze data. Evaluate and report findings to CSAF. 2. Provide AFTEC/CC Summary. 3. Provide inputs to Congressional Reports. 4. Present results of IOT&E and the AFTEC evaluation to DSARC 3. 5. Participate in JOTR and design review. 6. Identify issues/questions and objectives for FOT&E.	1. Prepare draft Test Directive for FOT&E. 2. Design, plan, conduct and report FOT&E. 3. Designate Operating and Supporting Commands for FOT&E. 4. Provide OT&E requirements for Class V Modification to HQ USAF/X00W and the Implementing Command. 5. Provide AF single point of contact for Joint Tests. 6. Provide OT&E data to Congressional Report. 7. Maintain Air Force OT&E Master Program.	
	1. Reviews ROCs for tech. feasibility; comments to HQ USAF. 2. Initiates planning	1. Program Mgt. Plan (PMP) prepared/issued. 2. RFP released to potential contractors for system	1. Prepares dev. contract which includes test dem. requirements. 2. Source selection for	1. Prepares production contract 2. Source Selection for prod. 3. DT&E continues on production items.	1. Follow on DT&E. 2. Provides tech assistance on MODS 3. May provide ranges

	internal design reviews.	IOT&E. 7. Provide IOT&E Test Plan to PM for inclusion in PMP. 8. Participate in prototype demonstration of hardware and other DT&E events. 9. Participate in JOTR. 10. Provide IOT&E inputs for Congressional reports.			
IMPLEMENTING COMMAND (USUALLY AFSC)	1. Reviews ROCs for tech. feasibility; comments to HQ USAF. 2. Initiates planning. 3. Calls Joint Oper. & Tech. Review (JOTR) (optional). 4. SPO cadre established 5. Develops preliminary test plans 6. May conduct advanced development demonstrations. 7. Prepares advanced procurement plan including Request for Proposal (RFP). 8. Program Office (SPO) formed	1. Program Mgt. Plan (PMP) prepared/issued. 2. RFP released to potential contractors for system definition (tech, cost, schedules & program characteristics) and/or prototype. 3. Source Selection for def'n contract(s). 4. DT&E, hardware proofing/ prototype demonstrations. 5. Evaluation of contractor studies & con./AF tests & demos. for full scale dev. recommendation. 6. JOTR (optional) prior to release of RFP for full-scale development. 7. Supports IOT&E.	1. Prepares dev. contract which includes test dem. requirements. 2. Source selection for dev. contract. 3. Conducts Prelim. Design Review (PDR). 4. Conducts DT&E incl. AFPES 5. Conduct JOTR (optional) mid-point of DT&E. 6. Conducts critical design review (CDR). 7. Submits test rpts and updated pgm. doc'tn to HQ USAF. 8. Supports IOT&E.	1. Prepares production contract 2. Source Selection for prod. 3. DT&E continues on production items. 4. Updates PMP. 5. Reports on DT&E. 6. Prepares for system turnover to Operating Cmd. and transition to AFLC. 7. Supports Follow-on OT&E. 8. Manages configuration changes.	1. Follow on DT&E. 2. Provides tech assistance on MODS 3. May provide ranges and facilities to support Follow-on OT&E. 4. Supports Follow-on OT&E.
SUPPORTING COMMANDS: AFLC ATC AFSS AFCS	1. Review ROCs; comment to HQ USAF. 2. Participates in planning studies. 3. Participate in JOTR. 4. Identify representatives in SPO.	1. Participates in prep/review of PMP along funct'l lines. 2. Participates in pln'g studies & source selection. 3. Participate in review of full scale development program mg't plans. 4. Participate in JOTR. 5. Log pln'g & determ'ntn of training requirements. 6. Participates in/observes hardware proofing tests/ prototypes demos.	1. Participate in DT&E and IOT&E. 2. Continue support planning & preparation. 3. Participate in JOTR. 4. Participate in design reviews. 5. Submit independent evaluations on supportability for presentation at DSARC III (when AFTEC does not conduct the IOT&E).	1. Participate in Follow-on OT&E. 2. Identify/take corrective actions on support problems identified by FOT&E. 3. AFLC prepares for/takes over system mg't. 4. ATC refines/continues training program.	1. AFLC provides system mg't & logistic support. 2. AFLC-dev/test/ installation MOD kits 3. ATC participates in major MODS to determine training impact. 4. AFLC provides logistic support for Joint or Comparison Tests.
OPERATING CMDS ADC MAC SAC TAC ALSO MAY BE AFSS AFCS ATC	1. Submits ROCs. 2. Participates in JOTRs. 3. Participates in planning studies, provides operational concept. 4. Identify SPO representation.	1. Coordinates on PMP, assists AFTEC in preparation pertaining to OT&E and concept of operations. 2. Participates in DT&E involving hardware proofing and prototype demonstrations.	1. Participates in DT&E/ IOT&E as directed. 2. Continues representation in SPO. 3. Participates in JOTR. 4. Participates in design reviews. 5. Prepares IOT&E reports (when AFTEC does not conduct the IOT&E).	1. Participate in FOT&E as directed by HQ USAF.	1. Participate in FOT&E such as: Test of new appl'tns Partic'tn in Joint Tests Comparison Tests Test and Evaluation of MODS (usually Class V to improve system capabilities).

R&D PROJECT TEST AND EVALUATION
DEVELOPMENT TEST AND EVALUATION (DT&E) — FOLLOW-ON DT&E

INITIAL OPERATIONAL TEST AND EVALUATION (IOT&E) — FOLLOW-ON OPERATIONAL TEST & EVALUATION (FOT&E), OPERATIONAL EVALUATION

SYSTEM/EQUIPMENT LIFE CYCLE TEST AND EVALUATION

NOTE #1. Acronyms used in Figure 2-1 are identified in the List of Acronyms in Appendix II.

NOTE #2. Word contractions used in Figure 2-1 are identified as follows:

appl'tns - applications	demos - demonstrations	incl - including	mods - modifications	prelim - preliminary
app'v'l - approval	determ'nt'n - determination	log - logistic	oper - operational	prep - prepare
c'mds - commands	dev - development	maj - major	partic't'n - participation	prod - production
con - contractor	doc't'n - documentation	memo - memorandum	pgm - program	r'p't - report
def'n - definition	funct'l - functional	mgt - management	pl'n'g - planning	tech - technical

Figure 2-1. Progression of Events

d. To designate the Deputy Test Director and provide (with augmentation from the major commands) the OT&E team for combined DT&E/IOT&E.

e. To conduct inter-service and joint OT&E programs.

As AFTEC operating procedures are developed and the major command directives are revised, it is expected that some adjustments will occur in the events of the weapons system life cycle described in this chapter.

2 - CONCEPTUAL PHASE -

The submission of a Required Operational Capabilities (ROC), (usually by one of the Operating Commands) in accordance with AFR 57-1 starts events which may lead to the acquisition of a new system (or item of equipment) for the operational inventory. This is the normal starting point for OT&E planning. Development Testing and Evaluation (DT&E) is an essential part of this process. DT&E measures the degree to which system objectives and goals are met. Results of DT&E and OT&E provide guidance to developers and decision makers in whether to proceed with the established plans, alter them, or terminate the procurement program. This process is presented in terms of a typical system such as a new combat aircraft; however, the processes for a major item of support equipment such as a bomb loader follow essentially the same route except for the level of review and decision.

In validating a ROC, HQ USAF obtains comments from the Air Force Systems Command (AFSC) normally the developer (Implementing Command), Air Force Logistics Command (AFLC) which provides logistics support, Air Training Command (ATC) which provides training support for operating and maintaining the system, USAF Security Service (USAFSS) which is involved with communications security (COMSEC), Air Force Communications Service (AFCS) who may provide necessary communications, and the Military Airlift Command (MAC) which may be the Operating Command which submitted the ROC or provide airlift necessary to support a new system. Copies of a ROC are sent to other Operating Commands and may be provided to other US military departments and allies if deemed appropriate.

If a ROC is validated, a formal directive initiates the Conceptual Phase study effort which is normally assigned to AFSC. The directive defines the task, designates Participating Commands, and specifies HQ USAF review and approval requirements, limits of authority. It also contains other items such as test and evaluation requirements as outlined in AFR 80-14. These T&E requirements include the critical issues that must be evaluated.

AFSC investigates alternatives for providing the capability to meet the need or threat and identifies and selects a preferred approach. This includes trade-off studies and determination of the economic, military and technical bases associated with the program. Before inputs to HQ USAF for inclusion in the initial draft of the Development Concept Paper (DCP), AFSC may call a Joint Operational and Technical Review (JOTR) as outlined in AFSC Regulation 800-18. These reviews provide the Commanders of AFSC, and the designated Operating and Supporting Commands with the relationships between operational and support concepts, system requirements, characteristics of the conceptual designs, technical difficulties, and estimated life cycle costs. System reviews are designed to avoid excessive acquisition costs by reducing marginal requirements, to protect those requirements essential to operational effectiveness, and to obtain support for the acquisition effort necessary to minimize lifetime cost of ownership (including the DT&E and IOT&E effort). During this phase, demonstrations of the feasibility and capability of key subsystems (and high risk items) identified in studies and analysis may be appropriate. By the time AFSC inputs are forwarded to HQ USAF, the cadre of a System Program Office (SPO) may be formed if required. In the case of an aeronautical system or aircraft, the SPO cadre would be formed at the Aeronautical Systems Division (ASD) at Wright-Patterson AFB, Ohio. At this time, information is available to the designated Operating Command responsible for conducting the OT&E to permit the start of preliminary test planning which may include identification of test units, facility requirements, time phasing, and contact points in the Implementing and Supporting Commands. In case the Operating Command has participated in or obtained reports on feasibility demonstrations of related hardware, these data can be used to refine preliminary test planning.

HQ USAF uses the AFSC input and other data in drafting a DCP. After Secretary of the Air Force (SAF) approval, the DCP is forwarded to the Office of the Secretary of Defense (SECDEF) for approval to proceed with the program. DCPs reflect program objectives, issues, risks and risk reduction approaches, accomplishments, and future plans. They highlight the major issues to be considered by the Defense Systems Acquisition Review Council (DSARC) before Secretary of Defense decisions. The plans for the next program phase will include any planned test and evaluation effort. Although any test and evaluation effort during the Validation Phase will be primarily DT&E, it provides important information for planning Initial OT&E (IOT&E).

The draft DCP is reviewed by the Offices of the Director of Defense Research and Engineering (DDR&E) and the Assistant Secretaries of Defense prior to DSARC I review. This review includes the Office of the Deputy Director (Test and Evaluation) for all test and evaluation planning. The program is next reviewed by the DSARC which is made up of the DDR&E and Assistant Secretaries of Defense for Installations and Logistics (I&L), Systems Analysis (SA), and Comptroller. The DSARC evaluation of the DCP provides the SECDEF with input to assist his decision. When an Air Force program is under review, the DSARC meeting is attended by the SAF and the Program Manager who can discuss any aspect of the program. If the SECDEF is satisfied he approves the DCP and provides guidance for the Validation Phase. The DCP now constitutes the program contract between DOD and the Air Force. After funding is available to HQ USAF the Air Staff issues a Program Management Directive (PMD) and a Budget Authorization (BA) to AFSC providing authority to obligate funds, and a Program Authorization (PA) which complements the BA and provides programming guidance to AFSC. The program now moves into the Validation Phase during which DT&E may, if prototyping is involved, begin with participation by Operating Command personnel who have the responsibility for OT&E.

During the program decision process a number of actions are taken by AFSC which must be considered in OT&E planning. After approval of the draft DCP by the SAF, a revised action directive is normally issued to AFSC directing actions in preparation for the Validation Phase. These actions, as well as others involved in the system acquisition process, are outlined in detail in AFSC Pamphlet 800-3 "A Guide for Program Management". These include: drafting the Request for Proposal (RFP) including the Work Statement for which potential contractors will submit proposals to accomplish the contractual requirements of the Validation Phase; issuance of priorities/guidance/directions to the appropriate AFSC Divisions, Centers, and laboratories; establishing the SPO; establishing the functional baseline which includes system performance objectives, operational and logistics concepts, cost estimates, a preliminary system specification defining the technical portion of the program baseline; preparation of the basic Management Plan, Cost Information Report Data Plan, Advanced Procurement Plan, Source Selection Plan, Real Property Facilities Plan, Logistic Support Plan, Production Plan, and a Test Plan which covers planning for the Validation Phase, and (as applicable) Full Scale Development Phase. Component, subsystem or system prototyping, and testing may be required prior to full-scale development. Through liaison with the SPO during this period, the designated Operating Command will be able to obtain information to initiate planning for IOT&E including, test design studies, and will consider appropriate simulations and will conduct preliminary planning for FOT&E.

3. VALIDATION PHASE -

The issuance of a Program Management Directive (PMD) plus a PA/BA to AFSC after all approvals moves a program into a phase which may involve hardware/prototype procurement or system definition based on studies normally prepared by contractors. With the availability of hardware and/or prototypes, AFSC will proceed with DT&E with participation of the Operating Command responsible for IOT&E and the Supporting Commands who will be responsible for preparing independent evaluations concerning supportability.

The PMD may necessitate changing some of the program documentation prepared during the Conceptual Phase. If not already directed, the PMD will probably direct preparation of the Program Management Plan (PMP).

The PMP is prepared, approved, and issued by the AFSC Program Manager with the cooperation and coordinated efforts of the designated Operating Command, AFLC, ATC, and other Air Force elements involved in the program. The PMP is the baseline document showing the integrated, time-phased tasks and resources necessary to complete the program. The test and evaluation section of the PMP addresses the overall test concept, participating organizations and their responsibilities, and planned DT&E and OT&E (particularly, IOT&E) activities. As the program progresses, the PMP is updated to reflect any changes in program guidance, schedules, and funding.

The Request for Proposal (RFP) for system studies, simulations and/or hardware development or prototyping is released by AFSC in accordance with the Advanced Procurement Plan. Contractor proposals are reviewed by a Source Selection Advisory Council who recommend contract award to the Source Selection Authority. Validation may be conducted primarily on the basis of studies leading to system definition or on the basis of studies and hardware or prototype demonstrations. Current trends stress hardware proofing and/or prototype demonstrations as a basis for system validation. DT&E and IOT&E are part of this process. A Full Scale Development Phase will be conducted later during which DT&E and IOT&E will continue to determine if the system should be placed into production.

Contracts for fabrication of hardware or prototype items for system validation purposes normally include the requirement to test and demonstrate functional and performance capabilities. Although this DT&E is conducted primarily by the contractor, it is closely followed by AFSC test personnel who observe test procedures, review results, establish procedures to assure program continuity into Full-Scale Development, and they may actively participate in some test operations or flights. Operating and Supporting Command T&E personnel also normally monitor these tests to gather information for their later IOT&E and independent evaluations required for the production decision. Participation by Operating Commands in OT&E during Validation varies from program to program. When competing contractors build prototypes, test and evaluation personnel from AFSC, the designated Operating Command, and Supporting Commands normally participate in a substantial portion of the demonstrations or "fly-offs."

During the Validation Phase, the SPO and contractor update the functional baseline of the system and prepare the design requirements (allocated) baseline for the Full-Scale Development Phase. Contractors also prepare proposed Statements of Work (SOW) reflecting the results of study and hardware/prototype demonstrations, and other reports as required by their contract for this phase of the program. Based on contractor inputs, studies conducted by Research and Development agencies, and test participation/observation, the SPO prepares RFPs for full-scale development and provides material to the Air Staff in updating the DCP. AFSC may conduct another Joint Operational and Technical Review (JOTR) at this time before the RFP is released for full-scale development of the system and before DSARC II. Again, the purpose of a JOTR is to examine the requirements and reduce the cost of ownership. At this point, AFSC and the designated Operational Command(s) and Supporting Commands will have better visibility on how the program should proceed and will be better able to make detailed tradeoff studies to improve the program.

HQ USAF is responsible for updating the DCP and other program documentation. After SAF approval it is forwarded to OSD where updating is completed. The DSARC reviews the updated DCP and other program documentation to assure that the proposed procurement method, contract type, management system, test result, etc. are matched with the job to be done, goals, risks, and uncertainties. The DSARC may add to the DCP direction for follow-on activity. Favorable DT&E reports based on hardware proofing of major components and subsystems or prototype demonstrations materially aid in reducing the risks and uncertainties and support an Air Force recommendation to contract for full-scale development. If recommendations of the DSARC and their review of program documentation concerning such things as technical, financial, and schedule aspects satisfies the SECDEF, he approves the DCP and the program proceeds into the Full-Scale Development Phase.

4. FULL-SCALE DEVELOPMENT PHASE -

Test and Evaluation conducted during the Full-Scale Development Phase is highly visible since it is during this period that a preproduction system which closely approximates the final product, including most support items, is designed, fabricated, and tested to determine its ability to meet requirements. The improved visibility must not be allowed to minimize the importance of test planning and preparation, including testing and evaluation involved in hardware proofing and prototype demonstrations, during the Conceptual and Validation Phases. Early test and evaluation is especially important since it reduces the risks and unknowns which become increasingly expensive to correct as system development progresses. Test and evaluation during Full Scale Development is conducted by the Air Force and the contractor. DT&E directed primarily for assessing technical adequacy is conducted by AFSC and the system contractors who are responsible for contract performance. IOT&E, which may have been initiated on a limited scale during the Conceptual and Validation Phases, continues and is completed during Full-Scale Development to support the independent evaluation of a new system by the Operating Command. The IOT&E report is required before a decision is made to place the system into production and deploy it. During IOT&E and DT&E, which are often conducted jointly because of resource and time limitations, technical and engineering as well as operational usability and supportability problems may be encountered and identified, even though these risks were carefully investigated during the preceding phases. Each such problem requires detailed analysis and specific consideration of tradeoffs between stated operational requirements, cost, and readiness date.

After the SAF forwards the DCP to the Air Staff with approved documentation for proceeding with the Full-Scale Development Phase, HQ USAF provides the direction to the Implementing and Participating Commands via an amended PMD along with BA/PA actions which provide funding for the Development Phase of the program.

The Full-Scale Development contract is negotiated and awarded in accordance with the approved DCP subject to source selection review procedures which may have been specified at HQ USAF and OSD levels. This contract will specify production of hardware and facilities required to support development and test and normally specifies test demonstration requirements expected of the contractor.

Early during this phase, management plans and documentation are revised to make them compatible with the approved Full-Scale Development Phase guidance. Preliminary Design Reviews (PDRs) are conducted for each Configuration Item (also known as a CI and refers to an aggregation of hardware or software, or any of its discrete portions, that satisfies an end use function and is designated for configuration management such as a complete aircraft, missile, or item of support equipment). These reviews assure that the approach is feasible and sound from a design, development, test, and deployment viewpoint and that the performance/design requirements developed during the Validation Phase can be effectively met. Participation of the Operating Command in the PDR provides the opportunity to identify operational factors to the designers to avoid difficult operating methods in the design and to accommodate operational characteristics which are best known to the operator.

IOT&E and DT&E during Full-Scale Development is paced by the rate of fabrication and delivery of the new system. During the initial part of this phase, continuing DT&E activities are primarily those performed by a contractor under SPO direction. Observation and monitoring may not be feasible. Personnel responsible for IOT&E planning should track the progress of these tests. Prior to full-scale system DT&E, a highly important contractor test activity is that connected with the large volume of component and subsystem qualification tests (sometimes referred to as "shake, rattle, and roll") required to demonstrate compliance with specifications and Military Standards and regulations pertaining to reliability, survivability/vulnerability, electromagnetic compatibility, etc. These tests are stressed because problems identified at this point can be corrected at a relatively modest cost or an alternate source sought. As development progresses, the first pre-production systems will be completed and more complicated events such as first flights, engine qualification and endurance runs, weapons launch and stores separation demonstrations, and other tests for verifying compliance with performance and design requirements in the system specification take place. Air Force participation in the DT&E process increases, and as discrete test objectives are completed, test results are evaluated and recommendations made to correct deficiencies and improve performance/design. These are implemented in the form of changes to the design, the contract, or to system specifications in accordance with contractual configuration management procedures.

During DT&E, Air Force Preliminary Evaluations (AFPEs) will be conducted to determine system performance, estimate potential, identify gross deficiencies, and to determine/estimate the degree to which contract specifications are being met (see AFR 80-14). AFPE data provides a source of planning information for IOT&E. The AFPE normally involves participation by the Operating and Supporting commands.

At approximately the midpoint of system level DT&E, another JOTR may be held by AFSC. This provides a most valuable system progress check point since it is based on actual test results. As with other JOTRs, Commanders from the Operating and Supporting Commands participate in the operational, technical, and trade-off presentations, and become party to working out mutually acceptable positions.

During DT&E and IOT&E, the rate of test accomplishment is a matter of concern closely related with attainment of the technical, operational, and supportability objectives. Technical difficulties, adverse weather, data loss, and resource or facility availability are among the types of problems which result in test — and thus program — schedule slippage. Generally, slippage in program schedules results in higher costs.

Although a fixed program schedule cannot be allowed to dominate testing with the result that insufficient data or time is taken to achieve test objectives, careful planning and consideration of alternatives early in a program can help avoid such problems. Since resources are nearly always limited, the need to combine IOT&E with DT&E to the maximum extent possible is recognized. The extent of participation by Operating and Supporting Commands in the testing to be conducted at AFSC test centers is established during joint planning. Test events which will provide operational or supportability data are included. Revisions in the test program may be necessary because of loss of test items or for other reasons already noted. This may result in the need to provide additional resources and higher priorities on facilities such as test ranges, computer facilities, and services of support agencies and personnel to support higher test sortie rates and make-up missions.

The critical design review (CDR) is significant to testing since it is normally conducted prior to or early during system level DT&E. The purpose of CDR is to review and approve the contractors design solutions prior to producing or acquiring the pre-production prototypes for testing. Continuity in participation by Operating and Supporting Command personnel involved in the IOT&E and who have participated in test activities, provides a better opportunity for identifying any changes that may be necessary to lessen or eliminate possible operational or supportability problems during later OT&E.

DT&E test results which provide a measure of the degree to which technical, engineering, and stated requirements have been met are presented in periodic progress reports, briefings and formal reports. They provide invaluable data for operational and supportability considerations in IOT&E since no aspect of performance should be considered in isolation. As DT&E is completed and the program moves toward the Production Phase test results are used in performing Functional Configuration Audits of all Configuration Items (CIs), updating the product baseline for the production contract RFP, and in preparing inputs to HQ USAF for updating the DCP.

The Operating Command representative observes the technically oriented DT&E in accordance with agreements with the SPO and the AFSC test center. He participates in DT&E and directs IOT&E. These observations and test data are used in FOT&E planning and in preparing the independently reported IOT&E required prior to the Production Decision. The Supporting Commands (principally AFLC and ATC) participate in both DT&E and IOT&E activities to obtain data for preparing the independent logistical and training supportability reports as required by DODD 5000.1 and AFR 80-14. Comments on communications security and communications services aspects may also be required from AFSS and AFCS.

The DCP is again updated by HQ USAF, coordinated throughout the Air Force as necessary, and submitted to OSD for a pre-DSARC planning meeting at least 30 days prior to DSARC III, the production/deployment decision.

The updated DCP is again reviewed in draft by DDR&E and the other Assistant Secretaries of Defense prior to DSARC review. Comments on the coordination draft are provided by OSD to the Air Force and the DSARC chairman within five work days of the DSARC. These comments are also provided to AFSC and the Program Manager for DSARC preparation.

The Program Manager recommends system production/deployment to the DSARC. In addition to the DSARC members, normally the SAF, CSAF, the appropriate Assistant Secretary, the DCS/R&D, and the Program Element Monitor (PEM) are in attendance. The DSARC III Test and Evaluation presentation includes assessments of test results from the Operating and Supporting Commands as to operational effectiveness and suitability and supportability. IOT&E results are presented by the operational test agency.

Following the DSARC III, the SECDEF decision is made known by a decision memorandum. HQ USAF then updates the DCP incorporating the SECDEF decision and distributes it to the DSARC principals, AFSC, the Program Manager, and others as appropriate.

5. PRODUCTION/DEPLOYMENT PHASE -

After the production/deployment decision has been made, DT&E continues on preproduction system items until test and evaluation of all technical aspects has been completed and adequacy of corrective actions initiated during Full-Scale Development have been verified. As production items become available, DT&E continues to verify technical adequacy and performance requirements. Operational test personnel continue participation in DT&E in preparation for FOT&E normally scheduled to begin when the first production items are delivered to an Operating Command.

Program directions and priorities for the Production/Deployment Phase are transmitted by HQ USAF and funding is provided through BA/PA actions. The RFP and other program documentation for the Production Phase were prepared during Full-Scale Development. These documents may require revision to reflect DCP guidance and will be used as the basis for negotiating and awarding the production contract.

As production gets underway, operational personnel continue participation in DT&E and should be aware of the test and contractor manufacturing procedures. Using

DT&E data and with knowledge of the production processes, including possible areas of difficulty, planning for FOT&E with production items is completed and potential problem areas are identified.

A Physical Configuration Audit (PCA) is conducted to assure that the hardware produced complies with contract specifications. The PCA establishes an approved configuration for system items which will be delivered to Operating Command units for the FOT&E.

Normally, FOT&E is planned to begin as soon as delivery of production items is made to the Operating Commands. FOT&E will continue at least until the first operational unit is equipped. Ideally, the test will be completed by the time the first unit is operationally ready. These tests are conducted in accordance with plans initiated early in the acquisition cycle and refined on the basis of data/information gathered in DT&E, the conduct of IOT&E, and updating with consideration for changes in threat and operating environment. These tests are conducted with operationally qualified crews under actual or simulated operational conditions. They include an assessment of the production system operational capabilities; development of tactics and procedures; evaluation of the logistics systems; crew training requirements and procedures; and adequacy of operating, maintenance, supply, and training publications.

FOT&E provides the first opportunity to collect operational data with a production system for refining logistic support estimates for AFLC and verifying adequacy of training courses for which ATC will retain responsibility throughout the life of the system. Thus, participation by AFLC and ATC as well as other Commands and agencies, which began during IOT&E, can be expected to continue during FOT&E.

During FOT&E with the first production items, the Operating Command Test Plan will require specific reports on any deficiencies significant to mission performance. Included are recommendations for modifications to be immediately processed through program channels. Such reports will normally be in addition to planned documentary test reports and may require specific follow-up actions to be taken. When possible, updating changes are made while production is still underway to correct any deficiencies revealed during this phase of testing. With proper planning, most of the system deficiencies can be identified during DT&E and IOT&E for the system configuration tested thereby making it possible to incorporate necessary changes prior to the production contract. Late identification of required changes will result in increased costs for updating changes or modifications thereby causing a serious impact on the programmed funds. Early identification and correction of deficiencies also reduces the need for Follow-on DT&E. Program benefits are also derived from OT&E by earlier stabilization of logistics support, crew training and qualification, and formal training courses, all of which tends to accelerate and stabilize the operational readiness date.

The event known as "system turnover" occurs when the Operating Command accepts responsibility for operation and maintenance of the first operating unit of a new system. Turnover includes acceptance of responsibility for related property accountability and inventory of subsequent units. FOT&E initiated during the Production Phase probably will not have been completed at this point. With proper consideration for system turnover in OT&E planning, the event will take place with no unexpected impact on test execution.

Another event during the system life cycle is the transitioning of AFSC program management responsibility from AFSC to AFLC. This is accomplished by means of a formal agreement approximately following completion of production and when the system design has stabilized and pattern deficiencies have been corrected. Upon completion of transition, AFLC assumes logistics support and management responsibility including engineering, procurement, and financial management. At this point the SPO is usually dissolved, and necessary technical support or T&E follow-up actions are resolved by mutual agreement between AFSC and AFLC. Program channels and relationships for OT&E on the first production systems are usually not affected since transition rarely takes place prior to test completion. Follow-up actions on deficiencies revealed during first production OT&E may not have been completed prior to transition, therefore, shift of responsibility would have to be included in the transition agreement and AFLC will then become the primary source of engineering and other system support for subsequent FOT&E.

6. SYSTEM INTRODUCTION INTO THE OPERATIONAL INVENTORY -

After all production items have been turned over to the Operating Command the system may remain in the operational inventory for many years. Follow-on OT&E for new applications (which may involve system modifications or the addition of new subsystems, and the determination of how to conduct operations in a new threat environment), contibute to extending the systems life cycle. In addition, FOT&E can be expected to include participation in joint exercises in which the system is employed in coordination with systems operated by other services and/or its capabilities may be measured against other systems operated to simulate a current or postulated threat.
Changes identified in FOT&E and recommended to improve system capabilities while in the inventory are designated and accomplished as Class V Mods in accordance with provisions of AFR 57-4 on retrofit configuration changes. Desired changes are submitted by the Operating Command to HQ USAF for approval via a ROC prepared in accordance with AFR 57-1. Class V Mods are normally recommended when significant improvements can be achieved in existing systems or equipment that will provide a new or improved operational capability. Installation of sensors which provide night weapons delivery capability is an example. Other examples might include changes in fuzing, warhead size, or sensor tracking gimbal limits in an air-to-air missile. The net result of such modifications is that scaled down operational testing approximating that involved in a new system acquisition is required. Even though management may be retained by AFLC throughout the program-AFSC may participate to provide specialized technical and DT&E type support. On extensive modifications requiring a large scale OT&E, AFLC and ATC personnel will normally be involved to obtain data required to assess the impact on logistic support and training.

OT&Es to determine system or equipment employment suitability for new applications and desired capability against a new threat are often originated by and conducted entirely within an Operating Command's own resources. In some instances OT&E may be conducted at the direction of HQ USAF or OSD to satisfy data requirements for force structure studies and system analysis. Tests of this type directed by HQ USAF or OSD frequently involve operations against threat systems (or simulators) possessed by another Service or Air Force Major Command whose participation has been directed by or established through the same authority. In the case where an Operating Command desires to conduct an OT&E against a threat environment (an acquisition radar simulator, for example) possessed by another Command, arrangements are usually made on the basis of a direct request sometimes involving the transfer of funds if the facility is industrially funded. Unless direct contact has been authorized, use of facilities and services possessed by another Service is normally arranged through HQ USAF.

Joint exercises directed and controlled by OSD are normally designed to measure system capabilities in a specific mode of operation and/or against a defined threat or target to provide test data for force structure analysis. HQ USAF directs Air Force participation which is usually conducted as an OT&E activity. These tests further evaluate current system capabilities and may reveal deficiencies which form the basis for recommended changes to improve system capabilities, or may lead to the formulation of a new ROC, beginning the cycle anew.

Also, during this period FOT&E may disclose latent deficiencies which were not identified until production items were turned over to be Operating Command.

Chapter 3

REVIEW DOCUMENTATION

1. SCOPE

This section provides both identification and source of review documentation as an aid to personnel responsible for the validation of the OT&E testing requirement, the preparation of both the Test Directive and the Test Plan, and in a more general application, to other personnel who have a need for information concerning the history of a defense system's conception, development and test history.

2. EXPLANATION OF TERM

Review Documentation - The term "review documentation" is used herein as a generic term referring to the documentary evidence of the validated need, development, test and operational employment of a system. For a given system, the term encompasses documentation related to:

a. The initiation of the material acquisition cycle for the system.

b. The progression of system development, including DT&E and OT&E, through the phases of the acquisition cycle.

c. The identification, refinement, and translation into test objectives of the critical questions and issues and the studies, analyses, and test programs responding thereto.

d. Other system history documents, as appropriate; the review of which provides a comprehensive view of the development of the system and its operational history.

3. APPLICATION

Background - The conduct of OT&E usually involves the commitment of considerable resources such as time, personnel, equipment and facilities. The commitment of resources means that OT&E costs money - a great deal of money - and therefore, OT&E programs must be carefully managed from the outset, based on valid test

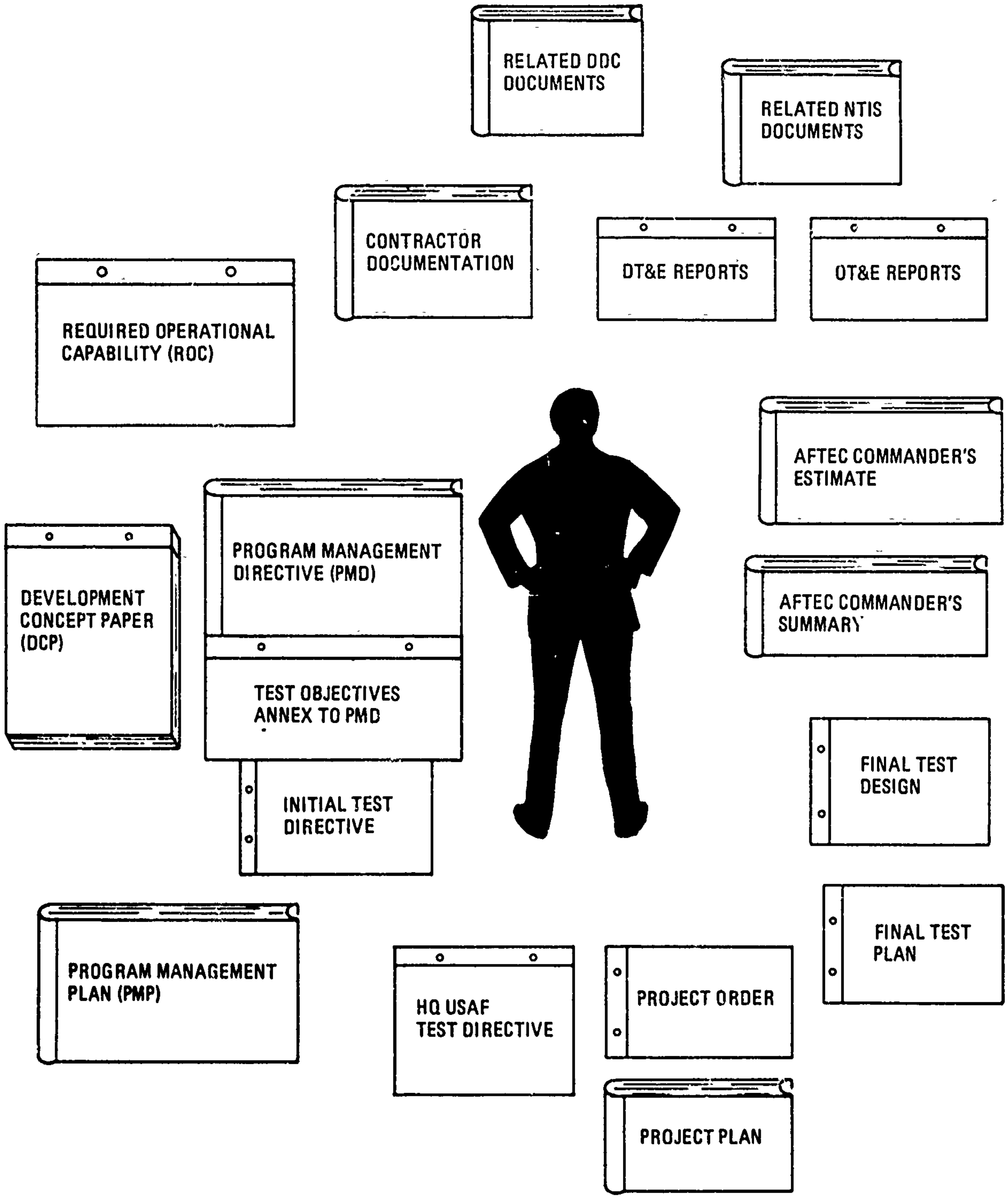


Figure 3-1. Sources of review documentation for the OT&E planner

requirements, and verified, advocated, planned and executed by knowledgeable, competent managers. It is important, in the preparation of the Test Directive and Test Plan, that responsible managers have a clear understanding of the current critical questions/issues, data requirements, and the test objectives. It is no less important that they be familiar with the system and how it operates, and with the past history of the system, especially in previous testing and evaluation. This is the function of review documentation.

Some of the review documentation data sources available to the test planner are depicted in Figure 3-1. The availability of specific documents, in most instances, will depend on how far along the system has evolved with regard to the material acquisition life cycle. For systems under conceptual study, review documentation will be limited to the Required Operational Capability (ROC) and such requirements documents and management and program documents as have been produced to date, as well as review documentation on similar systems. On the other hand, the review documentation for systems which have been deployed in operational use should be plentiful, ranging from the ROC through the various management and program documents and including documentation of testing during system deployment.

4. APPLICABLE DOCUMENTS

Program Documents. For purposes of this section on review documentation, the documents recommended herein for review by the test planner are listed below. Documents denoted by an asterisk are discussed in detail within this section.

*Required Operational Capability (ROC)

- *Combat ROC
- *Quick Reaction Capability (QRC)

*Development Concept Paper (DCP)

- *Program Memorandum (PM)

*Program Management Directive (PMD)

- *Test Objectives Annex to the PMD

*Program Management Plan (PMP)

*OT&E Test Directive

AFTEC Commander's Estimate

OT&E Test Plan

OT&E Test Report

Joint Test Plans and Reports

(OT&E Test Plan, OT&E Test Report, Joint Test Plans and Reports: detailed information available in Chapters 11, 15, and 16 of this document)

AFTEC Commander's Summary

Additional Review Documentation. Additional appropriate review documentation exists for systems which have progressed into the production and deployment phases of the acquisition life cycle. Among such additional review documentation may be:

(a) Contractor's reports and literature on the system.

(b) FOT&E Test Directives, Test Plans and Reports.

(c) Review documentation applicable to the system and current critical questions and issues. Also, documentation concerning the testing of similar systems, available from Defense Documentation Center (DDC) or National Technical Information Service (NTIS).

Note: Access to DDC documents is through a user number code which is easily acquired through appropriate channels. Bibliographic searches are also available from DDC which are initiated on the basis of a telephone call to the DDC center. AFR 80-14 mandates that Test Reports be submitted to DDC, therefore, it becomes one prime source of Test Reports as well as analysis studies, simulations, etc. Access to these DDC reports is by way of a key wording system governed by the DD 1473 form submitted by the Test Officer with his report and by DDC's Theasaurus of Scientific and Technical Terms. Many Air Force technical libraries have microfiche reports readily available.

5. THE UTILITY OF REVIEW DOCUMENTATION

Test Planning. Commencing with the tentative identification of a requirement for conducting OT&E, a coordinated effort is normally undertaken within the user command, the staff of the responsible MAJCOM, (and for issues involving Air Force management interests), AFTEC and HQ USAF. This coordinated effort, directed by a Project Officer within AF/X00WD, has as its objective the validation of the testing requirement, the formulation of test objectives, and the preparation of a scheme for conducting the OT&E, including the identification and allocation of resources and facilities required. Upon completion of the foregoing, and assuming the successful advocacy of the proposed OT&E a Test Directive is drafted by AFTEC, reviewed by HQ USAF, and upon approval at that level is published and distributed authorizing and directing the planning, conduct and reporting of the OT&E.

The Test Directive sets in motion the activities leading to the generation of a Test Design which is the basis for the Test Plan. The Test Plan provides a time-phased integrated and coordinated document which contains the complete detailed plan for the OT&E.

Each of these necessary OT&E preparatory functions involves the use of review documentation. For example, in the preparation of the Test Directive, the Project Officer is directed* to:

a. "Review previous attempts to collect data of the type requested and the results, if any, of such attempts."

b. "Determine the extent to which test objectives have already been fulfilled through conduct of IOT&E or other applicable tests."

c. "Review and analyze pertinent factors concerning the system, subsystem or equipment involved. This includes the mission, terrain, weather, human factors, environmental factors, and enemy capabilities/tactics applicable to the OT&E problem or requirement."

*(AF/X00WD Deputy Director Operating Instruction (DDOI) No. 10-2)

Other Levels. The utility of review documentation is not restricted to the preparation of the Test Directive. At all levels, prudent personnel concerned with the general planning and designing of the OT&E will find review documentation useful either directly or indirectly in the following matters of consideration:

a. Familiarity with the system (or comparable system) and its operation and performance characteristics.

b. Knowledge of what the system is designed to respond to (the operational deficiency or need).

c. The test history of the system including previously employed test design and methodology, and the data results of testing.

(The avoidance of redundant test and data collection).

(The identification of data from previous tests which is applicable to the problem at hand.)

d. The relationship or earlier system problems (critical questions and issues) to new.

e. Difficulties in previous testing and insights into avoiding such difficulties in current testing and data collection.

f. Estimating requirements for test resources, viz: facilities, personnel, time, money, test items.

g. The previous findings/conclusions of the developer, DT&E and OT&E tester, the user, trainer and logistician.

6. PRESENTATION OF REVIEW DOCUMENTATION MODULES

The following page presents modules providing a "quick-look" at the characteristics of specific review documentation. Figure 3-2 provides the modules in abbreviated outline form in chronological order of issuance. This is followed in Appendix A by the presentation of individual review documentation modules in somewhat more detail.

(1)

REQUIRED OPERATIONAL CAPABILITY (ROC)

PURPOSE. FORMALLY STATE AN OPERATIONAL DEFICIENCY OR NEED

CONTENT: 1. APPROXIMATELY 20 PAGES. DESCRIPTION OF DEFICIENCY OR NEED.

2. MISSION OR TASK AFFECTED.

3. HOW IDENTIFIED.

ORIGINATOR: ANY ECHELON OF AF OR DOD.

APPROVED BY: 1. GENERAL OFFICER BILLET INCUMBENT.

2. HQ USAF (RDQLM)

(2)

DEVELOPMENT CONCEPT PAPER (DCP)

PURPOSE. PROVIDE CONTINUALLY UPDATED CONTRACT BETWEEN SEC DEF AND HQ USAF. PRESENTS COMPREHENSIVE REVIEW OF PROGRAM ELEMENTS. FORMS BASIS FOR DECISIONS BY DSARC AND SEC DEF.

CONTENT: 1. ISSUES FOR DECISION.

2. PROGRAM PURPOSES.

3. ALTERNATIVES.

4. REQUIRED RESOURCES.

5. SCHEDULES (DEVELOPMENT/PRODUCTION).

6. RISKS/PROGRAM DIFFICULTIES/CRITICAL ISSUES.

7. THRESHOLDS.

8. DECISIONS/SIGNATURES.

ORIGINATOR: HQ USAF (PROGRAM ELEMENT MONITOR IN THE RESPONSIBLE DIRECTORATE)

APPROVED BY: 1. CSAF

2. SECDEF

(3)

PROGRAM MEMORANDUM (PM)

PURPOSE AN OSD DOCUMENT WITH SIMILAR FORMAT, AND COORDINATION AS THE DCP. A CONTR[...] DOCUMENT BETWEEN OSD AND AIR FORCE [...] CANT PROGRAMS, BUT WHICH ARE NOT SU[...] SPECIFIED DCP ACTION.

CONTENT: 1. ESSENTIALLY SAME AS DCP.

ORIGINATOR: HQ USAF (PROGRAM ELEMENT MONITOR IN [...] RESPONSIBLE DIRECTORATE)

APPROVED BY: 1. CSAF

2. OSD

(5)

TEST OBJECTIVES ANNEX (TOA) TO PMD

PURPOSE: PROVIDE BASELINE INFORMATION, GUIDANCE AND DIRECTION FOR THE DEVELOPMENT, CONDUCT AND REPORTING OF A TEST PROGRAM FOR A MAJOR SYSTEM IN ACQUISITION.

CONTENT: 1. CRITICAL QUESTIONS AND ISSUES.

2. BACKGROUND INFORMATION FOR EACH CRITICAL QUESTION AND ISSUE.

3. TEST OBJECTIVES.

4. TASKING OF COMMANDS FOR T&E.

5. TEST SUPPORT RESOURCES AND REQUIREMENTS.

6. GENERAL REPORTING REQUIREMENTS.

ORIGINATOR: 1. AFTEC PROVIDES THE OT&E INPUTS TO THE TOA.

2. THE IMPLEMENTING COMMAND PROVIDES THE OT&E INPUTS TO THE TOA.

APPROVED BY: DIRECTORATE LEVEL GENERAL OFFICER, DCS, R&D, HQUSAF.

(6)

HQ USAF TEST DIRECTIVE

PURPOSE: A DOCUMENT DRAFTED BY AFTEC FOR HQ USAF WHICH DIRECTS THE CONDUCT OF SPECIFIC OT&E BY OPERATIONAL ELEMENTS.

CONTENT: 1. BACKGROUND.

2. PURPOSE.

3. CONCEPT OF OPERATION.

4. TEST OBJECTIVE.

5. DESCRIPTION OF TEST ITEMS.

6. SPECIAL PLANNING FACTORS.

7. PARTICIPATING AGENCIES AND RESPONSIBILITIES.

8. GENERAL REQUIREMENTS.

9. MILESTONE SCHEDULE.

10. REPORTS.

ORIGINATOR: AFTEC DIRECTORATE FOR T&E.

APPROVED BY: HQ USAF/XOO

(7)

PROGRAM MANAGEMENT PLAN (PMP)

PURPOSE: PRINCIPAL MANAGEMENT BASELINE DOC[...] DEPICTS INTEGRATED TIME-PHASED TA[...] RESOURCES REQUIRED.

CONTENT: 1. PROGRAM SUMMARY, INTELLIGENCE.

2. PROGRAM MANAGEMENT.

3. SYSTEM ENGINEERING.

4. TEST AND EVALUATION (BASED ON C[...] ISSUES).

5. COMMUNICATION/ELECTRONICS.

6. OPERATIONS/MANPOWER/TRAINING/[...]

7. LIST OF BASELINE DIRECTIVES.

ORIGINATOR: PROGRAM MANAGER

APPROVED BY: PROGRAM MANAGER

(9)

AFTEC COMMANDER'S ESTIMATE

PURPOSE: PROVIDE HQ USAF AND INVOLVED AND INTERESTED MAJCOMS WITH PLANNING INFORMATION CONCERNING SCHEDULED OT&E.

CONTENT: 1. SUMMARY OF OT&E TEST OBJECTIVES.

2. SCHEDULES FOR OT&E.

3. RESOURCES.

4. ANTICIPATED PROBLEM AREAS.

5. OTHER SIGNIFICANT INFORMATION.

ORIGINATOR: AFTEC TEST DIRECTOR

APPROVED BY: 1. THE DIRECTOR OF TEST AND EVALUATION.

2. COMMANDER, AFTEC

(10)

OT&E TEST REPORT

PURPOSE: TO PRESENT THE RESULTS OF OT&E TO SATISFY THE DATA REQUIREMENTS OF THE TEST DIRECTIVE.

CONTENT: 1. REPORT OF TEST ACTIVITY
- a. INTRODUCTION
- b. PURPOSE OF OT&E
- c. METHOD OF ACCOMPLISHMENT
- d. DISCUSSION AND ANALYSIS
- e. SUMMARY

2. ANNEXES A-N, PROVIDING DETAILS OF: DESCRIPTION OF TEST ITEM/TACTICS/DOCTRINE, OT&E ENVIRONMENT, TEST METHODOLOGY, DESIGN, SUPPORTING DATA AND ANALYSIS, PRETESTING REQUIRED/ACCOMPLISHED, TEST ORGANIZATION, TEST OPERATIONS, MAINTENANCE, LOGISTICS AND SUPPLY, PERSONNEL, TRAINING, SAFETY, SECURITY, OTHER MATTERS.

ORIGINATOR: TEST DIRECTOR (AFTEC OR COMMAND AS APPROPRIATE)

APPROVED BY: 1. COMMANDER, AFTEC

2. HQ USAF

(11)

AFTEC COMMANDER'S SUMMARY

PURPOSE: PROVIDE THE AFTEC COMMANDER'S REVIEW, CONCURRENCE OR NONCONCURRENCE AND [...] OF THE FINAL REPORT OF AFTEC OR COMMA[...] CONDUCTED OT&E.

CONTENT: 1. OVERVIEW AND CRITIQUE OF THE OT&E [...]

2. RESULTS, RECOMMENDATIONS, CONCLU[...] ARE CONCURRED IN.

3. AREAS OF NONCONCURRENCE.

4. REMARKS, RECOMMENDATIONS, CONCLU[...] THE AFTEC COMMANDER CONCERNING THE [...] OF THE OT&E AS CONDUCTED AND REPORT[...]

ORIGINATOR: AFTEC TEST DIRECTOR

APPROVED BY: 1. AFTEC DIRECTORATE OF T&E

2. COMMANDER, AFTEC

(2)

DEVELOPMENT CONCEPT PAPER (DCP)

PURPOSE	PROVIDE CONTINUALLY UPDATED CONTRACT BETWEEN SEC DEF AND HQ USAF. PRESENTS COMPREHENSIVE REVIEW OF PROGRAM ELEMENTS. FORMS BASIS FOR DECISIONS BY DSARC AND SEC DEF.
CONTENT:	1. ISSUES FOR DECISION. 2. PROGRAM PURPOSES. 3. ALTERNATIVES. 4. REQUIRED RESOURCES. 5. SCHEDULES (DEVELOPMENT/PRODUCTION). 6. RISKS/PROGRAM DIFFICULTIES/CRITICAL ISSUES. 7. THRESHOLDS. 8. DECISIONS/SIGNATURES.
ORIGINATOR:	HQ USAF (PROGRAM ELEMENT MONITOR IN THE RESPONSIBLE DIRECTORATE)
APPROVED BY:	1. CSAF 2. SECDEF

(3)

PROGRAM MEMORANDUM (PM)

PURPOSE	AN OSD DOCUMENT WITH SIMILAR FORMAT, CONTENT, AND COORDINATION AS THE DCP. A CONTRACTUAL DOCUMENT BETWEEN OSD AND AIR FORCE FOR SIGNIFICANT PROGRAMS, BUT WHICH ARE NOT SUBJECT TO SPECIFIED DCP ACTION.
CONTENT:	1. ESSENTIALLY SAME AS DCP.
ORIGINATOR:	HQ USAF (PROGRAM ELEMENT MONITOR IN THE RESPONSIBLE DIRECTORATE)
APPROVED BY:	1. CSAF 2. OSD

(4)

PROGRAM MANAGEMENT DIRECTIVE (PMD)

PURPOSE:	THE OFFICIAL HQ USAF MANAGEMENT DIRECTIVE DURING ACQUISITION TO PROVIDE DIRECTION TO THE IMPLEMENTING COMMAND AND TO SERVE AS A CONTRACT BETWEEN HQ USAF AND THE IMPLEMENTING COMMAND.
CONTENT:	1. DEFINE AUTHORITY OF P.M. 2. RESPONSIBILITIES OF PARTICIPATING COMMANDS. 3. STATE REQUIREMENTS. 4. REQUEST STUDIES/ANALYSES. 5. INITIATE, APPROVE, CHANGE, MODIFY, TERMINATE. 6. SATISFY DOCUMENT NEEDS FOR ADVOCACY, DEVELOPMENT, PRODUCTION, MODIFICATION FUNDED BY RDT&E OR PROCUREMENT FUNDS. 7. MAY INCLUDE TEST OBJECTIVES ANNEX (TOA)
ORIGINATOR:	HQ USAF (THE PROGRAM ELEMENT MONITOR IS THE RESPONSIBLE DIRECTORATE).
APPROVED BY:	DIRECTORATE LEVEL GENERAL OFFICER, DCS, R&D.

(6)

HQ USAF TEST DIRECTIVE

PURPOSE:	A DOCUMENT DRAFTED BY AFTEC FOR HQ USAF WHICH DIRECTS THE CONDUCT OF SPECIFIC OT&E BY OPERATIONAL ELEMENT.
CONTENT:	1. BACKGROUND. 2. PURPOSE. 3. CONCEPT OF OPERATION. 4. TEST OBJECTIVE. 5. DESCRIPTION OF TEST ITEMS. 6. SPECIAL PLANNING FACTORS. 7. PARTICIPATING AGENCIES AND RESPONSIBILITIES. 8. GENERAL REQUIREMENTS. 9. MILESTONE SCHEDULE. 10. REPORTS.
ORIGINATOR:	AFTEC DIRECTORATE FOR T&E.
APPROVED BY:	HQ USAF/XOO

(7)

PROGRAM MANAGEMENT PLAN (PMP)

PURPOSE:	PRINCIPAL MANAGEMENT BASELINE DOCUMENT. DEPICTS INTEGRATED TIME-PHASED TASKS AND RESOURCES REQUIRED.
CONTENT:	1. PROGRAM SUMMARY, INTELLIGENCE. 2. PROGRAM MANAGEMENT. 3. SYSTEM ENGINEERING. 4. TEST AND EVALUATION (BASED ON CRITICAL ISSUES). 5. COMMUNICATION/ELECTRONICS. 6. OPERATIONS/MANPOWER/TRAINING/SECURITY. 7. LIST OF BASELINE DIRECTIVES.
ORIGINATOR:	PROGRAM MANAGER
APPROVED BY:	PROGRAM MANAGER

(8)

OT&E TEST PLAN

PURPOSE:	TO EXPAND THE OT&E TEST DIRECTIVE INTO A FULLY INTEGRATED TIME-PHASED PLAN FOR THE CONDUCT OF OT&E.
CONTENT:	(MAIN BODY) 1. INTRODUCTION. 2. TEST PURPOSE AND OBJECTIVES. 3. CONCEPT OF TEST OPERATIONS. 4. METHOD OF ACCOMPLISHMENT. 5. TEST SCHEDULE. 6. TEST MANAGEMENT AND ORGANIZATION. 7. RESPONSIBILITIES/SUPPORT. 8. PERSONNEL. 9. REQUIRED TEST REPORTS. 10. SAFETY.
ORIGINATOR:	AFTEC TEST DIRECTOR
APPROVED BY:	1. AFTEC TEST REVIEW GROUP 2. COMMANDER, AFTEC

(10)

OT&E TEST REPORT

PURPOSE:	TO PRESENT THE RESULTS OF OT&E TO SATISFY THE DATA REQUIREMENTS OF THE TEST DIRECTIVE.
CONTENT:	1. REPORT OF TEST ACTIVITY a. INTRODUCTION b. PURPOSE OF OT&E c. METHOD OF ACCOMPLISHMENT d. DISCUSSION AND ANALYSIS e. SUMMARY 2. ANNEXES A-N, PROVIDING DETAILS OF: DESCRIPTION OF TEST ITEM/TACTICS/DOCTRINE, OT&E ENVIRONMENT, TEST METHODOLOGY, DESIGN, SUPPORTING DATA AND ANALYSIS, PRETESTING REQUIRED/ACCOMPLISHED, TEST ORGANIZATION, TEST OPERATIONS, MAINTENANCE, LOGISTICS AND SUPPLY, PERSONNEL, TRAINING, SAFETY, SECURITY, OTHER MATTERS.
ORIGINATOR:	TEST DIRECTOR (AFTEC OR COMMAND AS APPROPRIATE)
APPROVED BY:	1. COMMANDER, AFTEC 2. HQ USAF

(11)

AFTEC COMMANDER'S SUMMARY

PURPOSE:	PROVIDE THE AFTEC COMMANDER'S REVIEW, CONCURRENCE OR NONCONCURRENCE AND EVALUATION OF THE FINAL REPORT OF AFTEC OR COMMAND-CONDUCTED OT&E.
CONTENT:	1. OVERVIEW AND CRITIQUE OF THE OT&E REPORT. 2. RESULTS, RECOMMENDATIONS, CONCLUSIONS WHICH ARE CONCURRED IN. 3. AREAS OF NONCONCURRENCE. 4. REMARKS, RECOMMENDATIONS, CONCLUSIONS OF THE AFTEC COMMANDER CONCERNING THE ADEQUACY OF THE OT&E AS CONDUCTED AND REPORTED.
ORIGINATOR:	AFTEC TEST DIRECTOR
APPROVED BY:	1. AFTEC DIRECTORATE OF T&E 2. COMMANDER, AFTEC

(12)

NOTES

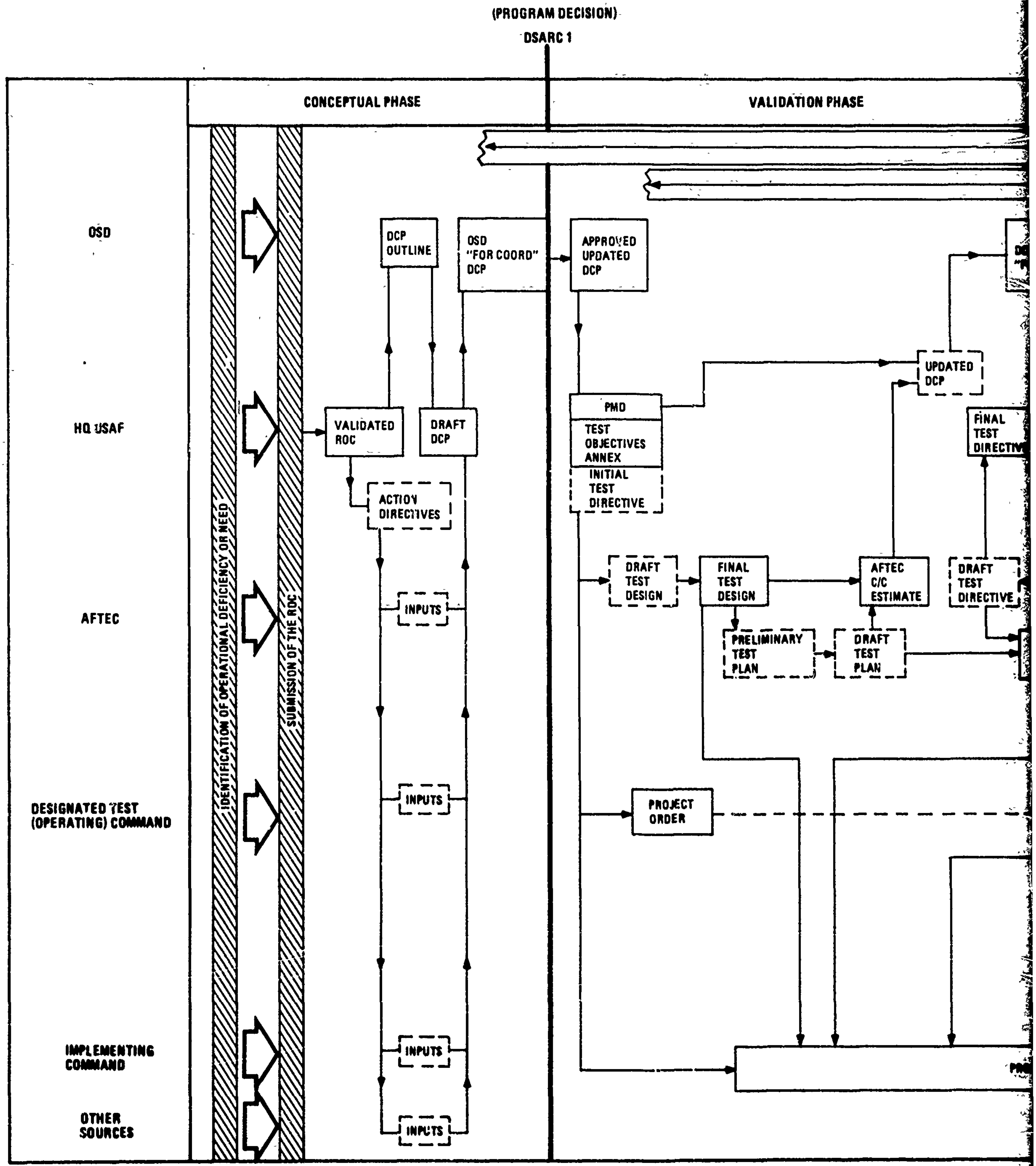
1. MANAGEMENT APPROACH AND PHILOSOPHY ARE ESSENTIALLY THE SAME FOR MAJOR AND MINOR SYSTEMS.
2. THE DIFFERENCES BETWEEN MAJOR/MINOR SYSTEM DOCUMENT FLOW ARE DETERMINED BY THE LEVEL FOR REVIEW/APPROVAL, PROGRAM RATIFICATION AND PRODUCTION DECISIONS.
3. IN SMALLER SYSTEM ACQUISITION WHERE PROGRAM THRESHOLDS DO NOT REQUIRE OSD ACTION, THE DOCUMENT FLOW WILL NOT REQUIRE THE DCP, ADVOCACY DOCUMENTS OR DSARC ACTIONS.

Figure 3-2. Quick Look Modules

7. SOME SOURCES OF REVIEW DOCUMENTATION

Sources of review documentation are widespread and depend somewhat on the nature of the system. The first search effort should be directed to command files (primarily, the technical library). Some other general sources of review documentation are listed as follows:

	Review Document	Source
a.	Program documentation for systems up to the DSARC 1 decision milestone (ROC, formal directives)	-HQ USAF/RDQ
b.	Program documentation for systems after DSARC 1	-HQ USAF/RDP
c.	Program Management Plans (PMP)	(1) For systems in development, refer to "INDEX OF ACTIVE USAF DIRECTED EFFORTS" RCS-HAF-RDE(0) 7103 for the identification of the responsible System Project Office under AFSC. (2) For deployed systems, refer to "INDEX OF ACTIVE USAF DIRECTED EFFORTS" for the identification of the responsible ALC under AFLC.
d.	OT&E Test Directives	(1) HQ USAF/X00WD (2) AFTEC (3) MAJCOM under which OT&E was conducted.
e.	OT&E Test Design/Test Plans/Test Reports (major and selected nonmajor programs)	(1) AFTEC
	(other nonmajor programs)	(2) MAJCOM under whom OT&E was conducted (3) DDC/NTIS



(PROGRAM DECISION)
DSARC 1
CONCEPTUAL PHASE
VALIDATION PHASE
OSD
HQ USAF
AFTEC
DESIGNATED TEST (OPERATING) COMMAND
IMPLEMENTING COMMAND
OTHER SOURCES
IDENTIFICATION OF OPERATIONAL DEFICIENCY OR NEED
SUBMISSION OF THE ROC
DCP OUTLINE
OSD "FOR COORD" DCP
APPROVED UPDATED DCP
VALIDATED ROC
DRAFT DCP
ACTION DIRECTIVES
INPUTS
INPUTS
INPUTS
INPUTS
PMD
TEST OBJECTIVES ANNEX
INITIAL TEST DIRECTIVE
UPDATED DCP
FINAL TEST DIRECTIVE
DRAFT TEST DESIGN
FINAL TEST DESIGN
AFTEC C/C ESTIMATE
DRAFT TEST DIRECTIVE
PRELIMINARY TEST PLAN
DRAFT TEST PLAN
PROJECT ORDER

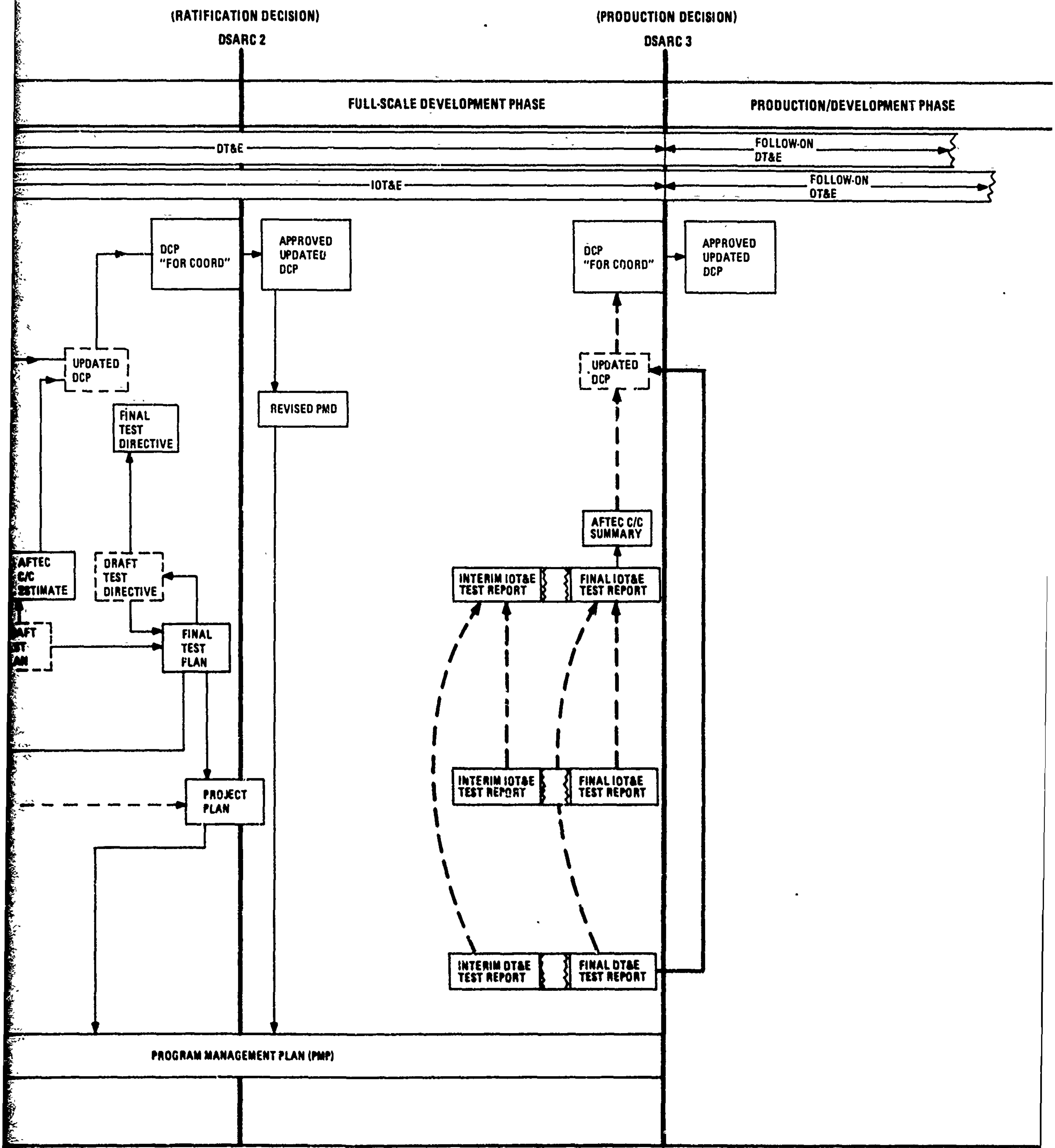


Figure 3-3. Flow of documentation applicable to IOT&E in the acquisition cycle

f.	DT&E Test Plans/Reports	(1) PMP for system (2) Appropriate SPO (3) Appropriate ALC (4) Contractor
g.	Contractor Documentation (technical reports, papers, etc)	(1) Contractor (2) DDC/NTIS
h.	AFTEC Estimate/AFTEC Summary	(1) AFTEC (2) Command Files
i.	Combined IOT&E/DT&E (CIDT & E) Test Plan	(1) Appropriate SPO (2) Contractor (3) DDC
j.	SPO - ALC Transition System Documentation	- Appropriate ALC
k.	Reliability and Maintainability (R&M) Studies (Estimates); LSC and LCC modeling efforts	- AFLC
l.	System Technical Orders (T.O. s)	- USAF Tech. Order Distribution System

CHECKLIST FOR REVIEW OF DOCUMENTATION

1. Identify and compile the referenced correspondence, reports of specific studies or analyses, and the HQ USAF Test Directive, if appropriate, which addresses the new requirement for data on a system/equipment.

2. Identify the critical questions and issues, if applicable to the data requirement, which provide the basis for the consideration for test.

3. As appropriate, identify from the Test Directive or formulate, if necessary, the test objectives which respond to the critical questions and issues and the data requirement.

4. Acquire review documentation on the same system from sources identified in the Chapter 3 supplemented as necessary with review documentation of like systems. Such review documentation should consist of the following, assuming that the system is deployed in operational use:

   a. The documentation identified in 1 above.

   b. Other requirements documentation (the Required Operational Capability (ROC) for the system.)

   c. PMD's issued by HQ USAF and command supplements thereto.

   d. The DCP/PM

   e. The Program Management Plan (PMP)

   f. The HQ USAF Test Directives for IOT&E and for tests conducted in FOT&E, as appropriate.

   g. The Test Plans and Test Reports for OT&E (both IOT&E and FOT&E) previously conducted on the system.

   h. The AFTEC Estimates and Summaries concerning previously conducted OT&E.

   i. DT&E Test Plans and Test Reports.

   j. AFLC generated reliability and maintainability (R&M) studies (estimates) and LCC and LSC modeling efforts.

   k. Air Force Technical Orders (T.O.s) applicable to the system.

5. Review the documentation in order to gain familiarity with the history of the system, its functions and characteristics.

6. From a review of documentation on previous testing, determine the following:

   a. Are the current issues addressed in previous documentation?

   b. Were attempts previously made to acquire the same or similar data to that which is currently needed? What were the results?

   c. What problems were surfaced in previous testing which relate to the current consideration for test?

      (1) Based on previous test history and capabilities of test facilities and commands, is the current test technology capable of providing the required data?

      (2) Does the system/equipment lend itself to a test scheme which will provide the data needed as indicated by previous testing?

   d. Are there requirements addressed in previous testing concerning cost, personnel, facilities, test items, time, etc. which are useful in the estimation of requirements relative to the current issue?

7. What constraints, evidenced by previous test history, have a possible relation to the current issue.

8. Are there test techniques, schedules, time-phased events, data collection/reduction/analysis tools or schemes used in previous testing which appear to have utility in the present consideration for testing?

9. Review the concept of operations for the system (it's in the PMP, among other places). Is the system designed to operate in the environment and in the performance envelope which the current test concept envisions?

10. Have you so familiarized yourself with the characteristics and operation of the system (or, in the absence of significant review documentation for a specific system, with a comparable system) and its test and operational history as to be confident in your ability to make competent decisions for which you are responsible?

Chapter 4

FORMULATION OF TEST OBJECTIVES

1. INTRODUCTION

The development of test objectives and related planning for Operational Test and Evaluations (OT&E) is a process in which all levels of the Air Force become involved. In the OT&E of major weapon/support systems this process is extended to OSD level through the process of Development Concept Paper (DCP) review, and Defense System Acquisition Review Council (DSARC) procedures. Because of operational interfaces, OT&Es may include participation of other military services or civilian agencies in test planning and the setting of test objectives. The military significance of most OT&E makes it critical to the national defense. Accordingly, planning and test objective development must be timely, thorough, and clearly communicated to all participants to assure that the operational information and data obtained from an OT&E is that needed to satisfy the basic purpose of conducting the test and for making production and/or operational employment decisions.

Development of test objectives is one of the most important steps in planning OT&E. No matter how expertly OT&E is designed and performed, the operational information and data necessary for development, production, support, and operational decisions will not be obtained unless the right test objectives were specified at the onset. Planning and development of test objectives must proceed together to assure that all planning actions support OT&E objectives and that the objectives established are achievable within the allowable time frame and resources available.

The USAF has recently activated an Air Force Test and Evaluation Center (AFTEC) which will be involved in the planning and development of test objectives.

In discussing the planning process and development of test objectives, it must be recognized that OT&E as practiced in the Air Force varies greatly as to who initiates requirements, organizational and individual responsibilities and involvement, lead time, resources, complexity, and relative importance. The common elements of planning and development of test objectives for the various types of OT&E activities are discussed below. Variations in the details of planning and test objective development will be discussed which address OT&E in general groupings of activities associated with one of the following:

1. Acquisition of new weapon/support systems and equipment. These testing activities are given primary emphasis in DODD 5000.1, DODD 5000.3, AFR 80-14, and other Air Force regulations and directions.

2. Modification of existing operational systems and equipment to provide a new or improved combat capability in accordance with the provisions of AFR 57-4.

3. Comparison of operational capabilities and interfacing of weapon/support systems possessed by two or more of the military services. In some instances, such comparisons may be limited to Air Force systems or equipment having similar capabilities. When several services are involved, these OT&Es are usually referred to as Joint Tests and originate with a DOD or JCS level directive. Other Joint Tests are directed which involve the exercise of operational systems in collection of basic operational performance and capability data for use in conducting studies and analyses, and for identifying needs and characteristics desired in future systems.

4. Investigation and development of new tactics, techniques, and applications of systems and equipment in the operational inventory.

OT&E programs generally proceed through a preliminary planning stage followed by a detailed planning stage during which the Test Plan is prepared, resources assembled or programmed, and all necessary actions are taken to initiate test execution. Often the initial planning on high priority programs overlaps and merges into the detailed planning phase. For convenience, however, this process will be assumed to be essentially sequential.

2. PRELIMINARY TEST PLANNING

Preliminary planning takes place during the preparation of acquisition documentation, test directives, and orders. It identifies and initiates long term test resource programming actions. It is usually a DOD, HQ USAF and AFTEC, or Major Command headquarters function. Test personnel from operational units may be called upon to contribute and to provide information for use in planning and developing objectives. Preliminary planning includes analysis of and action on the following:

1. The basic requirement and overall purpose for conducting the test and evaluation; and the operational mission of the system/equipment including the mission tasks/factors involved. This is done to confirm the need for the test and to assist in focusing planning and test attention on the development of objectives which will produce an OT&E designed to provide the operational information and data needed for the decision process.

2. General test and evaluation objectives specified by DOD Directives, AFR 80-14 and other Air Force regulations/directives; and critical questions and issues in applicable Development Concept Papers must be translated into operational test objectives. Since DCPs are drafted at Air Staff level (with AFSC inputs), these critical questions and issues are normally prepared by members of the staff responsible for OT&E matters. Relative priority for answering each objective and critical question/issue should be established at this time. Additionally, determination of which test objectives should be addressed during Initial OT&E (IOT&E) and which should be covered during Follow-On OT&E (FOT&E) as defined by AFR 80-14 is necessary.

3. Information and data available from earlier tests of similar systems/equipment, operational experience, or during Developmental Test and Evaluation which would satisfy any of the OT&E objectives.

4. Test objectives in earlier tests of similar systems/equipment which are currently valid; and which, if used, would tend to standardize and expand the data base.

5. The environment in which the test should be conducted to provide the required orientation, quality and quantity of operational data. The test environment normally includes: appropriate target(s)/terrain characteristics and climatic conditions in which operations will be conducted, interface with systems/operations of other Air Force elements and/or Service(s), simulated enemy threat, defined qualification and skills of operating and supporting personnel, and the operational/maintenance doctrine and procedures of organizations responsible for employing the system/equipment and tactic/technique.

6. Test resources required and available to include: test items, including support systems; spares; maintenance and logistic support; test and test support personnel; test ranges; practice and maneuver areas; and any special funds over and above system/project or Command O&M funds.

7. Review of inputs and actions required in connection with Development Concept Papers (DCP), Program Management Directives (PMD), Program Management Plans (PMP), OT&E Test Directives, and Test Orders.

8. Long range training requirements for test and test support personnel.

9. Relative priority of the OT&E effort in comparison with other OT&Es which may be in competition for the particular test resources involved.

10. Data collection/reporting requirements.

11. Coordination requirements and identification of the primary Operating Command and all participating Operating/Supporting Commands. Except for OT&E of support equipment such as life support items, communications gear, etc., identification of the primary Operating Command that should be directed to conduct the OT&E is self evident or readily identified by review of requirements and other programming documentation.

12. Schedule and location for test coordination meetings and identification of participants.

13. Requirements for mathematical modeling and computer simulation techniques for: obtaining information and data needed to achieve evaluation objectives; to guide conduct of the OT&E by prediction of performance capabilities and limitations; or, to identify critical areas of investigation.

14. Applicability and desirability of retaining Initial OT&E test team members to provide continuity in the Follow-On OT&E with production items.

15. Schedule for presenting and publishing test planning data in DCPs, PMDs, PMPs, OT&E Test Directives, and Test Orders.

16. Requirement for an initial combat (or operational unit/theater) introduction test and evaluation phase of the Follow-On OT&E.

While the foregoing list of preliminary planning activities is not all inclusive, it provides an indication of the range of items which must be considered and the actions which must be taken to assure a smooth transition to detailed test planning and development of specific test objectives by the operational unit assigned responsibility for publishing the Test Plan, conducting the test, and reporting test results. Early identification and action on as many planning factors as possible works to an advantage in several ways. For example, it might be found that all desired test objectives probably could not be achieved because of early identifiable limitations of resources and time. Under such circumstances, test objectives can be readily modified at the test directing level in contrast to the substantial program delays and wasted efforts that might result if a problem were not recognized until detailed planning was underway in an operational unit. Arrangement for participation by another Military Service is another example of an action that should be identified and initiated during preliminary planning since long lead times and high level coordination are usually involved.

3. DETAILED TEST PLANNING

Detailed test planning is accomplished in connection with preparing the Test Plan and normally is performed by the operational test unit assigned responsibility for conducting the test. In event that participation in initial or preliminary planning has taken place, detailed test planning may overlap and time will be gained for programming resources, training test personnel, and refining test objectives. Generally, detailed planning, subject to variations in types of OT&E, will include the following:

1. Critical review, refinement, and development of the test objectives contained in directive documentation into specific objectives that test designers can translate into Measures of Effectiveness and measurement/analysis requirements. Recommendations should be made to the test directing authority for deletion of any test objectives that are not realistic or not achievable because of resources, cost or time limitations. Additional test objectives should be recommended if expansion of the OT&E appears desirable to investigate any additional areas of operational effectiveness and suitability.

2. Organizing the test design team and establishing planned completion dates for an outline of test methodology including appendices thereto, such as; flight profiles, summaries of test measurements to be accomplished during each mission, data collection and processing plans, and analysis techniques. Early initiation and completion of this effort is mandatory because it is directly related to nearly every test planning and programming effort.

3. A pretest planning meeting conducted as soon as practicable after receipt of test directing documentation. It will be convened and chaired by the Test Director or project officer and attended by working-level personnel from the operational and supporting units who will be involved in the test to clarify and detail the method of conducting the test.

4. Development of a test program schedule to include initiation and completion dates for:

    a. Pretest planning meeting

    b. Each planning activity

    c. Completion of test design

d. Resource programming and availability dates

e. Test Plan preparation, coordination, and approval

f. Training and familiarization

g. Safety review and pretest coordination meeting

h. Physical testing

i. Interim and final reporting

j. Introduction team deployment, return, and report

5. Assignment of responsibilities for drafting each section of the Test Plan and development of a schedule for draft completion, coordination, and publication.

6. Coordination with AFLC, ATC, and other Supporting Commands and participating Services to determine their specific data requirements and desired participation in IOT&E and FOT&E.

7. Identification of aircrew and other project personnel by name and making arrangements for appropriate training and familiarization. Identification of test support requirements and preparation of operating procedures related to the handling of test aircraft and scheduling of mission support.

8. Identification of logistics support requirements, and preparation of procedures related to maintenance and supply support and collection of maintainability and reliability data. Planning should include identification of organizational-level maintenance and supply support operations, and those normally performed at base level or above.

9. Estimating resource costs (including test range costs).

10. Estimation of TDY requirements by the test agency and submission of funding requirements.

11. Identification of long-lead-time test support requirements that are not possessed by the test unit and submission of requests to test support agencies. Because of the time periods involved, it is desirable to initiate these actions based on participation in preliminary planning prior to receipt of test directive documentation provided authority to do so has been given or can be obtained. These test support requirements include:

a. Technical support by specialized organizations or individuals who are required to assist in test planning, design, conduct, and reporting.

b. Operational support, both airborne and ground-based, required in conduct of the test. This support is normally available from field units and can be obtained through Operating Command channels, but will require requests through higher headquarters if support from overseas commands, other Services, or other governmental agencies is necessary to duplicate intended operational environment of the system/equipment/concept being tested.

c. Threat and target support.

d. Technical facility and services support.

e. Logistic facilities and support inherent to the support base from which the test system/equipment will be operated, which will be required by units supporting test operations on TDY basis, or which are unique to the design of the test and resulting test operations.

12. Finalization of personnel requirements - skills, numbers, experience, and in-place dates - in the areas of:

a. Aircrew and system/equipment operations.

b. Maintenance and technical support.

c. Test support.

d. Data reduction and analysis

13. Identification and scheduling of equipment modifications (primarily Class II) required for: instrumentation such as data boxes, recorders, telemetry packages, beacons, cameras; and installation of provisions for carrying and releasing special stores.

14. Development of an Information Plan.

15. Development of safety plans and conduct of safety review/pretest coordination meeting chaired by the Test Director and attended by representatives of all agencies participating in the test plus representatives from test units' Office of Safety.

16. Development of a plan for introducing the new system/equipment/concept to other operational units and overseas theaters. Planning should include: identification of team members; units and overseas bases to be visited; operational information and data collection requirements; resource requirements; and initiation of programming actions and theater clearance procedures.

17. Preparation of an Environmental Impact Statement for tests not included within the established ranges.

NOTE: See Chapter 11 for the actual standard Test Plan format and instructions for its preparation.

4. DEVELOPMENT OF OT&E OBJECTIVES

Development of test objectives constitutes the initial step in the preliminary as well as in detailed planning of an OT&E. In practice, it is a continuing process in which objectives are refined as test planning and design considerations progress and required coordination, review, and approval procedures take place. Test objectives set forth during preliminary planning are usually general in nature. As preliminary planning progresses into detailed planning by an OT&E agency, the general test objectives are broken out into the specific objectives published in a Test Plan and provide test designers with the basis for development of Measure of Effectiveness, data requirements, and designing test missions. New planning factors generated as a result of analysis of mission results during test execution, unforeseen test item or test support limitations, or special interest taken in the test at higher organizational levels may result in deletion, revision, or addition of test objectives.

Basically, OT&E objectives are infinitive statements (i.e., to measure, to observe, and/or to evaluate) for the purpose of answering critical questions and issues, and to provide a basis for making decisions affecting development, production, support, and employment of a weapon/support system or item of equipment.

The following infinitive statements were extracted from past USAF OT&E reports and are typical of OT&E objective key phrases:

To observe the degradation. . .

To collect data on. . . .

To obtain information. . .

To provide early system knowledge. . .

To determine feasibility of. .

To evaluate . . .

To confirm proper operation of . . .

To determine the capability to perform . . .

To evaluate ability to meet the requirement . . .

To identify and document deficiencies . . .

To determine characteristics of . . .

To compare characteristics . . .

To determine if possible to change characteristic . . .

To develop concept . . .

To determine method for . . .

To determine bounds on employment . . .

To determine amount of support for . . .

To determine compatibility with . . .

To determine degradation when . . .

To evaluate trade-offs of . . .

To select replacement for . . .

To measure the effect of . . .

The use of "To Verify" should be avoided as it implies a preconceived result. Review of the broad spectrum of test and evaluation activities covered by OT&E reveals that complete standardization of objectives is not possible; however, opportunities appear to exist for greater standardization in the development of objectives within each of the several areas of OT&E activity. These opportunities are to be found primarily in the proper application of the current regulations, directives, and procedures which provide the principal sources of OT&E objectives. Better standardization within each general grouping of OT&E should result in greater predictability of data and information yield; a more thorough coverage of data and information requirements; and a better data base for force structure studies, comparison of systems, and analysis of their capabilities in various operational environments.

The principal sources for the development of test objectives are listed and discussed below.

Objectives Prescribed In Regulations and Directives

Basic regulations and directives usually considered are DOD D 5000.1 "Acquisition of Major Defense Systems," DOD D 5000.3 "Test and Evaluation," AFR 80-14, other Air Force functional area regulations, directives, and office instructions. Added to these are numerous Major Command regulations and guidance documents that are applicable, in whole or part, to OT&E. The DOD directives noted above emphasize data and information needs for use in making development and production decisions during the acquisition of major weapon and support systems. The primary source of such information and data is the initial phase of OT&E specified in DOD 5000.3 and identified as Initial Operational Test and Evaluation (IOT&E) in AFR 80-14. AFR 80-14 is basically consistent with DOD 5000.3. AFR 80-14 specifies general test objectives for DT&E and OT&E which is divided into two parts: IOT&E and FOT&E. AFR 80-14 notes that "OT&E may continue through the life cycle of a system or item of equipment as hardware improvements are developed, new uses are devised, modifications have caused a major change in performance, or new operating environments are encountered." Specific OT&E objectives listed in AFR 80-14 cover a much more limited portion of the system life cycle. They appear primarily appropriate to the continuation of weapon/support system OT&E as it progresses from IOT&E with a limited number of test items during system validation to full-scale system testing with production configured items entering the operational inventory.

Many other regulations, directives, and guidance documents specify collection of data and information for which appropriate test objectives must be developed. Examples are AFR 800-8 on integrated logistics support, AFR 80-46 on personnel subsystems, AFR 80-5 on reliability and maintainability, AFR 80-18 on transportation and handling, AFR 80-38 on survivability, and MIL-STDs pertaining to system acquisition such as MIL-STD 781 which specifies conduct of reliability and maintainability demonstrations during test programs conducted in accordance with AFR 80-14. In addition, internal Air Staff instructions specify responsibilities and procedures for developing critical questions and issues to be answered during testing which are included in DCPs and which are translated into test objectives in PMDs and OT&E Directives. More will be said about critical questions and issues as a source of test objectives in the next discussion section.

Upon initial review, it may appear that Air Force OT&E planners could be overwhelmed by the number of testing requirements and test objectives set forth in the current framework of regulations and objectives. Test planners and designers will find, however, that many of the test requirements and objectives in regulations and directives do not apply to the OT&E for which they are responsible. Generally, a greater number of the regulations and directives have application to large weapon or support system tests than apply to tests of limited scope such as would be involved in investigating a new tactic or technique for employing a system in the inventory. If test planners consistently make a thorough review of all applicable regulations, directives, and standards during test planning, a significant step in standardizing test objectives will have been accomplished and requirements for obtaining essential data and information are less likely to be overlooked.

Objectives From Critical Questions and Issues

AFR 80-14 specifies that critical questions and issues to be addressed during test and evaluations will be reflected in the Initial Development Concept Paper (DCP), Program Memorandum (PM), Program Management Directive (PMD), and Program Management Plans (PMP). It further specifies that Test Plans will translate these critical questions and answers into critical test objectives and make provisions for their accomplishment. These procedures are identified primarily with acquisition of major weapon and support systems and are not a source of test objectives for other OT&Es. DCPs are drafted by the Air Staff in accordance with HQ USAF HOI 800-1. Usually, AFSC and Operating/Supporting Commands are called upon to make inputs to the draft DCP or draft revisions. This participation plus the Joint Operational and Technical Review conducted prior to publication of the initial DCP provides an opportunity for test planners and designers in Operating and Supporting Commands to help develop realistic critical questions and issues for translation into test objectives. One of the approaches used for deriving these questions and issues is by identification of the tasks that the system is expected to perform.

After the DCP has been drafted by the Air Staff it is coordinated throughout the Air Force and submitted to OSD for coordination prior to presentation before the Defense System Acquisition Review Council. Following presentation to the DSARC, the Secretary Defense decision is incorporated into an updated DCP which provides a basis for developing the PMD which may include a Test Objectives Annex (see HQ USAF H0I 800-2). PMD directions are reflected in the Program Management Plan (AFSC Pamphlet 800-3) prepared by the AFSC Program Manager and in the Test Plan for the OT&E (usually IOT&E) prepared by the responsible Operating Command.

Test objectives derived through the process outlined in this discussion are standardized to the extent that they are developed by a formalized procedure which allows for coordination and input at all levels of participation in a major acquisition program. Although the process is fallible to the extent that human beings are involved, chances for gross oversights are minimized by the extensive review and coordination process. An additional degree of discipline is injected by the requirement for presenting results of Test and Evaluation in relation to critical questions and issues at DSARC presentations. Of particular significance from an OT&E standpoint is the independent presentation of the results of Initial OT&E, usually by the OT&E test agency, which is required for use in determining if demonstrated operational effectiveness, suitability and supportability has been sufficiently adequate to commit the systems to production.

Objectives Specified in Program Management and Test Directive Documentation

Program Management Directives noted in the preceding discussion are normally used in directing the conduct of IOT&E which is a part of the normal system acquisition process. As noted, test objectives in the PMD are derived primarily from the statements of critical questions and issues; however, expansion to include objectives specified in AFR 80-14 and Air Force functional regulations is not precluded. OT&E Test Directives are often issued to direct OT&Es and joint or comparison tests specifically directed by HQ USAF. The test objectives contained in these directives may cover the broad range specified in AFR 80-14 and other regulations for

OT&E or production systems; however, objectives met during IOT&E of the pre-production items conducted during system validation are normally not repeated unless modification and improvements incorporated in the production system suggests the need for updated information and data. On the other hand, test objectives in OT&E Test Directives for items of support equipment, joint, or comparison tests usually require a more specific type of test derived information and data. OT&E associated with modified systems and equipment however, may approach the objectives contained in PMDs and OT&E Directives and are patterned, in part, after previous OT&Es on like systems or items of equipment. This tends to achieve a degree of test objective standardization and aids in developing a data base from which comparative information and data can be withdrawn for analysis and study.

PMDs and OT&E Test Directives are distributed to Major Command headquarters. At the Operating Command headquarters, a designated OT&E staff agency may prepare a Test Plan which includes detailed test objectives based on the PMD or OT&E Test Directive and transmit it to operational units or special test units for execution. An alternative is for the Operating Command headquarters to translate the PMD or OT&E Test Directive into a directive document designated as a Test Order which is used by test units as the basis for developing detailed test objectives and Test Plans.

Operating Command Test Orders are also used to direct OT&Es originated by the Command as well as those directed by HQ USAF. The test objectives for tests originated by an Operating Command may be developed by a headquarters OT&E staff agency or they may be developed in draft by one of the Command's test agencies, coordinated and approved by the Command headquarters.

Generally, the greatest degree of standardization can be expected in test objectives developed for major weapon/support systems because of their origin primarily at Air Staff level where they are fit into the framework of acquisition/test regulations and directives, and are subjected to an extensive review and high level approval process. Other OT&Es, particularly those originated and conducted by Operating Commands, are less likely to be standardized because of the **fewer** number of staffs involved, less extensive coordination and approval procedures, and the diversity of test activities involved; however, the existence of Command regulations and guidance documents and the tendency to pattern new tests in the same general mold as completed tests of a like nature tends to maintain a degree of standardization within the Command.

5. OBJECTIVES FOR OT&E OF NEW WEAPON/SUPPORT SYSTEMS AND EQUIPMENT

The planning of an OT&E for a new weapon/support system and the development of objectives for the test normally involves the full range of actions outlined in the preceding sections. These are the OT&E's given primary emphasis in DOD Directive 5000.1 and 5000.3, AFR 80-14 and other Air Force regulations and directives on testing, and many of the Command regulations and publications.

The planning of OT&E for new items of equipment (bombs, fuzes, life support items, cargo loaders, specialized stores, and the like) and the development of test objectives for a test follows a track similar to system OT&E but generally vary significantly in the scale of effort involved. In the past, acquisition has followed a number of management procedures which have varied from an R&D laboratory proof test to a small scale system approach; however, current trends tend toward a more standard approach in general accordance with AFR 800-2 "Program Management."

6. OBJECTIVES FOR OT&E OF MODIFIED SYSTEMS AND EQUIPMENT

Class V modifications as defined in AFR 57-4 "Retrofit Configuration Changes" are made on systems and equipment in the inventory to provide a new or improved operational capabilities. Normally, a Class V modification originates with the submission of a Required Operational Capability (ROC) document in accordance with AFR 57-1 by an Operating Command. AFR 57-1 specifies that if the best solution to an operational requirement appears to be a Class V modification that certain technical information and recommendations shall be provided including recommendations on testing, kit proofing, and prototyping. After AFSC and AFLC review, and HQ USAF approval, action is directed by means of a PMD directing establishment of a modification project. Class V modifications are managed and documented in accordance with AFR 57-4 rather than

AFR 800-2 "Program Management" and normally AFLC rather than AFSC is the Implementing Command.

Both AFR 57-4 and AFR 80-14 specify that testing will be conducted in connection with modification programs; however, only AFR 80-14 addresses the need for OT&E on modified systems and equipment. In practice, procedures for verifying operational effectiveness/suitability and supportability of modified systems and equipment varies from program to program. OT&E for a large modification program such as conversion of several cargo aircraft to a special reconnaissance version may be directed by an OT&E Test Directive issued by HQ USAF and includes a range of operational/support investigations approaching those involved in the OT&E of a new system. On the other hand, operational and support considerations, including training, may be combined into a test of a modified system or item of equipment conducted by AFLC.

Although OT&E of modified systems and items of equipment is not as well defined as OT&E for new systems and equipment, the characteristics of modification programs that should be considered during test planning and development of test objective include:

a. Requirements for conducting OT&E on modified systems and equipment as stated in AFR 80-14.

b. Operating Command involvement in modification programs, including making recommendations on testing requirements, commences with submission of a ROC.

c. AFLC rather than AFSC is usually the Implementing Command.

d. Testing requirements may be included in the PMD issued by HQ USAF directing modification programs. OT&E requirements may be amplified by an OT&E Test Directive issued by HQ USAF.

e. Participation by other Services in the OT&E may be desirable if the modified system or equipment interfaces with the operations of another Service.

f. Cost, limited test items, and time considerations will probably necessitate conducting developmental or qualification testing in conjunction with operational testing.

g. Presentation of test plans, objectives, and test results for DSARC review and approval will not be required since DSARC procedures normally do not apply to Class V modification programs.

h. Test objectives normally will be directed toward investigation of only those operational capabilities and supportability aspects which are a result of, or are affected by the modification.

i. Test completion normally will be scheduled prior to delivery of modified systems/equipment to operational units.

j. HQ USAF will normally exercise review and approval authority for findings and recommendations which would result in modification program changes.

k. A data base for test planning purposes will exist which includes test reports and other documentation associated with the original acquisition of the system or equipment, reports on demonstrations or tests of prototypes for subsystems or equipment involved in the program, AFR 57-1 and 57-4 documentation, reports of any similar modifications, and studies.

l. Coverage of the equipment acquisition program by a documentation and action chain which includes ROC's, PMD's, and development plans. Development plans may be tailored along the lines of a PMP or they may be part of PMP that covers a family of equipment acquisition efforts in areas such as life support (includes oxygen masks, helmets, flight suits, and the like). Certain established equipment acquisition efforts are documented by means of the Research and Development Planning Summary (DD 1634).

m. Basic test objectives are derived from PMD's and OT&E Test Directives issued by HQ USAF, and testing requirements specified in DOD Directive 5000.3, AFR 80-14 and other AFR's.

n. Normally, HQ USAF test directions will be transmitted to operational test units by means of a Test Order or Test Plan prepared in the headquarters of the Operating Command involved. These documents may expand upon the test objectives specified in PMD's and OT&E Test Directives.

o. Normally OT&E of equipment items will not be subject to DOD/DSARC review and approval procedures since HQ USAF will usually have production decision authority.

p. OT&E of equipment items usually will be divided into IOT&E and Follow-On OT&E, similar to system testing.

q. Usually large test teams will not be required; however, representation from AFLC, ATC, AFCS and AFSS as well as from other Services may be required. In most instances AFLC representation or participation will be appropriate.

r. Completion of IOT&E and submission of independent reports by Operating and Supporting Commands prior to the production decision will usually be required. Completion of a Follow-on OT&E with production items will usually be scheduled for completion early during initial production runs.

s. Valid requirements for introduction teams will exist, particularly for items such as new fuzes, bombs, and other items which may be introduced to operational units before formal training and crew qualification courses have been placed into operation. Team composition and introduction team objectives will be similar to those noted for system programs.

7. OBJECTIVES FOR COMPARISON OR JOINT TESTS

Comparisons of the operational capabilities of Air Force systems and equipment having similar or overlapping capabilities; and comparisons of Air Force systems/equipment with those possessed by another Service are OT&E activities which have recently been given increased emphasis. This emphasis is expected to continue as escalation of military hardware and personnel costs force reduction in duplicative capabilities. Added to these tests carried out jointly by the Air Force and another Service in order to compare system/equipment capabilities, are Joint Tests performed in order to develop an operational data base for use in studies and analysis. Examples include detailed investigation of command and control system capabilities, target acquisition range investigations, and studies of time of exposure to hostile fire during selected weapon delivery operations. Joint Tests, as noted, usually are not involved in OT&E associated with an acquisition program for a new system or item of equipment; however, it should be noted that tests conducted during system acquisition may involve comparison with systems/equipment already in the inventory as well as participation by other Services because of operational interfaces. Comparison tests discussed in this section usually are the result of high level interest (congressional, OSD, JCS, or HQ USAF) in determining comparative capabilities of a proposed system, in prototype or pre-production status, with other systems/equipment in the inventory having similar or overlapping operational capabilities. Joint Tests are usually directed from OSD or JCS level for conduct with operational systems in collection of basic operational performance and capability data which can be used in conducting studies and analyses, and to identify operational capabilities and characteristics desired in future systems.

Comparison Tests

As noted earlier, comparison tests may involve the Air Force only or may be conducted jointly with another Service whose system or equipment is being compared. Although these tests do not follow a set pattern, some of the characteristics which influence test planning and development of test objectives include:

a. Basic test direction originates at the OSD (Deputy Director for Test and Evaluation, DDR&E) or HQ USAF level.

b. Test concepts may be developed by WSEG/IDA for those programs of specific interest to the JCS.

c. HQ USAF transmits directions to an Operating Command for conduct with an OT&E Test Directive.

d. Operating Command may transmit directions to the operational test unit with a Test Order or a Test Plan prepared at the Command headquarters.

e. Participation by other Services will be defined in the OT&E Test Directive.

f. Use of prototype or pre-production equipment from an acquisition program may be directed.

g. Lead times for test planning and assembling resources usually will be much shorter than for OT&E associated with system/equipment acquisition.

h. Test objectives will not cover as broad a spectrum of investigation as in a full-scale system OT&E. Usually will be confined to comparative performance capabilities and, possibly, cost of operation.

i. Contract agencies may be called upon to provide scientific and technical assistance.

j. No introduction team requirements.

k. A data base for planning and test standardization purposes that may include reports of tests on similar systems/equipment, acquisition documentation, DT&E and OT&E reports on the systems/equipment being tested, and studies prepared by the Services or contract agencies.

Joint Tests

By definition, Joint Tests indicate participation by two or more Services. While Joint Tests do not follow a set pattern, some of the characteristics which influence test planning and development of test objectives include:

a. Basic test direction originates at the JCS or OSD (Deputy Director for Test and Evaluation, DDR&E) level.

b. Test concepts may be developed by WSEG/IDA for those programs of specific interest to the JCS.

c. HQ USAF transmits directions to an Operating Command with an OT&E Test Directive.

d. Operating Command may transmit directions to the operational test unit with a Test Order or a Test Plan prepared at the Command headquarters.

e. Participation by other Services and government agencies will be defined in the OT&E Test Directive.

f. Testing program is normally not associated with system acquisition.

g. Test operations may be conducted at an R&D facility if special data collection requirements are involved, otherwise ranges and facilities possessed by an Operating Command probably will be utilized.

h. Operational equipment and crews will perform test missions.

i. Lead time for test planning and assembling resources will not be as lengthy as for OT&E's associated with acquisition programs.

j. Test objectives will be confined to a comparatively narrow spectrum of investigation.

k. Contract agencies may be called upon to provide scientific and technical assistance.

l. No introduction team requirements.

m. A data base for planning and test standardization purposes that may include test reports on the systems/equipment involved, combat and training reports, selected studies and simulations bearing on the area of investigation involved.

8. OBJECTIVES FOR INVESTIGATIONS OF NEW TACTICS, TECHNIQUES, AND OPERATIONAL APPLICATIONS

During OT&E of a new system, particularly OT&E with production items operated in squadron or other basic combat organizational elements, tactics and techniques for operations against known or postulated threats are investigated. Those tactics and techniques recommended for maximum combat effectiveness are described in the OT&E report and are demonstrated and/or briefed by the introduction team. After a new system enters the inventory and has been deployed, it is almost inevitable that new threats and other changes in the operational environment will occur. These changes necessitate continuous capability assessment and the development of new and improved employment tactics and techniques. They also may require investigations and development of procedures for operating a system in a new role or operational application. Operational testing involved may consist primarily of conducting test missions and collecting data sufficient to prove or disprove a proposed solution; or it may be necessary to conduct an extensive series of test missions for the purpose of obtaining basic quantitative data for use in formulating proposed solutions which then would be the subject of additional testing. Although these tests are not associated with the acquisition process, R&D as well as OT&E, test facilities and other test support resources used during full scale system testing probably will be required unless test objectives can be constrained and limited. This area of testing covers a broad range of activities such as: investigation of air combat maneuvers for use in engaging a new interceptor threat, development of optimum ECM formations for causing a threat radar to break lock, determination of radar detection and acquisition ranges against new target complexes, and determination of optimum procedures for visual detection of low altitude intruders. Broad characteristics of this type of test activity which influence test planning and development of test objectives include:

a. Test requirements usually originate at the headquarters of an Operating Command, possibly upon the recommendation of an operational test unit or a combat wing.

b. Test directions are transmitted to operational test units by means of a Test Order or Test Plan prepared in the headquarters of the Operating Command involved.

c. Direct involvement in testing by AFSC, AFLC, and ATC or other Supporting Commands is not likely; however, test results might indicate new equipment, maintenance and logistic support, or training needs which are reported and communicated through normal Command or requirements channels.

d. DSARC procedures are not applicable.

e. Test objectives are normally limited to the specific area of investigation and do not address logistical and support aspects except, possibly, specialized crew training.

f. Findings and recommendations which would lead to adoption of new tactics, techniques, and operational applications are approved at Operating Command headquarters unless implementation involves HQ USAF controlled resources.

g. A valid requirement for an introduction team may exist for the purpose of demonstrating a new tactic, technique, or operational application to combat units and for obtaining data on problems involved during introduction and crew qualification.

h. A data base for planning purposes will exist which includes test reports associated with acquisition of the equipment involved, training and combat reports, intelligence reports, and reports of similar tests and investigations.

9. PROCEDURES FOR DEVELOPING OPERATIONAL TEST OBJECTIVES

Many of the test objectives in current regulations and directives are stated in terms of broad guidelines such as "determination of operational effectiveness and suitability." A logical method is needed for translating these broad guidelines into critical questions and issues, and test objectives which are appropriate for system advocacy, test planning and test design purposes. A successful approach to this problem used by both Air Force and Army OT&E agencies is to conduct an analysis of the operational mission structure in which the system, item of equipment, or tactic/technique will be employed, and to examine the premission, mission, and postmission tasks and subtasks involved in mission support and execution. Each task is then examined in terms of factors which influence how and under what conditions the task will be performed. This procedure will not necessarily identify all the test objectives during coordination and approval; however, it does provide a rational approach for stating questions and issues, and test objectives as they relate to essential operational mission accomplishment.

For example, the operational mission structure, tasks, and mission factors which have been found useful for tactical air operations are listed below:

Tactical Operations Structure

Counter Air	Search and Rescue
Interdiction	Refueling
Close Air Support	Forward Air Control
Fire Suppression	Special Air Warfare
Combat Air Patrol	Battlefield Illumination
Electronic Warfare	Airlift
Reconnaissance	Command and Control

Operations Tasks

Premission

- Planning system check-out and status reporting
- Scheduling, loading and arming
- Briefing
- Scheduled maintenance
- Unscheduled maintenance

Mission

- Aircrew equipment preflight and postflight checks
- Take-off (or launch)
- Climb-out
- Navigation/cruise
- Refuel
- Search
- Acquisition
- Identification
- Evaluation and assignment
- Conversion
- Tracking
- Ordnance or stores delivery
- Return to base
- Countermeasures

Postmission

- Debriefing
- Damage assessment
- Scheduled maintenance
- Unscheduled maintenance
- Operations analysis

Operations Factors (which affect performance of operations tasks)

Pre-takeoff and post-landing

Operations

- Personnel in terms of numbers, experience, etc.
- Training
- Support equipment such as simulators, technical data, etc.
- Supply
- Intelligence
- Facilities in terms of space, layout, etc.

Maintenance and Armament

- Personnel in terms of numbers, experience, etc.
- Training
- Support equipment such as AGE, common test equipment, technical data, etc.
- Supply
- Facilities in terms of space, layout, etc.
- Climatic environment

Environment

- Target
- Weather & climate, includes temperature, humidity, etc.
- Enemy defenses, countermeasures, etc.

Aircrew

- Human factors
- Aids such as maps, pads, slide rules, charts, escape kits, etc.

Airborne Components

- Sensors such as visual, radar, IR, etc.
- Communications such as voice, data link, etc.
- Computer such as air data, weapons release, etc.
- Armament or weapon auxiliaries
- Navigation such as loran, TACAN, inertial, etc.
- Flight control
- Airborne
- Propulsion

Ground and Airborne Command and Control Components

- Sensors such as visual, radar, IR, etc.
- Communications such as voice, data link, etc.

Mission tasks and factors can be matched as appropriate to system, equipment, tactic, or technique under consideration and how they are related to the flow of mission events. Based on examination of the tasks and factors involved, critical questions and issues which should be answered by OT&E and the test objectives necessary to find these answers can be developed in general terms during preliminary planning and, later, in the specific terms required in detailed planning, Test Plan preparation and test design.

Steps in the Development of Objectives

The first step is to develop a clear definition of the purpose of the OT&E. It will normally fall into one of three general types: (1) OT&E of a major weapon system, supporting system or item of equipment, (2) an OT&E devised to solve an operational problem, or (3) a mission enhancement OT&E designed to explore and examine a new application of an existing or modified system. The statement of purpose is the basis for the general test objective(s) and for all detailed objectives.

The second step is to conduct an analysis of the mission structure involved in the OT&E. This constitutes a top level structure of the scenario with identification of the players, and general description of the conditions of the OT&E. An ordered approach to test progression is structured to derive an overall concept of test and provide an approximation of the scope of test.

The third step is to conduct an analysis of the purpose and general test objective(s) and to divide the purpose into two categories: (1) mission, and (2) support. The mission oriented parts will be derived from the mission analysis, related operations structure, and operations factors. The support parts will be derived from the normal support requirements related to the mission analysis, augmented by the unique support characteristics of the OT&E. Specialized assistance is recommended in those unique areas of logistics, training, advanced technologies and new applications.

The fourth step identifies detailed objectives derived from the prior analysis. Detailed objectives identified must relate to the general objective(s) and to each other. Areas of duplication in detailed objectives should be reduced to a minimum and each set assigned a priority.

The fifth step is constraining the objectives to realistic dimensions. The objectives developed are to this point relatively unconstrained by practical economic and time limitations. Ideally, this is the way it should be. However, time and economics are real world conditions and form the boundaries within which the ideal objectives must be fulfilled. The best method used for constraining objectives is judgment. You can employ two screening processes: (1) a top level screening which identifies each objective as primary or secondary, and (2) a screening process for each set of objectives which established a confidence level and sample size for each set of test conditions involved in satisfying the objective. From this a first cut of the actual scope of effort is established and trade-offs can be conducted to determine which objectives survive and those which are labeled "nice-to-have".

After having developed a list of test objectives for a particular OT&E which satisfies the requirements of applicable regulation and directives, answers critical questions and issues, is consistent with data base requirements, and has survived coordination procedures, the test planner should still conduct a careful review of the objectives before issuing a directive or preparing a Test Plan.

CHECKLIST FOR FORMULATION OF THE OBJECTIVES

1. Are all critical questions and issues translated into test objectives?

2. What are the primary and what are the secondary test objectives? Can any of the secondary objectives be satisfied as by-products of the primary objective or through add-on data collection operations performed on a non-interference basis?

3. Are all test objectives consistent with the primary purpose for conducting the OT&E?

4. What objectives should receive primary consideration during IOT&E and which ones can be deferred until OT&E?

5. In system acquisition OT&Es are the critical questions and issues in the DCP, test objectives in the PMD or OT&E Test Directive, test objectives, in the PMD, test objectives in the Test Order, and detailed test objectives developed for the Test Plan all consistent?

6. Which test objectives can be satisfied by evaluation and which can be satisfied only by test?

7. Are the objectives realistic in terms of time, cost, and resources? What changes can be made and still satisfy information and data requirements?

8. Can the test objectives be readily converted into Measurements of Effectiveness? If not, which test objectives should be revised?

9. Have OT&E requirements for Supporting Commands been covered in such areas as maintenance, logistics, training, communications, and communications security?

10. Do test objectives cover area of interface with other Air Force systems, and/or systems operated by other Services?

11. Are the test objectives consistent with OT&Es of similar equipment so that the data and information obtained will readily fit into the existing data base?

12. Do test objectives cover all of the information and data needs in controversial areas?

13. Are objectives consistent with existing and/or planned tactics, techniques, and doctrine?

14. Do test objectives reflect Operating/Supporting Command inputs?

15. Do test objectives cover consideration of organizational and force structure aspects including all of the interfaces involved?

16. Are test objectives consistent with the planned mission of the system? Are secondary missions covered?

17. Do test objectives cover development of tactics and techniques for proper operational employment?

18. Do test objectives cover information and data needs for combat/operational introduction teams? Are test and evaluation objectives for the combat/operational introduction test and evaluation phase included? Are there any unique support aspects which should be covered by the test objectives? Has coverage of environmental aspects been covered? Are cost data requirements covered? Are mobility/transportability aspects covered?

19. Can the test objectives be broken down into sub-sets to the point where the last level is sufficiently simple to be considered a data requirement? Is each simple enough that it can be answered in a clear manner by a measured value or a simple one word answer or checkmark on a questionnaire?

20. Are the objectives oriented to operational missions as opposed to system performance per se (a DT&E Objective)?

Chapter 5

SELECTION OF AN APPLICABLE TEST CONCEPT

1. INTRODUCTION

Several times in the course of the life of an OT&E program it may be necessary to conduct a tradeoff study to determine the best test concept or approach to be used on the program. This best approach is invariably closely allied to available resources. There is an ever present need to be alert to the possibilities of reducing the magnitude of the OT&E effort in order to keep within the resources available and the time allowed. The first use of the tradeoff study may come very early in the cycle when higher echelons consider the need and advisability of subjecting the test item to a Joint Test, or employing it in a large scale exercise or perhaps conducting a simulated operation with multiple test items in order to more realistically represent the actual tactics of employment. Another use of the tradeoff study may occur later in the program and at lower echelons to determine the best option from one of a more limited span of choices. In both cases the basic principles used in the tradeoff studies are the same. Therefore, for simplicity and brevity, the tradeoff study will be described as a standardized process covering a wide range of application rather than two processes in different time frames each addressing a smaller range of candidates.

The first step in the tradeoff study process is to develop a set of test concepts that are suitable candidates for the OT&E under consideration. The set should be arranged in a logically graduated sequence from the simplest candidate to the most complex.

2. TEST CONCEPTS

A list of hypothetical test concepts is presented in the following paragraphs.

Evaluation only (no physical testing). This candidate makes use of existing data compiled from available sources. Data included may be from previous operational tests, development tests, contractor data, data from other Services, or even data from similar enemy systems if available. Theoretical analyses may be considered if they are particularly appropriate and other data were not available. Some obvious advantages for this candidate are that it can normally be accomplished in a relatively short time with a few people and at slight cost. However, some clear disadvantages also exist.

The quality and limitations on the "borrowed" data are seldom as well defined as is desirable, hence misapplication or over extension of the data is generally a matter of concern. The results of the evaluation may lack realism because of the absence of the specific hardware test item, specific operational personnel, and the operational environment. All these dangers can be reduced by careful examination of all the data and accepting for the evaluation only the most applicable.

Simulation. This candidate also avoids physical testing with the attendant advantages of reduced time, limited personnel requirements, and moderate cost (particularly if an existing model can be used for the simulation). If a model must be devised and perhaps a computer program written some of these advantages may shrink or even disappear entirely. Other advantages are the wide range of test conditions that may be explored, and the ability to simulate future conditions or environments that are presently beyond the state of the art to physically represent. Disadvantages include the constant concern over simplifying assumptions and idealizations that may be unrealistic, and the need for validation of the model over the range of variables with which it will be used.

All of the remaining candidates include physical testing to various degrees of complexity.

One form of practical simulation is the "dry run" i.e. test firing runs without ammunition release.

Targets of Opportunity. This type of test, for example, may be applicable to communications/electronics hardware. In the case of a surveillance radar the test item may be operated in an area containing high density civilian air traffic. A wide variety of target sizes, ranges, altitudes, speeds, and aspects may be investigated without the cost of providing scheduled targets.

Scheduled Targets. By using the radar test example, more specific data on the radar may be obtained by providing scheduled targets of the type expected to be tracked in operational use. These targets can be programmed and controlled in realistic flight profiles and more comprehensive and critical data (such as probability of detection) obtained with realistic operator and maintenance personnel supporting the system.

Live Ordnance. In the case of aircraft tests the next step may be the use of live ordnance with the increased scope of using different tactics for the ordnance delivery.

Interfacing equipment. All of the previous tests treated with a single test item without interfacing equipment. In tests of interoperability, interfacing equipment will be required to obtain responses to the test objective. The complexity of the test increases accordingly with increased time, personnel, and costs.

Multiple Test Items. In tests involving tactics and tactical formations several test items may need to be operated simultaneously with the associated ever increasing costs.

Two-Sided Tests. These tests contain an adversary component and as a result are frequently confused with Joint Tests. An airborne radar system may be tested to determine how well it supports a penetration mission in the presence of enemy electronic countermeasures. If both the airborne radar and the simulated enemy jammers are under control of the Air Force no other Service is required. If, however, an Air Force aircraft is tested to determine how well it can accomplish ordnance delivery in the presence of simulated enemy ground based anti-aircraft fire provided by the Army, the test will become a Joint Test. Two-sided tests (although more realistic) increase in effort, cost, and time because of their increased complexity and their need for more stringent safety measures to protect the players and test support forces.

Joint Tests. These tests involve the commitment of equipment and personnel of more than one Service. Generally, the adversary role is present. The increased realism is a clear advantage. They may suffer from the same disadvantages as two-sided tests in addition to the added difficulty of working into the separate and parallel chain of command of the other Service.

Exercise. Items undergoing OT&E may be inserted into a large scale exercise to take advantage of the realistic operational environment and the availability of targets. The main purpose of the exercise is training and as such a "no-interference posture" of the OT&E must be maintained. This frequently results in less instrumentation (and thus less data) than the USAF would like. Also, the OT&E personnel have much less control over events which again may result in less data (if missions cannot be delayed when the test item is malfunctioning). The cost of participating in the training exercise may be modest, but since the volume of resulting data is limited, the cost per data element may be much higher than for other candidates.

This representative list of test concept candidates can be increased or altered as the test item type and the objectives of the test vary.

3. CRITERIA FOR DECISION

Against this spectrum of candidates must be considered the criteria that influence the decision making process. The criteria are those major considerations that form constraints or describe characteristics that are of primary concern to the tester.

Availability of Time. Seldom can a test be conducted on a leisurely basis. The test is generally only a portion of a much larger problem or an acquisition program which has a firm schedule that is interlocked with the developer's schedule, with user's requirements, and with higher echelon budgets. Within these limits the OT&E must be accomplished. The various test concepts require different lead times to arrange and different execution times.

Availability of the Range. Range schedules can, at times, become overloaded. Keen competition will exist for any openings that do appear. Therefore, tests that can be accomplished with minimum range time are very attractive. In view of range costs this is always a good idea in any case.

Availability of Personnel. Personnel are required to operate and maintain the test item, as well as to provide other support. Quite apart from the cost of these personnel is the problem of obtaining their efforts for the period required. Obviously the greater the number (and complexity) of pieces of equipment involved in the test and the longer the test period extends the le[illegible] the likelihood of being able to obtain and retain their service when desired.

Availability of Equipment. There is generally an acute shortage of the amount of equipment needed to perform OT&E. IOT&E is a particularly difficult case with perhaps only one or two test items in existence. The situation gets even more critical in a combined DT&E/IOT&E with one test item serving both the developing command and the operating command. Tests with interfacing equipment and test with multiple test items also may pose difficulties.

Risk. The risk associated with accomplishing the test on time and with good results varies with the test concept. The simple, smaller tests normally get higher grades on this criterion than do the longer more complex tests.

Data Quality and Quantity. This criterion is interesting in that it changes in an irregular way across the spectrum of test concepts. The grade may be low on both ends of the spectrum. In an evaluation (without physical test) use must be made of whatever data is in existence. This invariably results in less quality and quantity than is desired. Similarly, the exercise provides limited quantity and quality because of non-interference with the training mission. The simulation rates higher than the evaluation because of the large quantity of data that can be obtain ven though the realism of the data may be limited.

Cost. This crucial criterion is often difficult to apply because of the general lack of detailed cost related information on OT&E. Perhaps the most reliable approach to examining the costs of any test concept is to consider the entire range of elements that make up OT&E costs and then examine any cost increments that are related to the test concept. The OT&E cost may be divided into the following separate elements: pre-test activities (planning, simulation and modeling), the test item, the use of a test platform (aircraft), the use of the test location (test range), personnel (both military and civilian), targets and threat simulation, special purpose instrumentation and support equipment, post-test activities (data processing, data analysis, report preparation) and contingencies. In general, it appears that all of these costs increase as the test concept increases in size and complexity. In the absence of recorded data for these cost elements (and much will be absent) personal contact should be made with Test Directors on current or previous tests to obtain the benefit of their experience. From the examination of the cost increments, relative cost comparisons can be obtained for the spectrum of test concepts.

4. TRADEOFF MATRIX

A sample array in matrix form of the criteria and test concepts is shown in Table 5-1. The form has been filled with a hypothetical qualitative rating for each test concept for each criterion. (These ratings can be argued with, but that is of little importance in this example of methodology.) Any test concept that has all high ratings is a very attractive one (such as targets of opportunity in Table 5-1). In the squares where a low rating appears, it may be necessary to get a rough quantitative measure to insure that the candidate is not eliminated from consideration by being below the threshold of available funds or whatever other resource is being considered.

In the real case the matrix structure will be modified by the type of system being tested, the objectives of the test, and the importance of the criteria related to the specific conditions of the test.

An overall review of the ratings may permit one candidate to be selected. If a single candidate cannot be selected the review will help in the elimination of candidates that are not good choices. A more detailed investigation can then be made to arrive at the single choice.

TABLE 5-1. TRADEOFF MATRIX

Test Concepts

Criteria	Eval. Only	Simulation	Targets of Opportunity	Scheduled Targets	Live Ord.	Inter-facing Equip.	Multiple Test Items	Two-Sided Test	Joint Test	Exercise
Availability of Time	High	High	High	Med	Med	Med	Med	Low	Low	Low
Availability of Range	NA	NA	NA	High	Low	Med	Med	Low	Low	Low
Availability of Personnel	High	High	High	High	High	High	High	Med	Med	Med
Availability of Equipment Test Items & Support	High	High	High	High	High	Med	Med	Low	Low	Low
Risk	High	High	High	Med	Med	Med	Med	Med	Med	Med
Data Quality & Quantity	Low	Med	High	Med	High	High	High	High	High	Low
Cost	High	High	High	Med	Med	Low	Low	Low	Low	Low

Chapter 6

MEASURES OF EFFECTIVENESS

1. INTRODUCTION

Within the Air Force and the rest of the DOD, there exists many different ways in which the term Measures of Effectiveness (MOE's) is used. This is because the term "MOE" has not been defined. MOE's are used to mean any attribute of a system such as power output of a transmitter or the cost of executing a specific mission with the system. MOE's are also used to mean how well a system accomplishes missions. MOE's even include any evaluation of the system such as "I like it." Whatever way the term is used, the MOE's are seldom, if ever, defined in such a way that the reader knows exactly what is being measured. Further confusion is generated because many common terms (such as CEP) have no standard definition and the tester uses defined terms (such as mission or system) to mean things different than the approved DOD definitions. To further compound the problem, test objectives for similar systems with the same mission are not standardized.

This lack of standardization leads to disadvantages in the present application of MOE's. There is difficulty in communicating between decision levels (tester, Command HQ, AFHQ, DDR&E, SEC DEF and Congress) since often there is no way for any decision maker to find out what is meant by the terms used. The main reason for having MOE's is to aid management in making decisions; therefore this communication difficulty needs correcting. Confusion is further increased by having many MOE's for a given type system for a single specific mission. While these MOE's may at times be somewhat similar, it is seldom possible to go from one to the other without more information. This information is often lacking.

Another problem caused by the lack of standard definitions and information, is the inability to use data from one test for other tests. This leads to duplication in testing and extra expenditures of resources. With standardization, the effect of modifications will be easier to detect. The resulting improvement or degradation in the system's operation is easier to determine (that is, takes less data) if the same MOE is used throughout the system's life cycle.

Unless it is quite clear what was measured, the implications of the results of a test can easily be misunderstood. Also, it is nearly impossible to compare a new system with an old one when different or undefined MOE's are being used. When the quality of the data gathered is not changing rapidly, using standard MOE's for old and new systems will make it easier to compare them as long as the test conditions are essentially the same.

The main advantage of not changing the present situation in MOE's is that no one would have to change his modi operandi.

The least change to the present situation in MOE's would result from 1) a requirement that all terms and measures be defined and 2) many different MOE's still be retained for a system for a single specific mission. With these changes, communications would improve and there would be some decrease in the duplication of testing. However, these advantages would not be optimized. The rest of the disadvantages mentioned above, would remain.

The MOE's scheme being proposed goes one step further. Not only are defined standard terms to be used, but there will be only one MOE for a system for each specified scenario. The conditions under which the system is to be tested will be defined. The MOE will address the operational characteristics of the system at the user level. That is, the MOE addresses mission success, not higher level successes such as outcome of the battle or the war. Using MOE's in this manner would give the best communications between decision levels. The different decision levels can assign their own weights to different aspects of the system as are appropriate to them and know what and how they are weighting. Data from one test can be easily used on, or combined with, another as appropriate. The applicability of the data is easy to ascertain. Improvements or degradations to a system due to modifications can be determined more easily than at present. This is even more true since much of the data can be easily obtained during normal operational use. This FOT&E data can then be used as a base line against which the modified system is to be compared. These advantages (combining data and easily noticed changes) result in the greatest saving in resources of all the MOE's schemes considered since testing and duplication in testing are minimized. The scheme for MOE's being proposed is the one with which new testers can most easily become familiar. This results from the ready availability of the meaning of terms and the overall logical approach to MOE's. It will be possible to make direct comparisons between old and new systems when such comparisons are appropriate. That is, when the scenarios are the same. The standard MOE's will make the meaning of test objectives clear to the tester as well as decision makers. The disadvantage of standardizing MOE's is that changes would be required from the way things are presently done.

Since MOE's, as the term states, are used to measure the effectiveness of a system, the MOE's used will be addressing system effectiveness at the user level. At higher levels, other considerations besides MOE's are used to make management decisions. These other considerations include, but are not limited to, life cycle cost, urgency of the need, priorities and politics. Thus MOE's are one in a series of factors going into the final decision process.

Before defining what will be meant by MOE's, several other terms must be defined. While many of the definitions to be given are taken directly from Mil-Standards, it is important to include them here to be sure that there is no misunderstanding of what is being said. Up to this point in the discussion of MOE's, it has made little difference if the exact meaning of the terms used was clear or not. Henceforth, clarity is essential.

Mil-Std 499 (USAF) defines a "system" in the manner in which it will be used in further discussions concerning MOE's. It says, "A system is a composite of equipment, skills, and techniques capable of performing and/or supporting an operational role. A complete system includes all equipment, related facilities, material, software, services, and personnel required for its operation and support to the degree that it can be considered a self-sufficient unit in its intended operational environment."

Another term that will be used often is "mission". AFM 11-1, Volume 1 defines mission as "the task, together with the purpose, which clearly indicates the action to be taken and the reason therefore."

Before a measure can be defined, the property being measured must be defined. Therefore, before Measure of Effectiveness can be defined, the meaning of effectiveness (a property) must be agreed upon. Also, what it is that has this measurable property must be defined. OT&E is the area of concern, therefore it is a system that possesses some level of effectiveness which is to be determined. The definition of "system" is given above. Mil-Std 499 (USAF) defines system effectiveness as follows:

> "System effectiveness is a measure of the degree to which a system achieves a set of specific mission requirements. It is a function of availability, dependability, and capability."

Now three more terms must be defined, namely "availability", "dependability", and "capability". Mil-Std 499 (USAF) refers to Mil-Std 721B for these definitions. The latter defines "availability" as:

> "A measure of the degree to which an item is in the operable and committable state at the start of the mission, when the mission is called for at an unknown (random) point in time."

It also defines "dependability" as:

> "A measure of the item (sic) operating condition at one or more points during the mission, including the effects of Reliability, Maintainability and Survivability, given the item condition(s) at the start of the mission. It may be stated as the probability that an item will (a) enter or occupy any one of its required operational modes during a specified mission, (b) perform the functions associated with these operational modes."

And finally, "capability" is defined as:

> "A measure of the ability of an item to achieve mission objectives given the conditions during the mission."

Capability and dependability are conditional measures as their definitions say. That is, capability requires that the fact of dependability be given. Similiarly, dependability requires that availability be given. The problem still exists of deciding on the scale (units) to be used for availability, dependability and capability. Since Mil-Std 721B states that dependability may be stated as a probability, logically it is desirable to state the other two as probabilities. Hence, Availability (A) is the probability that an item is in an operable and committable state at the start of a mission when the mission is called for at an unknown (random) point in time; Dependability (D) is the probability that an item will (a) enter or occupy any one of its required operational modes during a specified mission, (b) perform the functions associated with those operational modes given the item's Availability; and Capability (C) is the probability that an item will achieve the mission objectives given the Dependability. Thus D and C are conditional probabilities (see Appendix G). When capitalized, Availability (A), Dependability (D) and Capability (C) mean the probabilities as defined above. With these definitions, A, D, and C are statistically independent. That is, if an item is said to be available for a mission, its dependability can be addressed. If it is unavailable, no information about its dependability can be gathered. Looked at from another point of view, the results of measuring D given A are the same as the result of measuring D with no statement about the magnitude of A but given the fact the item was available for D to be measured.

With the above definitions, it follows that a Measure of Effectiveness of an item is a parameter which evaluates the extent of the adequacy of the item to accomplish an intended mission under specific conditions. It is a function of Availability, Dependability, and Capability. Thus MOE's are expressed as probabilities since A, D, and C are probabilities.

There are certain qualities any good MOE should have:

a. The MOE should be sensitive to all variables affecting the item. By this it is meant that anything that affects the item's effectiveness should appear as an input to the MOE is some fashion.

b. The MOE should be precisely defined. This prevents decision makers and others from misunderstanding the implications of the MOE.

c. The MOE should not be overly broad. The MOE should address the effectiveness of the item in question and not include other items not relevant to the issue. For example, a MOE addressing the outcome of an all out nuclear engagement is too broad to evaluate a new bomb/navigation subsystem for strategic bombers even though that subsystem may affect the outcome of the engagement.

d. The MOE's, as well as their input measures of performance, should be mutually exclusive. This prevents one aspect of the item from being counted several times and weighting the MOE heavily for this aspect.

e. The MOE should have exhaustive inputs. This assures that all aspects that can affect the item's effectiveness are included in the inputs.

f. The MOE should be relevant to the mission. This assures that the proper effectiveness is being measured.

g. The MOE should have inputs that are relevant to the design issues. This assures that the issues are investigated.

h. The MOE should be expressed in terms meaningful to the decision maker. Since the purpose of MOE's is to aid the decision maker, it is important to have the MOE's meaningful to him.

i. The MOE should have inputs that are measurable. If the inputs are not measurable, the MOE cannot be evaluated.

j. The MOE and its inputs should be quantifiable if at all possible. Qualitative evaluations should be used only for aspects that cannot be measured. This is almost always only the man-machine interface.

In the discussions that follow, the proposed MOE's, measures of performance (A,D, and C), and the data elements satisfy all of the above qualities of good MOE's.

For a single scenario, the tester will make measurements from which the data elements can be derived. The data elements are combined to give A, D, and C. Having these numbers, the tester obtains the MOE for that scenario by multiplying A, D, and C together. Much of the data used to calculate A and D are already being taken for use in the study of the logistics support for the test item.

There is no intention to limit the tester to only those measurements described herein. If he chooses to, he may make additional measurements or analyses as he sees fit.

2. ASSUMPTIONS AND GROUND RULES

As with any area of technical endeavor, certain assumptions or ground rules must be made for MOE's. The rationale for these assumptions and the assumptions themselves are discussed in the following paragraphs. No attempt has been made to put them in any particular order of importance.

a. Standard MOE's will be at the user level. Since these MOE's are inputs at all decision levels, it is only here that is is possible to start standardizing. It would be impossible to standardize MOE's at some other, higher level, if the inputs to the user level were not standardized. Thus, with the lowest decision level using standard MOE's, it is easier to standardize at all other levels.

b. There will be a separate MOE for each scenario for which a system has a mission capacity. If the MOE's were combined into some grand ensemble MOE, it would be impossible to separate the MOE for the most important (or most likely) mission from the least important (or least likely) one. Also, the importance or likelihood of a mission is time dependent whereas the MOE should be a constant as long as the system and scenario remain constant.

c. The mission for which the system is to be tested must be defined before the measurement is made. For example, the effectiveness of an aircraft will be different for an air-to-air engagement than for an air-to-ground engagement.

d. The scenario must be explicitly stated. If the scenario is not prescribed in the documentation received by the tester, he must define it. The scenario includes the following information:

(1). The mission to be executed

(2). A complete definition of the system whose MOE is to be determined.

(3). For a test of one system against a second system (i.e., a two-sided test) a complete definition of the second side system including such things as target aspect angle for radar systems.

(4). The tactics to be used in the test. This includes the second side's tactics.

(5). The type and number of personnel to be used in the operation of the system.

(6). The level and type of training of the personnel in the operation of the system.

(7). The level of the engagement. For example, one-on-one or N-on-M where N and M are integers.

(8). The use rate. For example, for an aircraft it might be one sortie a day or a maximum sortie rate.

(9). The sequence of events in the mission profile. For an aircraft, this would be flight profile.

e. The detail of the tester's data elements should stop before the subassembly level. This prevents having to change data elements for a given type subassembly whenever the designs are different. Subassembly, as used here, means (per MIL-STD-280A):

> "Two or more parts which form a portion of an assembly or a unit replaceable as a whole, but having a part or parts which are individually replaceable. (Examples: Gun mount stand, window recoil mechanism, floating piston, telephone dial, IF strip, mounting board with mounted parts, power shovel dipper stick.)"

The data elements for Availability and Dependability will be based on Technical Order (T.O.) checklists whenever possible.

f. All quantifiable data elements and measures of performance (A, D, and and C) are to be stated and measured as probabilities. This facilitates joining the measurements mathematically and giving them a physical interpretation individually as well as when combined. Since the probabilities (data elements) are either independent or conditional, their product has the same meaning and value as obtained by only determining the value of measures of performance (A, D, and C).

g. There will be a single, well defined, scale for qualitative evaluations. Qualitative evaluations should be used for man-machine interfaces only. All other evaluations should be quantitative.

h. No weights should be assigned by the tester to the different measured quantities or qualitative evaluations. Each should be observed or measured and reported. The reason for this is that each person using the data will have different weights. Even if a Delphi process were to be used, the opinions would change when there is a change of circumstances. The data should be presented in such a fashion that it is useful individually and for any circumstances. A Delphi process is a means of trying to obtain a concensus on the assignment of weights from a group of experts. Weighting can be useful to decision makers but should not be used by testers since testers should report data that is supportable by measurements and will not change as the values of the weights vary over a period of time. For example, the mission for the item tested may decrease in importance but the item's MOE for that mission remains constant.

i. When testing a subsystem or lower level item the tester should select all data elements that apply to his item plus all other data elements that address its interoperability with other items of the whole system. He does not need to test the complete system each time. For example, if the test item were a new transceiver in an aircraft, then the tester would select the data elements for the transceiver, check for RFI affecting the Availability, Dependability and Capability of other electronic equipment and the Availability and Dependability of the electric power source in the aircraft. The tester reports the value of the data elements for the new item and the corresponding data element for the old (or alternate) item. The ratio (new to old) of those data elements that address Availability measures the effect of the item on the system's Availability. Similarly, how the item affects the system's Dependability and Capability is measured by the ratios (new to old) of the data elements for Dependability and Capability.

j. All MOE's are to be the product of Availability (A), Dependability (D), and Capability (C) as defined earlier. A, D, and C are time sequenced in the sense that D requires the system to have been available and C requires the system to have been dependable.

k. The MOE, the measures of performance (A, D, and C), and the data elements should always be reported. By so doing, future testing is minimized when modifications are made to the test item or different scenarios (which include missions and tactics) are used. Often these changes may only affect some of the data elements or measures of performance. In that case only those changed data elements need be evaluated. Also, as is often the case, low values of A, D, and/or C can be traced to a few data elements. The decision maker should be well aware of the cause of the low A, D, and/or C so that appropriate actions can be taken.

l. According to MIL-STD 721B, "failure" is the inability of an item to perform within previously specified limits. From an operational point of view, if a failure leads to an abort of a mission, it makes no difference if the failure is real or if the operator just thought there was a failure. Therefore, in all cases, failures that lead to an abort of the mission will include both real and imagined failures. The tester should keep a record of each kind of failure as the recommended fix is different for each type. According to AFM 11-1 Volume I, abort is the "failure to accomplish a mission for any reason other than enemy action. It may occur at any point from initiation of operation to destination."

m. The tester should only call for checks of items when they would normally be done when the system is being used operationally. Two reasons for this are 1) all of the extra testing can put an added load on the test item, the operators and the rest of the system, and 2) if the check is normally performed during the Dependability phase, a discrepancy could lead to an abort where if it were checked much earlier (an unusual time), the discrepancy might be fixed and no change in the MOE would be recorded.

3. ADVANTAGES OF THE PROPOSED MOE SCHEME

The above assumptions and ground rules lead to certain advantages for the MOE scheme proposed. Some of these advantages are:

a. Data from one test can be combined with data from other tests since data elements are defined and standardized.

b. Data collection for A and D is easy. It is collected by exception since, if a system is available for a mission (scenario given) all of the subsystems that are checked and are needed must have been available. This data is presently collected and used for the Air Force standardized maintenance data collection system. The most readily available aid to the tester in this area is the maintenance officer assigned to the test.

c. Problem areas are highlighted by their low value contributing to a low A, D, and/or C.

d. Since the value of most of the data elements for A and D can be obtained during normal operations, changes in these values can be noticed early in testing. It is easier (requires less data) to notice a change in value than to determine the value of a data element.

e. Data elements are combined into meaningful concepts so that the number of measures are minimized. That is, the data elements combine to give the required information rather than requiring another measurement (or series of measurements) to obtain the required information.

f. Setting limits (upper and lower) on A, D, and/or C as well as specific data elements means testing can be stopped earlier by showing the item will satisfy the test objectives or cannot pass them. This early decision will reflect in a savings of resources. For example, if the A of an item is less than 0.50 with 90% confidence (see Appendix G) it may be prejudged unacceptable and test can stop. Similarly, if A is greater than 0.95 with 90% confidence, it may be prejudged to be acceptable and testing for A can stop.

g. The test objectives, data elements, and test results are logically tied together in such a way that tests should not have to be repeated with the current method. This has happened because of inconsistencies between what was really wanted and what was tested.

4. STANDARD QUALITATIVE EVALUATION

Qualitative evaluations should be used only when it is impossible to have a quantitative evaluation. It should, therefore, be used mainly for evaluating man-machine interfaces and not the resulting response of the machine to the interface. The evaluation becomes the man's personal estimate of the man-machine interface. Care should be exercised in the interpretation of the results of qualitative evaluations since, when numbers are assigned to them, the meaning of such things as the average evaluation and its standard deviation can have meanings different from quantitative data averages and standard deviations.

There are many non-standard qualitative evaluation schemes. Some of these schemes are very restrictive in their applicability such as evaluating automatic landing systems by test pilots. None of these was designed to be a standard qualitative evaluation scheme. One of the major problems with many of the schemes is that the order of the descriptors used (from best to worst in quality) is neither obvious nor clear.

One of the problems has been that there was no standard with which the evaluator was to compare the item under test. The proposed qualitative evaluation scheme is based upon the requirement that a "Standard" be defined. This Standard is either the average of similar items from the evaluator's experience or the item to be replaced by the item under test. Which of these Standards is to be used should be explicitly stated by the tester in his instructions to the individual evaluators.

Using the appropriate definition of the Standard, the evaluator is asked to assign a number evaluation (Q-factor) using the descriptor, and the number correlated with that descriptor as shown in Table 6-1. The fact that what one evaluator calls "Slightly Above the Standard", another may call "Above the Standard" is really of no consequence. What is important is that the order of the descriptors is never confused. That is, there is no question that going from 0 to 10, the corresponding descriptors are describing better and better items.

The fact that the qualitative evaluation is now quantified does not, in any way, convert the qualitative data into quantitative data. The only reason to quantify the qualitative data is to aid in the analyses and manipulations of them.

TABLE 6-1. QUALITATIVE EVALUATION SCHEME

Q-Factor	Descriptor
10	Perfect
9	Outstanding
8	Well above the Standard
7	Above the Standard
6	Slightly above the Standard
5	Same as the Standard
4	Slightly below the Standard
3	Below the Standard
2	Well below the Standard
1	Unacceptable
0	Of no value

Standard - 1. Comparable to the average of similar items or

2. Comparable to the item to be replaced by the item under test.

Care must be taken in the interpretation of the data that is obtained by qualitative evaluations. Normally, for quantitative data, essentially the same instrumentation is used over and over. When this is done, the mean and standard deviation usually relate to the mean and deviation of the performance of the item under test. This is true when the variation in the measuring instruments is much less than the variation in the performance of the item being tested. In qualitative evaluation, the mean and deviation from a single evaluator corresponds to the mean and deviation from instrumentation. That is, it addresses what that evaluator thinks of the item (on the average) and how erratic is the performance of the item in the evaluator's judgment. The other type mean and deviation is derived from many evaluators. Here the mean is an estimate of what the next evaluator will, on the average, say about the performance of the item. The deviation is an estimate of how the item is evaluated from the viewpoints of different individuals.

It should always be kept in mind that qualitative evaluation should be restricted to (1) those aspects of the item's effectiveness that are not quantifiable and (2) evaluations of the man-machine interfaces. Examples of the latter are ease with which the person can successfully accomplish a task with the item, physical comfort while interacting with the item, and psychological aspects of the interaction.

5. STANDARDIZED MOE's SCHEME

The standardized MOE's scheme being discussed in this document can perhaps best be understood by use of a rather simple example. After the simple example is understood, the full MOE's scheme will be easier to comprehend.

The mission of the example system is "Point-to-Point Telecommunications." The system is a Communication/Electronics System. The hypothetical system is made up of two field telephones (model AN/PCC-3) connected by two miles of field telephone wire (type AN/UTC-14) with one person on each telephone.

The scenario for which the system's MOE is to be evaluated is as follows. The system is in a tropical environment. The two miles separation is all overland. The system is not under enemy attack nor is it in an EW environment. The operators are an average aircraft maintenance supervisor and an average base supply attendant. The type of messages to be sent are orders for replacement parts by the aircraft maintenance supervisor and reports of spare parts on hand information by the base supply attendant. A message consists of an order for or the stock level of one part. Each message is sent twice by the talker and said once by the listener after the second transmission. This repeating of the message is for verification. In case of an error of any type, the message is repeated once more by the talker. If any error still persists, the system failed in that "Point-to-Point Telecommunication". The time allocated to the complete communication is the time required for the speaker to say the message three times plus the time for the listener to say the message once plus one minute (the time allocated for call setup). There will be 10 aircraft maintenance supervisors, each with 5 different but typical messages. There will be 10 base supply attendants, each with 5 different but typical messages. This results in a total of 100 different typical messages. Each sender works with each of the 10 receivers once through his list of typical messages. Thus, there is a total of 1000 messages sent by the system. The decision maker wishes to decide if this existing system can be used for this new purpose.

The test objectives are to estimate the system's Availability, Dependability and Capability and give 95% confidence limits on these estimates.

Normally, at this point the tester would go through the list of data elements for a Point-to-Point Telecommunications System and select those that apply to his system, scenario and test objectives. In the example, this step is deleted and only the results of this selection will be covered.

For generality, all Communications/Electronics Systems have been defined as being composed of Transmitting Subsystems, Propagation Medium Subsystems and Receiving Subsystems. Not all systems contain all three subsystems, but in this example all three are represented. A general description of what is included in each subsystem will be found in the discussion of standard MOE's for Communications/Electronics Systems (see page 100).

For purposes of illustration, only those components represented in the example will be discussed at this point. Each Transmitting Subsystem consists of those portions of the field telephones (type AN/PCC-3) which make up the power supply (batteries), the mouth piece, the transducer that converts the spoken words into electrical signals, the connectors to the Propagation Medium Subsystem and the human speaker of the messages(talker).

The Propagation Medium Subsystem consists of the two miles of field telephone wire (type AN/UTC-14) connecting the two AN/PCC-3 field telephones.

The Receiving Subsystem consists of the terminals connecting the Propagation Medium Subsystem to the Receiving Subsystem, the transducer which converts the electrical signals to sound, the earpiece and the human listener to the messages.

The Availability of the System is the product of the Availability of the Transmitting, Propagation Medium and Receiving Subsystems. Each of these subsystem's Availabilities is just the probability that the subsystem is not in corrective maintenance or preventive maintenance, undergoing modifications or in a delay state. (See Figure 6-1 which is the time relationship diagram taken from MIL-STD-721B). The definition of each of these time elements can be found in MIL-STD-721B and Appendix H. The reason that the subsystem's Availability is dependent only on the above mentioned items is that there is no check made of the system prior

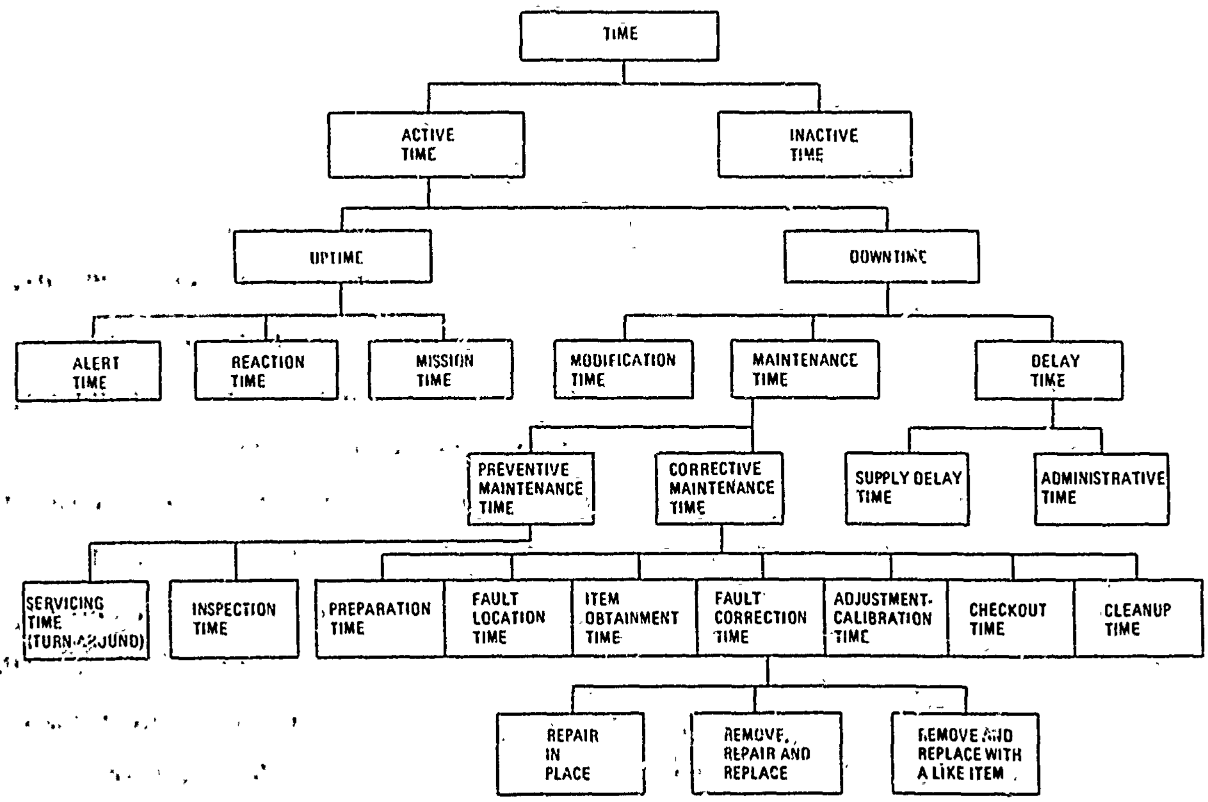


Figure 6-1. Time relationship

to its use for this mission. For example, for aircraft there are aircrew and maintenance preflight checks which can affect an aircraft system's Availability.

The Dependability of the System is the product of the Dependability of the Transmitting, Propagation Medium, and Receiving Subsystems. The Dependability of the Transmitting Subsystem is the probability of not aborting the mission (a Point-to-Point Telecommunication) due to a malfunction of the Transmitting Subsystem. This probability is made up of the following seven probabilities: The probability of not aborting the mission due to a malfunction in the Transmitting Subsystem's

a. Operators,

b. Modulation,

c. Noise Level,

d. Output power level,

e. Distortion level,

f. Connectivity, and

g. Power Supply

The first of these probabilities refers to the probability of not aborting the mission for some reason such that the operator (talker) could not complete the communication (e.g. he became sick). The second of these probabilities is the probability that the modulation level does not go below some specified level which will result in an abort. The third of these probabilities is the probability that the noise level does not exceed a specified value and result in an abort. The fourth probability is the probability that the power output will not go below a specified minimum level and result in an abort. The fifth is the probability that the distortion introduced by the transmitting subsystem does not go above a specified maximum level resulting in an abort. The sixth probability is the probability of not aborting the mission due to the Transmitting Subsystem's connection to the Propagation Medium Subsystem. The last probability is the probability of not aborting the mission due to the system's batteries.

The Dependability of the Propagation Medium Subsystem is the probability of not aborting the mission due to a malfunction in the Propagation Medium Subsystem. This probability is the product of the probabilities of not aborting the mission due to a malfunction in the Propagation Medium Subsystem's

a. Continuity,

b. Attenuation level,

c. Noise level,

d. Distortion level, and

e. Connectivity.

The first of these probabilities is the probability of not aborting the mission due to the propagating medium losing continuity during the communication. The second is the probability of the mission not being aborted due to the attenuation of the propagating medium increasing above a specified level. The third is the probability of not aborting the mission due to the noise level introduced by the Propagation Medium Subsystem being above a specified minimum level. The fourth probability is the probability of not aborting the mission due to the distortion level introduced by the Propagation Medium Subsystem being above a specified minimum level. The last is the probability of not aborting the mission due to the connection of the Propagation Medium Subsystem to the Transmitting and Receiving Subsystems during the communication.

The Dependability of the Receiving Subsystem is the probability of not aborting the mission due to a malfunction in the Receiving Subsystem. This probability is made up of the product of the probabilities of not aborting the mission due to a malfunction of the Receiving Subsystem's

a. Operators,

b. Demodulation,

c. Noise level,

d. Gain level,

e. Distortion level, and

f. Connectivity

The first of these probabilities is the probability of not aborting the mission for some reason such that the operator (listener) could not complete the communication. The second is the probability of not aborting the mission due to a malfunction in the Receiving Subsystem's demodulation circuitry. The third is the probability of not aborting the mission due to the Receiving Subsystem's noise level exceeding a specified upper limit. The fourth is the probability of not aborting the mission due to the Receiving Subsystem's gain being below a specified level. The fifth is the probability of not aborting the mission due to the Receiving Subsystem's distortion level being above a specified upper limit. The last (sixth) of these probabilities is the probability of not aborting the mission due to the Receiving Subsystem's failure to connect to the Propagation Medium Subsystem.

All references to "specified levels" mean that the tester must allocate such things as gain, noise, and distortion levels to each of the subsystems. This allocation can be obtained from equipment specifications if no other source exists. However, prior experience with similar systems, or an alternate system's performance, would be a better source for determining how to allocate these various levels. In all cases, short duration excursions which exceed the set levels but which cause no interference with the communication, do not result in an abort of the mission. Therefore, they are not counted against the system. It should be kept in mind that the gain may be less than one in any subsystem and is, therefore, an attenuation. However, for generality, the Receiving Subsystem is referred to as having a gain and the specified gain level is less than one in this example.

The number to be used for the Availability and Dependability of the operators (talker and listener) can be obtained from past experiences with similar systems or can be assumed to be some value. It is explicitly called out so that, when the tester's data is used for such things as force analysis, different values may be assumed but the value used by the tester will be known. This results in the analysis made by persons other than the tester being more realistic or not including some aspects (personnel Availability and Dependability) more than once.

The Capability of the System is made up of two probabilities. These are (1) the probability of the system being capable of handling the message and (2) the probability of the timeliness of the system. In this simple example, the probability of the system being capable of handling the message is the probability of the system having the required fidelity. This probability is made up of two other probabilities: (1) the probability of the system having voice-intelligibility, and (2) the probability of the system having voice naturalness. The first is measured by the ratio of the number of messages received with no errors to the total number of messages sent. The scenario described what constitutes an error. The probability of having voice naturalness can be obtained from the ability of the listener to recognize the voice of the sender or from a measurement of frequency shifts, cutoffs and non-linearity in the frequency response. The probability of timeliness is the probability that the time required for the message is less than the specified maximum value (the time allocated to the message). This maximum value was stated in the scenario.

The tester should determine all of the above probabilities and report them in such a fashion that good and bad points of the system are easily determined.

Assume that out of the 1000 messages the following are the results. The Transmitting Subsystem was available 981 times, the Propagation Medium Subsystem, 992 times; and the Receiving Subsystem, 987 times. The overall System was available 960 times. Thus (rounded to three significant figures):

			95% Lower limit	95% Upper limit
Availability of the Transmitting Subsystems	=	.981	.972	.989
Availability of the Propagation Medium Subsystem	=	.992	.986	.997
Availability of the Receiving Subsystem	=	.987	.979	.993
Availability of the System	=	.960	.947	.971

Assume that out of the 960 times that the system was available, the following were the results of the Dependability measurements.

Subsystem	Source of Malfunction	Number of Malfunctions Out of 960 Attempts
Transmitting	1) Operators	0
	2) Modulation	2
	3) Noise level	1
	4) Output power level	11
	5) Distortion level	3
	6) Connectivity	4

Subsystem	Source of Malfunction	Number of Malfunctions Out of 960 Attempts
Propogation Medium	1) Continuity	7
	2) Attenuation level	1
	3) Noise level	2
	4) Distortion level	2
	5) Connectivity	5
Receiving	1) Operators	1
	2) Demodulation	2
	3) Noise level	2
	4) Gain level	6
	5) Distortion level	2
	6) Connectivity	2

These malfunctions lead to the following Dependabilities and 95% confidence limits.

Subsystem/System	Dependability	95% Confidence Limits
Transmitting	.978	.986 .968
1) Operators	1.00	1.00 .999
2) Modulation	.998	.9998 .994
3) Noise level	.999	1.00 .996
4) Output power level	.989	.994 .981
5) Distortion level	.997	.999 .992
6) Connectivity	.996	.999 .991

Propogation Medium	.982	.990 .973
1) Continuity	.993	.997 .986
2) Attenuation level	.999	1.00 .996
3) Noise level	.998	.9998 .994
4) Distortion level	.998	.9998 .994
5) Connectivity	.995	.998 .989
Receiving	.984	.991 .976
1) Operators	.999	1.00 .996
2) Demodulation	.998	.9998 .994
3) Noise level	.998	.9998 .994
4) Gain level	.994	.998 .988
5) Distortion level	.998	.9998 .994
6) Connectivity	.998	.9998 .994
Complete System	.945	.958 .929

In the above calculations it is assumed that no two malfunctions occurred simultaneously. If this were to happen, each measure should count against the appropriate data element (e.g. noise level and gain level) but only count as one subsystem failure.

Assume that out of the 907 times that the system was available and dependable, the following were the results of measurements of the system's Capability. Of the 907 messages sent, there were 48 with errors. There were 12 times when it was impossible for the listener to recognize the speaker but 10 of these cases corresponded to messages with errors. In only two cases, the messages took longer than the time allowed. These failures did not correspond to either messages with errors or lack of naturalness. Thus the following are the results of Capability measures, each with the 95% confidence limits.

1) The probability of the system having voice intelligibility is

$$\left(\frac{907 - 48}{907}\right) = .947 \qquad \begin{matrix} .961 \\ .932 \end{matrix}$$

2) The probability of the system having the required naturalness is

$$\left(\frac{907 - 12}{907}\right) = .987 \qquad \begin{matrix} .993 \\ .978 \end{matrix}$$

3) The probability of the system having the required timeliness is

$$\left(\frac{907 - 2}{907}\right) = .998 \qquad \begin{matrix} .9998 \\ .994 \end{matrix}$$

The system's Capability is

$$\left(\frac{907 - 52}{907}\right) = .943 \qquad \begin{matrix} .957 \\ .927 \end{matrix}$$

The fact that only 52 failures are counted comes from the fact that, as was pointed out, ten of the naturalness failures were coincident with intelligibility failures.

Thus, the system's measures of performance (with 95% confidence limits) are:

A	=	.960	.971
			.947
D	=	.945	.958
			.929
C	=	.943	.957
			.927

The system's MOE is the product of the above, i.e., A x D x C.

MOE	=	.855	.876
			.833

In the Test Report, the tester should report all of the above results. MOE, A, D, and C should be emphasized with the other data elements perhaps in tabular form. If the table is too extensive, the data elements should be included as an annex to the Test Report. Since none of the data elements makes a major contribution to a low A, D, or C, none need be called out for special attention in the Test Report. If one, or several, had been much lower than the rest, it should be discussed in detail. The discussion should include the cause of the low value, possible fixes and what further testing, if any, is recommended.

If there is an alternate system which can be used for the mission as described in the scenario, then the tester should report its MOE, A, D, and C. The values of the data elements should also be reported so that the decision maker knows exactly where there are differences between the two systems.

The following sections will discuss the determination of MOE's for Communications/ Electronics Systems, Defense Suppression Systems, Aircraft Systems, and Missiles.

6. MEASURES OF EFFECTIVENESS FOR COMMUNICATIONS/ELECTRONICS SYSTEMS

To obtain standard MOE's for Communications/Electronics (CE) Systems, it must be clear what is, and is not, included in CE. According to AFM 11-1 Vol. III entitled "Communications-Electronics Terminology", "Communications-Electronics is the broad field of activity encompassing the functions of program formulation, policy planning, inspection, and direction of communications-electronics operation and maintenance. It includes supervisory and technical responsibilities for the construction, installation, operation, and maintenance of communications and electronics systems and equipment. It further includes all radio, wire, and other means used for the electrical and visual transmission and reception of information or messages in the clear or by cryptographic means; all radar and radiation aids to air traffic control and navigation and enemy aircraft warning and interception; all ground electronic devices and systems for the control and tracking of aircraft and guided missiles, electronic weather equipment, electronic countermeasures devices, and related electronic systems and equipment." Recall that a complete system includes equipment, related facilities, material, software, services and personnel. For a CE system the equipment includes all the materiel listed above.

It is possible to make a simplifying assumption about all CE Systems, and that is, that each one is always made up of Transmitting, Propagation Medium, and/or Receiving Subsystems. It is not necessary that all CE systems have all three subsystems but they always have at least one of the three.

A Transmitting Subsystem is the subsystem that takes the information to be sent, processes it and inputs it into the propagation medium. It therefore includes its power supply, cooling, operators, transmitter input (microphone, key, sensors, A-to-D converter, etc.), encrypting, modulating, oscillator, power amplifier and feed network (antenna, antenna drive, output matching network, etc.). It will, by definition, include the power supply of systems that have a single supply for both the transmitter and receiver.

A Propagation Medium Subsystem is that subsystem used to connect the output of the Transmitting Subsystem to the input of the Receiving Subsystem. It therefore includes space, earth, and water, as well as coaxial cable, field wire, waveguides, light pipes, etc. Natural media must be included in the system since, for example, ionospheric conditions can prevent communications even when the equipment is working.

A Receiving Subsystem is that subsystem that takes the output from the Propagation Medium Subsystem, processes it and returns it to useful information. It therefore includes the input network (antenna, antenna drive, input matching network, etc.), tuning, amplifying, demodulating, decrypting, output equipment (speakers, printers, visual displays, etc.), and cooling. It includes its power supply only when the Receiving Subsystem has a power supply separate from a Transmitting Subsystem.

With these definitions, it is possible to have a small set of data elements for Availability and Dependability that hold for all CE Systems as will be seen from the discussion of a CE System's Availability and CE System's Dependability. Care must be taken to insure that all items of the system are included, but only once, in the subsystems.

Figure 6-2 shows the relationship between a CE System's MOE and its Availability, Dependability, and Capability. As stated earlier, this breakdown of MOE's into the measures of performance A, D, and C is to be done for all systems. Availability addresses the probability of a system being in an up condition but in an uncommitted state. Dependability addresses a system when it is in a committed state.

The CE System's Availability is broken down into the Transmitting, Propagation Medium, and Receiving Subsystem's Availabilities as shown in Figure 6-3.

Those data elements that make up the Transmitting Subsystem's Availability are shown in Figure 6-4. The Operator's Availability is the probability that the required operators for the Transmitting Subsystem are available. The probability that the Transmitting Subsystem is not down is obtained as is shown in Figure 6-4 where the terms are as defined in MIL-STD-721B (see also Figure 6-1) except for "Scope Creek"

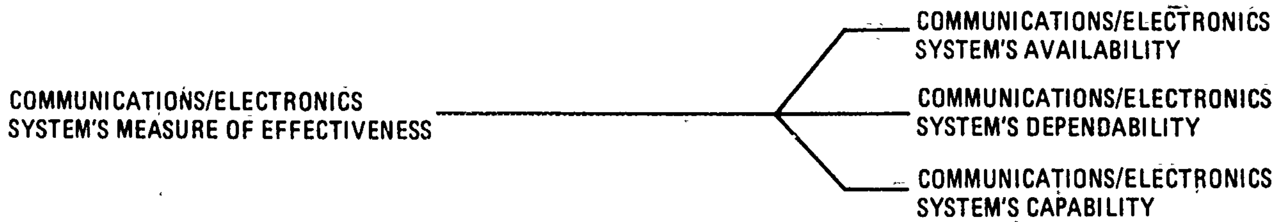


Figure 6-2. CE System's MOE

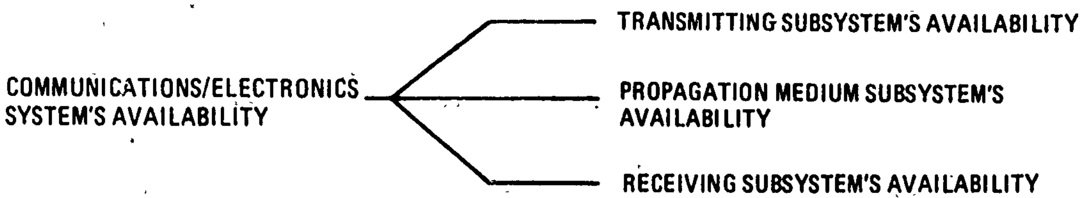


Figure 6-3. CE System's availability

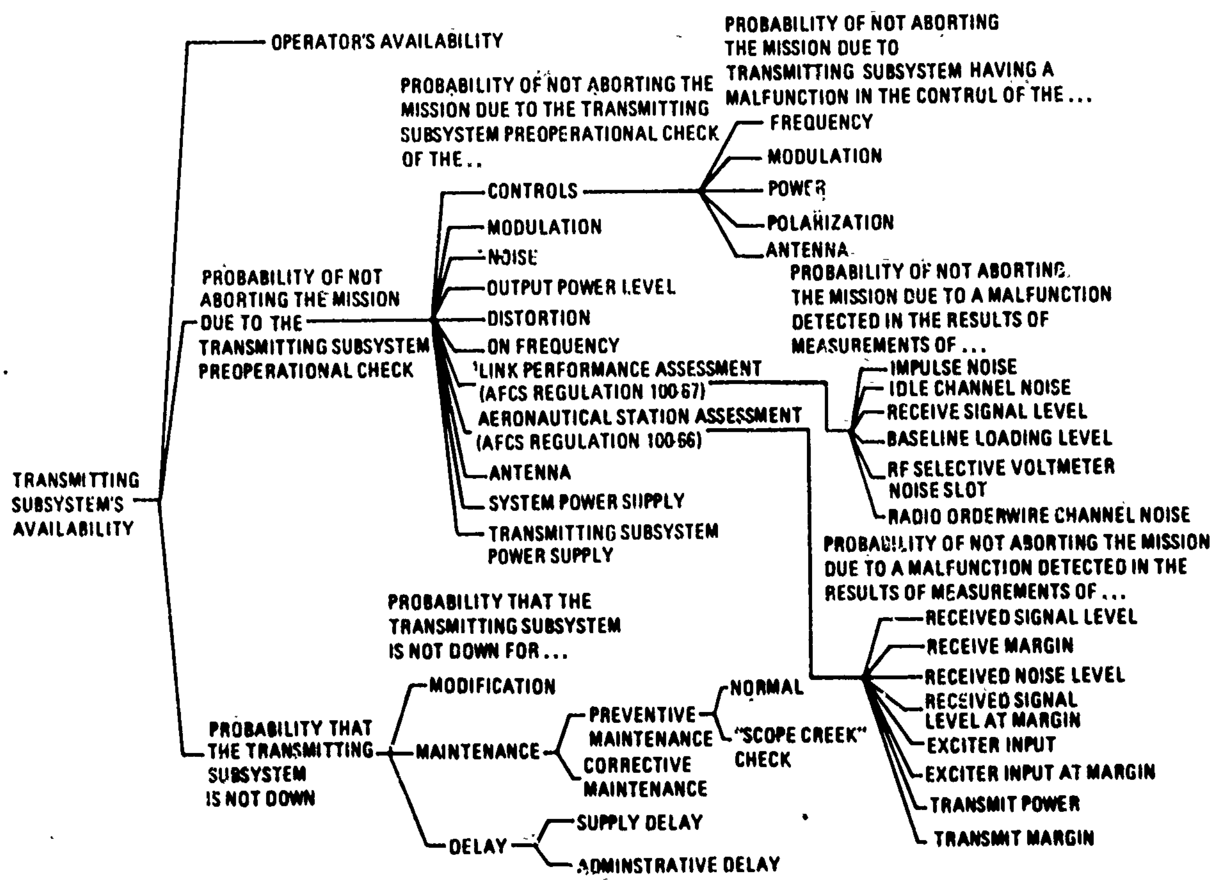


Figure 6-4. Transmitting subsystem's availability

Check. This term refers to the fact that certain checks of a Transmitting Subsystem are included in and may be directed by regulations (e.g., AFCSR 100-66 or 100-67), a general check of some, or all, of the items shown in Figure 6-4, or even no check at all. The tester should select the appropriate data elements for the normal operation of the CE System under test.

The preoperational check of the Transmitting Subsystem's Controls is broken down into frequency, modulation, power, polarization and antenna. This means that the operators have selected the desired frequency, modulation, power, polarization and antenna (choice, direction, tuning, etc.) For example, the transmitter may be working well having been turned to 105 MHz but 150 MHz was what had been requested. This results in the system not being available but there was no malfunction in the equipment, only in the system since it includes the operators. For studies, it is important that this type failure be included.

It must be kept in mind that the data elements (in Figure 6-4, for example) which are not appropriate to the system under test or the scenario should be ignored. The lists in this section are to cover any scenario and CE System, therefore they are, in general, more extensive than will be required by any single test. The tester discusses in the Test Report which ones of this list are selected to be measured due to their applicability to the test objectives and which are selected to be ignored or given assumed values by the tester. All of the above statements of guidance for the tester apply also to the Propagation Medium and Receiving Subsystems.

The Availability of the Propagation Medium Subsystem is determined by the data elements shown in Figure 6-5. Needless to say, the type of Propagation Medium Subsystem (wires, space, light pipes, etc.) has a large influence on the types of tests to be made as well as the data elements to be chosen.

The data elements that make up the Receiving Subsystem's Availability are covered in Figure 6-6. The general guidelines and statements covered under Availability of Transmitting Subsystems apply to Receiving Subsystems also, including the remarks about the subsystem's controls.

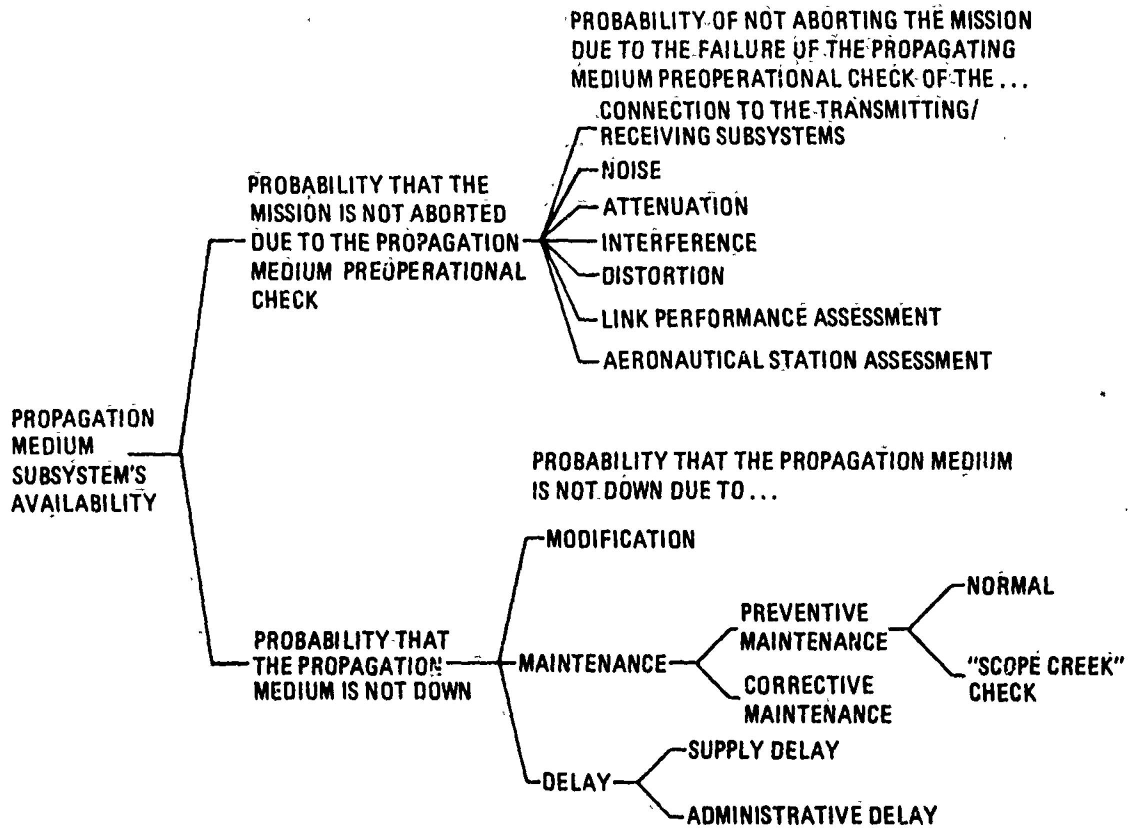


Figure 6-5. Propagation medium subsystem's availability

Having the Availability of the Transmitting, Propagation Medium, and Receiving Subsystems, the CE System's Availability is the product of these three quantities where there are no coincident failures in the subsystems. This is the same result as is obtained by calculating the CE System's Availability under those conditions, since there is to be no overlap in the different subsystems and every item in the system is to be included only once in one or another of the subsystems. Care must be taken to assure that what is to be included in each subsystem is explicitly stated. Also, when evaluating changes and comparing with previous results, the items included in each subsystem should be the same. Where there are coincident failures the tester can calculate the CE System's Availability directly or adjust the inputs to take out the coincident failures in such a manner that they only count as a single abort.

The data elements that make up a CE System's Dependability are shown in Figure 6-7. As stated earlier, the Dependability state starts when the system is in a committed

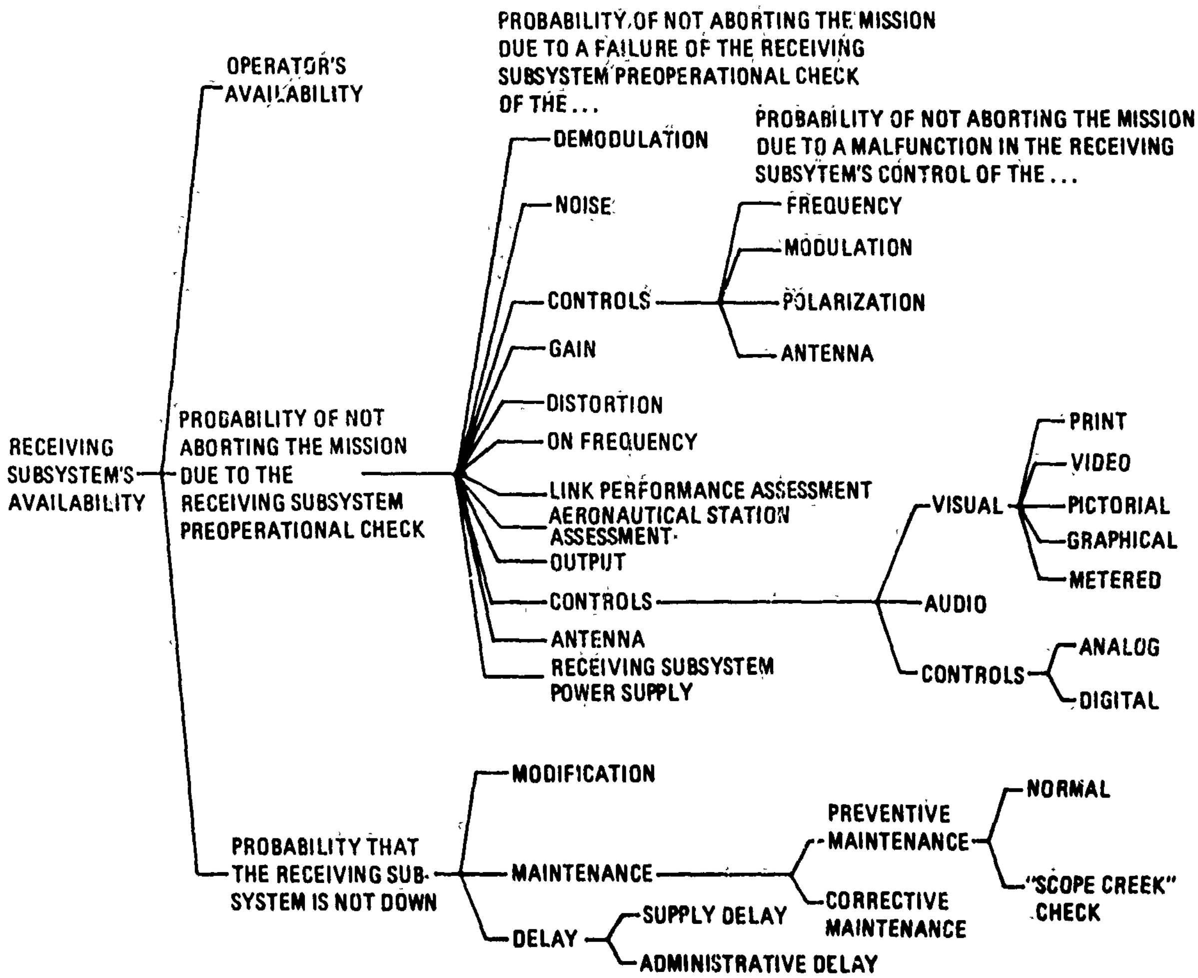


Figure 6-6. Receiving subsystem's availability

state. The comments about assuming values for or ignoring data elements discussed earlier (under CE System's Availability) apply to the Dependability data elements in Figure 6-7. Also, the tester should select only those data elements which apply to his test item, scenario and test objectives as was done for Availability data elements. The data element "Vulnerability" in each subsystem covers subsystem malfunctions due to an unnatural hostile environment to which it should be invulnerable. The "Survivability" data element means that the subsystem has a greater degradation to the man-made hostile environment than should be expected when the subsystem is operating normally. According to AFM 11-1, vulnerability means "The characteristics of a system which causes it to suffer a definite degradation (incapability to perform the designated mission) as a result of having been subjected to a certain

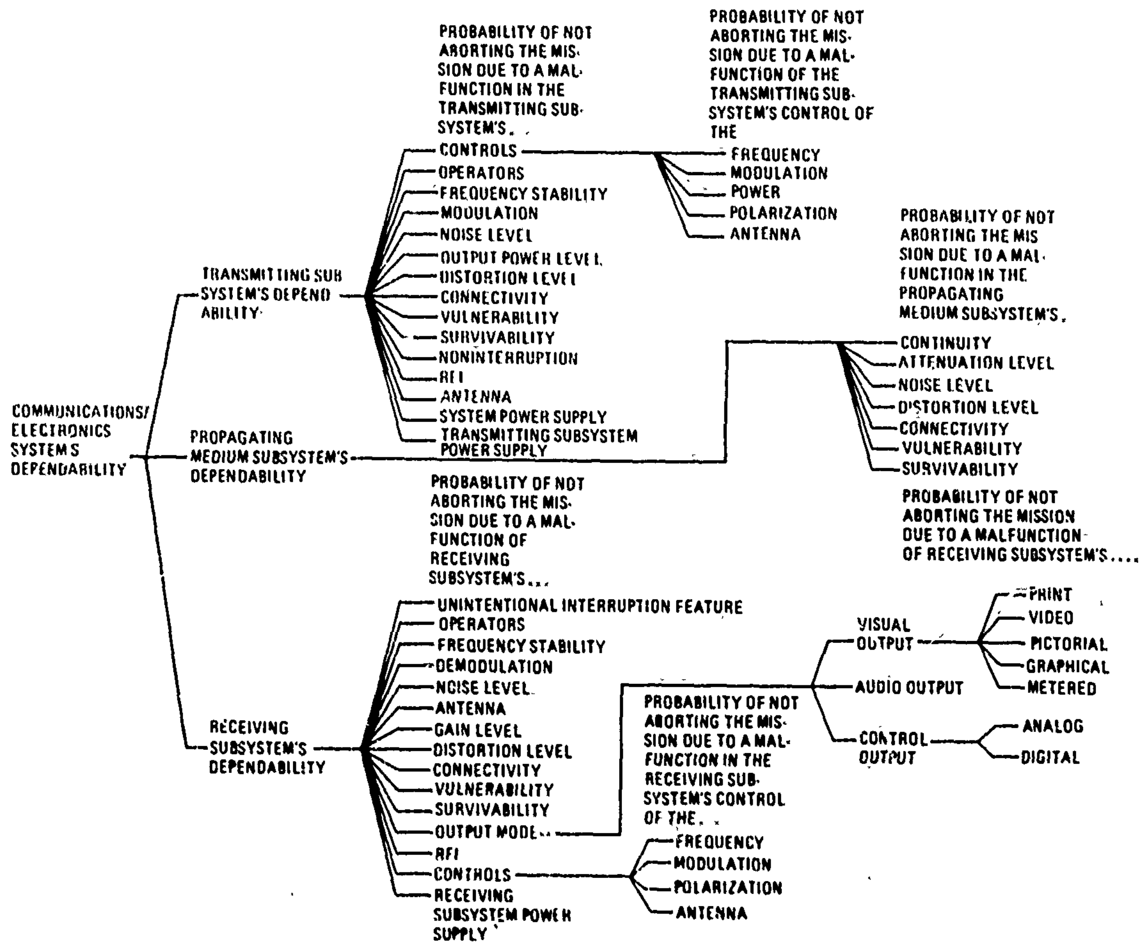


Figure 6-7. CE System's dependability

level of effects in unnatural (manmade) (sic) hostile environment." It also defines survivability as, "The capability of a system to withstand a man-made hostile environment without suffering an abortive impairment of its ability to accomplish its designated mission."

The Capability of a CE System depends upon its mission. In Figure 6-8 there is a listing of the missions of CE Systems for which they have a Capability. As will be seen shortly, many of these missions have very similar data elements. For example, Capability for Air Traffic Control is made up of Point-to-Point Telecommunications

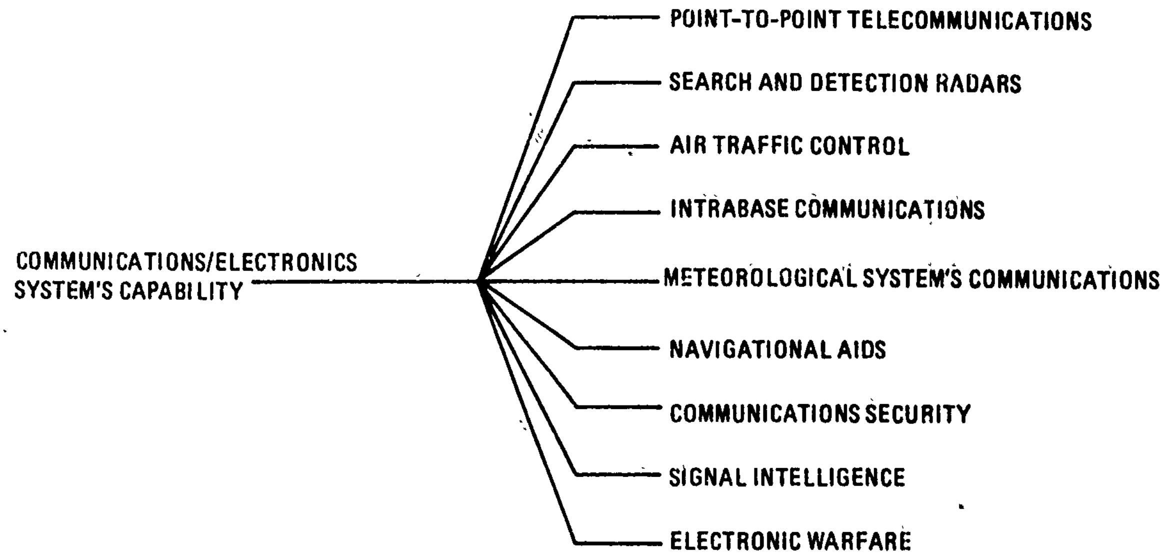


Figure 6-8. Communications/Electronics System's capability

Capability and Search and Detection Radar Capability. The latter is further detailed as Radar Capability for Aircraft Control. Some of the definitions of the terms used in these missions are:

a. Communications Security is the protection resulting from all measures designed to deny unauthorized persons information of value which might be derived from the possession and study of telecommunications, or to mislead unauthorized persons in their interpretation of the results of such possession and study. COMSEC includes: cryptosecurity, transmission security, emission security, and physical security of communications security materials and information (Ref. AFM 11-1 Vol. IV).

b. Signal Intelligence is a generic term which includes both communication intelligence and electronic intelligence. Also called SIGINT. See also intelligence (Ref. JCS Pub. 1).

c. Communications Intelligence is the technical and intelligence information derived from foreign communications by other than the intended recipients. Also called COMINT (Ref. C1, JCS Pub. 1).

d. Electronic Intelligence is the intelligence information product of activities engaged in the collection and processing, for subsequent intelligence purposes, of foreign, noncommunications, electromagnetic radiations emanating from other than nuclear detonations and radioactive sources. Also called ELINT (Ref. C1, JCS Pub. 1).

The discussion of CE Systems Capability will start with Point-to-Point Communications Capability, and Search and Detection Radar Capabilities. Figure 6-9 shows the data elements for Point-to-Point Telecommunications Capability. Since these data elements measure different attributes of the system, they should be unweighted and reported separately. The probability of the system being capable of handling the message depends upon the system's fidelity, the message not getting lost, and the system having the required service features. Timeliness is not considered in message handling capability. Fidelity means the exactness with which the information in the output of the Receiving Subsystem represents the information in the input to the Transmitting Subsystem. For voice transmission, probability of fidelity is made up of probabilities of voice intelligibility and voice naturalness. Probability of voice intelligibility is the probability of sentence (or word) recognition as determined by test listeners when standard text is spoken by test talkers (See Appendix G). Since naturalness is the degree to which the received speech sounds like the unprocessed human voice, the probability of voice naturalness is the probability of the listener recognizing the talker's voice or the probability that the system has frequency shifts, cutoffs, and non-linearity of frequency response within specified limits. For non-voice transmissions, probability of fidelity is the probability that the degradation is less than the specified maximum allowed. How this specified maximum is obtained will be discussed later. The degradation may be due to one, or more, of the following:

a. The sensor calibration being off due to drift, miscalibration, nonlinearity, etc;

b. The bit error rate in digital transmissions;

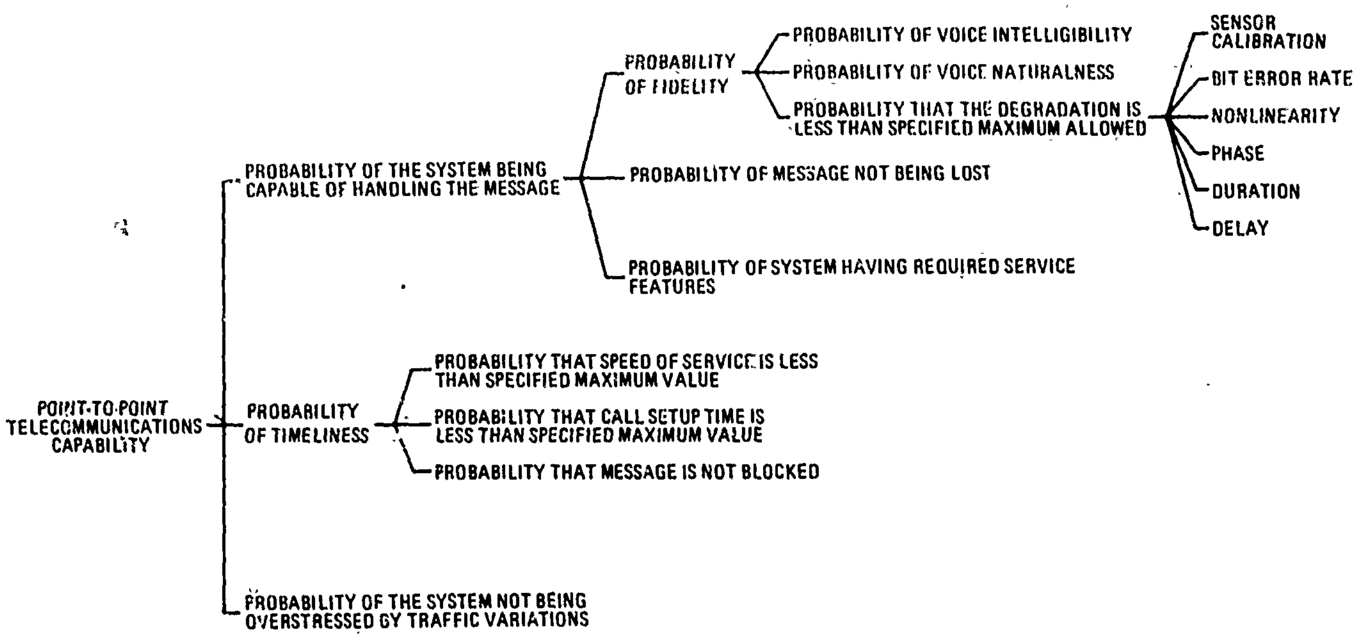


Figure 6–9. Point-to-point telecommunications capability

c. Nonlinearity in the system;

d. Phase response where phase contains the information;

e. Duration where signal duration contains the information; and

f. Delay where the delay contains the information.

The probability of the message not being lost is the probability that the messages or calls that are accepted by the system are not lost or misrouted. Probability of the system having the required service feature is always either zero or one. For example, if the message requires a secure link and the system does not have one, the capability of the system for handling the message is zero.

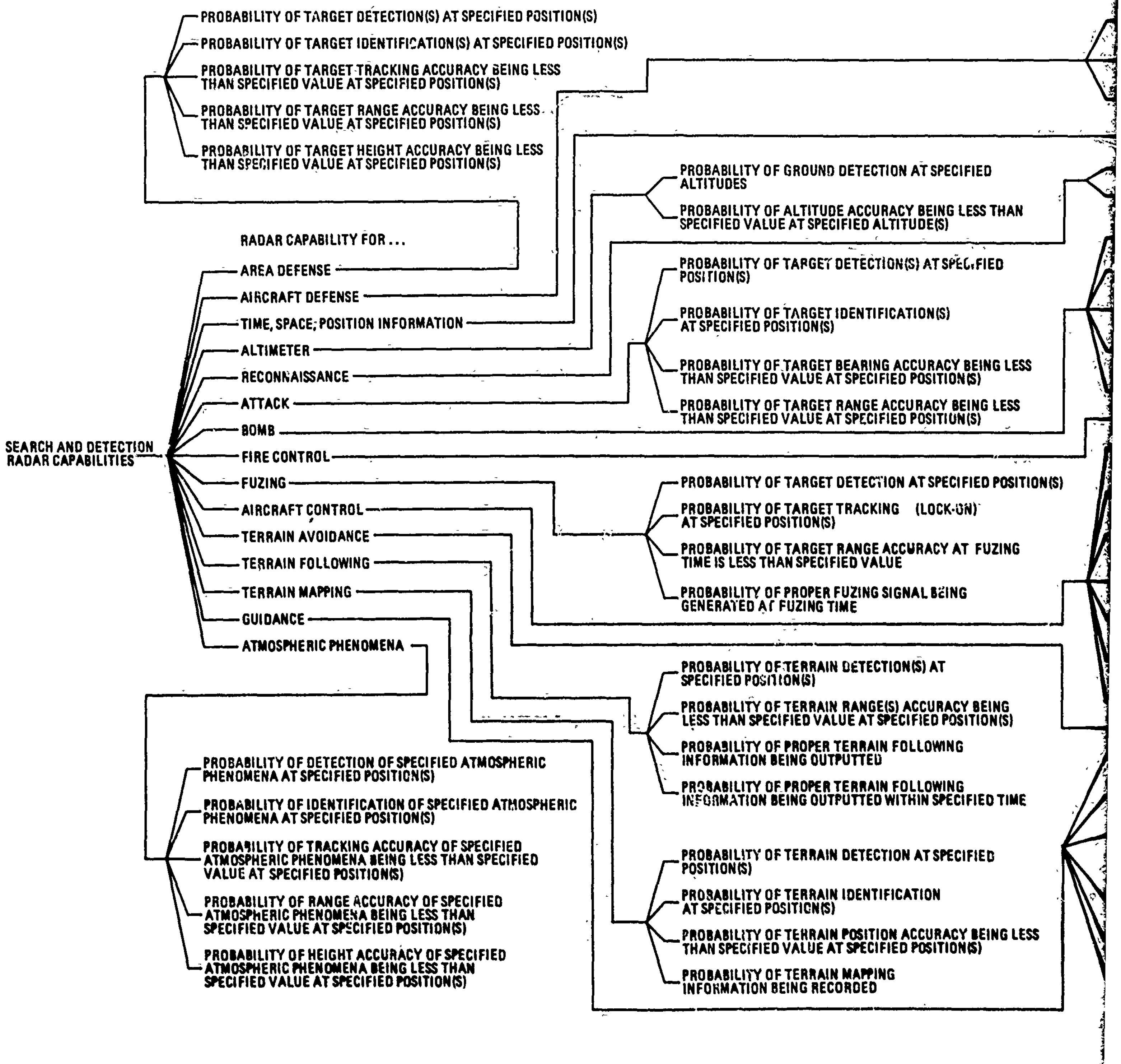
Probability of timeliness is made up of three probabilities;

a. The probability that the speed of service is less than the specified maximum value,

b. The probability that call setup time is less than the specified maximum value, and

c. The probability that the message is not blocked.

The speed of service is measured by the time a message requires to move through the system from the first bit into the Transmitting Subsystem to the last bit out of the Receiving Subsystem or its equivalent. Call setup time is measured by the time required to establish a complete circuit from the sender to the receiver given the call is not blocked. For example, from the last digit dialed by the talker until the listener is on the other end of the circuit. The precise meaning must be given explicitly by the tester. A message is blocked when the system is unable to handle the message due to the system's existing load. The probability of the system not being overstressed by traffic variations is the probability that the system's capability is not degraded by variations in traffic characteristics which are sufficiently short term that they must be served by the inplace system. This includes such things as variations in the volume, the mix, or the geographical and time distributions of the traffic. Figure 6-9 shows all of the data elements that make up the Capabilities of Point-to-Point Telecommunications. The tester should select those that apply to his system, scenario, and test objectives. He should set, or have set for him, before the test, all specified values called for in the data elements (e.g. maximum degradations and times) and state the source of these limits. The source can be system specifications, an alternative new system's values, present system values, desired values, or reasonable values from past experience. Many of the other Communications/Electronics Systems utilize Point-to-Point Telecommunications as subsystems so that the data elements must apply to many missions.


SEARCH AND DETECTION RADAR CAPABILITIES
RADAR CAPABILITY FOR . . .
AREA DEFENSE
AIRCRAFT DEFENSE
TIME, SPACE, POSITION INFORMATION
ALTIMETER
RECONNAISSANCE
ATTACK
BOMB
FIRE CONTROL
FUZING
AIRCRAFT CONTROL
TERRAIN AVOIDANCE
TERRAIN FOLLOWING
TERRAIN MAPPING
GUIDANCE
ATMOSPHERIC PHENOMENA
PROBABILITY OF TARGET DETECTION(S) AT SPECIFIED POSITION(S)
PROBABILITY OF TARGET IDENTIFICATION(S) AT SPECIFIED POSITION(S)
PROBABILITY OF TARGET TRACKING ACCURACY BEING LESS THAN SPECIFIED VALUE AT SPECIFIED POSITION(S)
PROBABILITY OF TARGET RANGE ACCURACY BEING LESS THAN SPECIFIED VALUE AT SPECIFIED POSITION(S)
PROBABILITY OF TARGET HEIGHT ACCURACY BEING LESS THAN SPECIFIED VALUE AT SPECIFIED POSITION(S)
PROBABILITY OF GROUND DETECTION AT SPECIFIED ALTITUDES
PROBABILITY OF ALTITUDE ACCURACY BEING LESS THAN SPECIFIED VALUE AT SPECIFIED ALTITUDE(S)
PROBABILITY OF TARGET DETECTION(S) AT SPECIFIED POSITION(S)
PROBABILITY OF TARGET IDENTIFICATION(S) AT SPECIFIED POSITION(S)
PROBABILITY OF TARGET BEARING ACCURACY BEING LESS THAN SPECIFIED VALUE AT SPECIFIED POSITION(S)
PROBABILITY OF TARGET RANGE ACCURACY BEING LESS THAN SPECIFIED VALUE AT SPECIFIED POSITION(S)
PROBABILITY OF TARGET DETECTION AT SPECIFIED POSITION(S)
PROBABILITY OF TARGET TRACKING (LOCK-ON) AT SPECIFIED POSITION(S)
PROBABILITY OF TARGET RANGE ACCURACY AT FUZING TIME IS LESS THAN SPECIFIED VALUE
PROBABILITY OF PROPER FUZING SIGNAL BEING GENERATED AT FUZING TIME
PROBABILITY OF TERRAIN DETECTION(S) AT SPECIFIED POSITION(S)
PROBABILITY OF TERRAIN RANGE(S) ACCURACY BEING LESS THAN SPECIFIED VALUE AT SPECIFIED POSITION(S)
PROBABILITY OF PROPER TERRAIN FOLLOWING INFORMATION BEING OUTPUTTED
PROBABILITY OF PROPER TERRAIN FOLLOWING INFORMATION BEING OUTPUTTED WITHIN SPECIFIED TIME
PROBABILITY OF TERRAIN DETECTION AT SPECIFIED POSITION(S)
PROBABILITY OF TERRAIN IDENTIFICATION AT SPECIFIED POSITION(S)
PROBABILITY OF TERRAIN POSITION ACCURACY BEING LESS THAN SPECIFIED VALUE AT SPECIFIED POSITION(S)
PROBABILITY OF TERRAIN MAPPING INFORMATION BEING RECORDED
PROBABILITY OF DETECTION OF SPECIFIED ATMOSPHERIC PHENOMENA AT SPECIFIED POSITION(S)
PROBABILITY OF IDENTIFICATION OF SPECIFIED ATMOSPHERIC PHENOMENA AT SPECIFIED POSITION(S)
PROBABILITY OF TRACKING ACCURACY OF SPECIFIED ATMOSPHERIC PHENOMENA BEING LESS THAN SPECIFIED VALUE AT SPECIFIED POSITION(S)
PROBABILITY OF RANGE ACCURACY OF SPECIFIED ATMOSPHERIC PHENOMENA BEING LESS THAN SPECIFIED VALUE AT SPECIFIED POSITION(S)
PROBABILITY OF HEIGHT ACCURACY OF SPECIFIED ATMOSPHERIC PHENOMENA BEING LESS THAN SPECIFIED VALUE AT SPECIFIED POSITION(S)

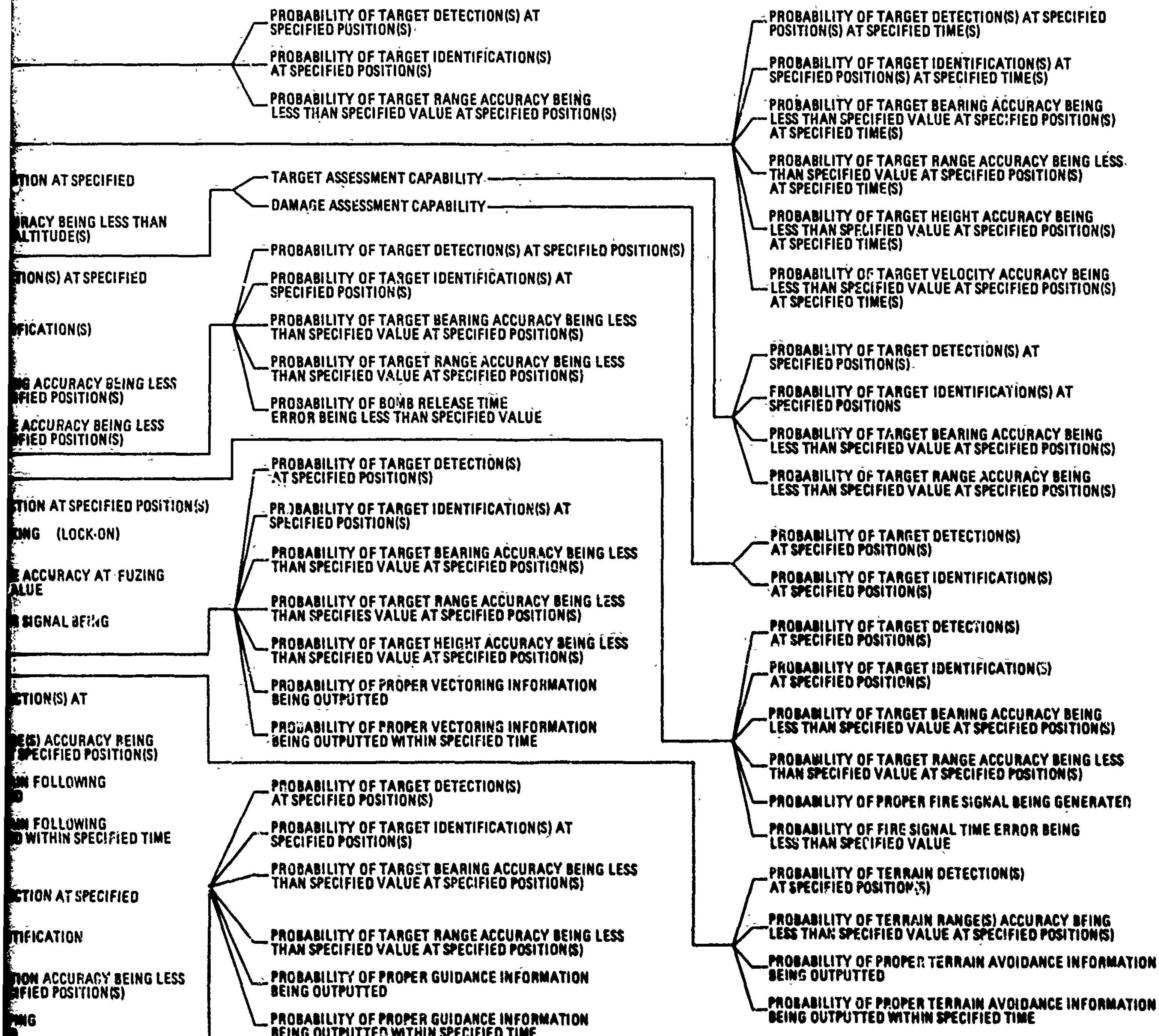


Figure 6-10. Search and Detection Radar Capabilities

A second type CE System that is used for many missions as well as being a subsystem in other CE Systems is radar. The missions and the data elements that radars utilize are shown in Figure 6-10. Search and Detection Radar Capabilities are broken down into fifteen different missions. The meaning of the titles of the missions is self evident except, perhaps Radar Capability for Time, Space, Position Information. This capability applies to radars such as are used on ranges for the tracking of aircraft, missiles, etc. In the data elements, position means the location, in three dimensions, of an item, with the origin of the coordinates taken as the system under test. For systems that cover a significant area, the origin should be defined, by the tester, as some significant point within the confines of the system. The scenario should state the target aspect(s). Then a separate capability is measured for each aspect. Again, the tester must set, or have set for him, all of the specified values (e.g., positions, accuracies, times, etc.) The source for these specified values should be, as before, specifications, other systems' values, values desired, or past experience. Although there appears to be a large number of data elements shown in Figure 6-10, about 70% of them are made up of probabilities for detection, identification, tracking, and position measurement. The reason for the repetition is to make the data elements for each mission complete. Many CE Systems refer back to special applications of radar capabilities.

The Capability diagrams for Air Traffic Control, Intrabase Communications, and Meteorological System's Communications are shown in Figure 6-11. As indicated, the tester selects the appropriate data elements from Point-to-Point Telecommunications Capability for all three cases and Search and Detection Radars Capability for two of them. The reason for this is, for example, because Air Traffic Control is just a special mission (application) of Point-to-Point Telecommunications and radars. Radars are covered under Search and Detection Radar. Those aspects of a Meteorological system included in the data elements are their response time, calibration, range and nonlinearity. Response time of the sensors is part of the speed of service since the sensors are part of the Transmitting Subsystem. The dynamic range of the sensor is part of its nonlinearity since its range is usable as long as it is linear.

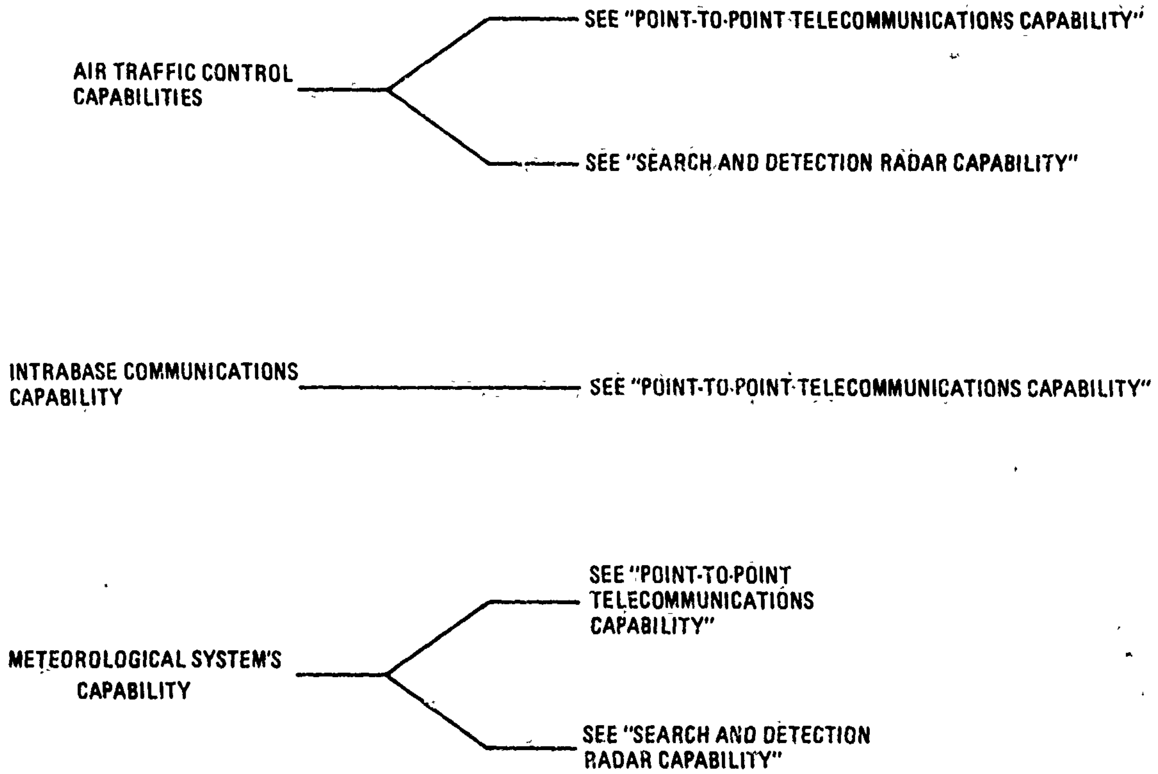


Figure 6-11. ATC control, intrabase communications and meteorological system's capabilities

A major use of CE Systems is as aids to navigation. Figure 6-12 shows the data elements for Navigational Aids Capability. Each type of Navigational Aids System uses data elements based upon its method of operation. Again, the specified values must be set before the test and from the sources mentioned. Where the tester is referred to Point-to-Point Telecommunications, he selects those data elements in Figure 6-9 that apply to the method of operation of the system under test.

The next mission of CE Systems to be considered is Communications Security (See AFM 11-1 Vols I and IV for definitions of the terms used). This mission may apply to any of the other CE Systems. If it is a consideration in other CE systems, then the tester selects those aspects of Communication Security Capability shown in Figure 6-13 that apply to the system under test. For example, a secure TWX transmitter/receiver system might include all of the capability statements in Figure 6-13 except

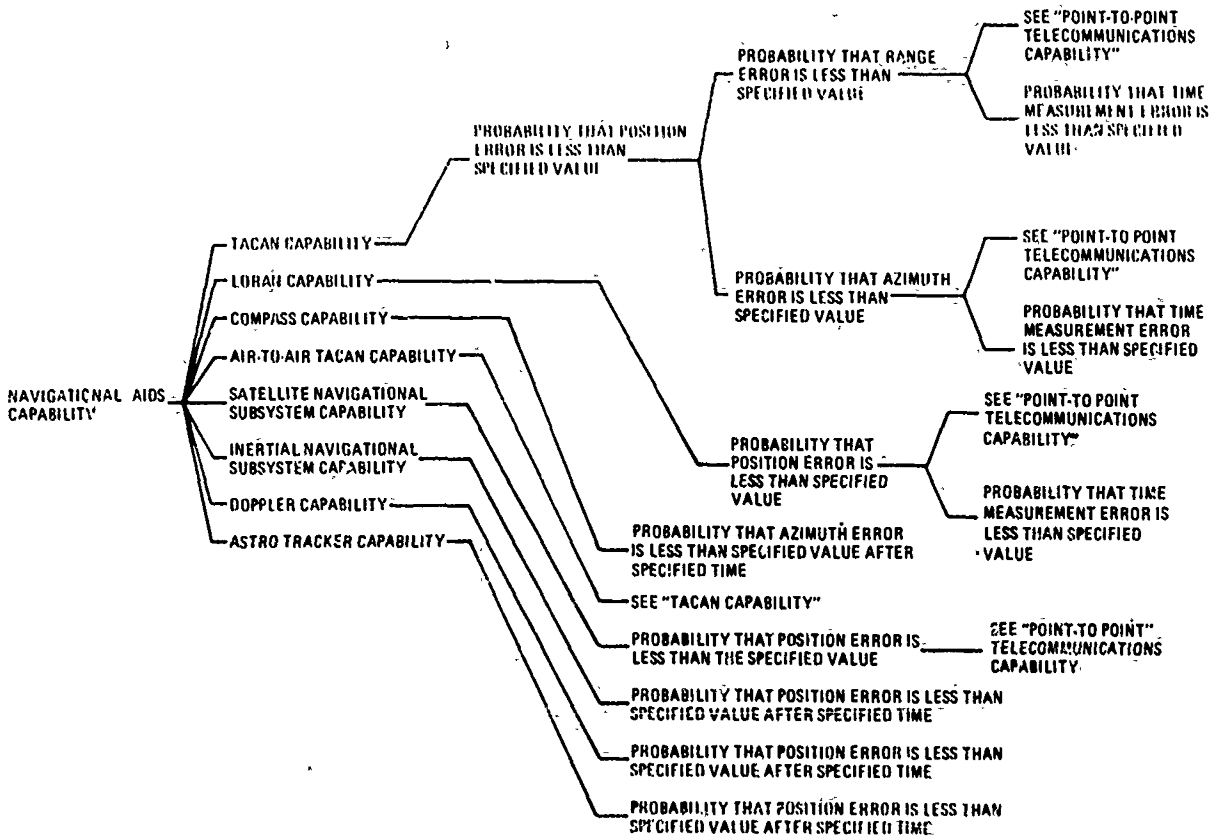


Figure 6-12. Navigational aids capability

Transmission Security Capability. Again, the tester must state what the specified values are and their source. The probabilities listed for Cryptosecurity, and Transmission, Emission and Physical Security are, in general, not obtainable from tests but are derived from evaluations of the system's design.

Signal Intelligence (SIGINT) Capability is covered by the data elements shown in Figure 6-14. SIGINT is either Communications Intelligence (COMINT) or Electronic Intelligence (ELINT). Each of these depends upon obtaining the signal and then obtaining the information in the signal. Again, these data elements may not be derivable from testing but from evaluations. In either case, the tester should make clear the source of the values that he reports. If these data elements are the result

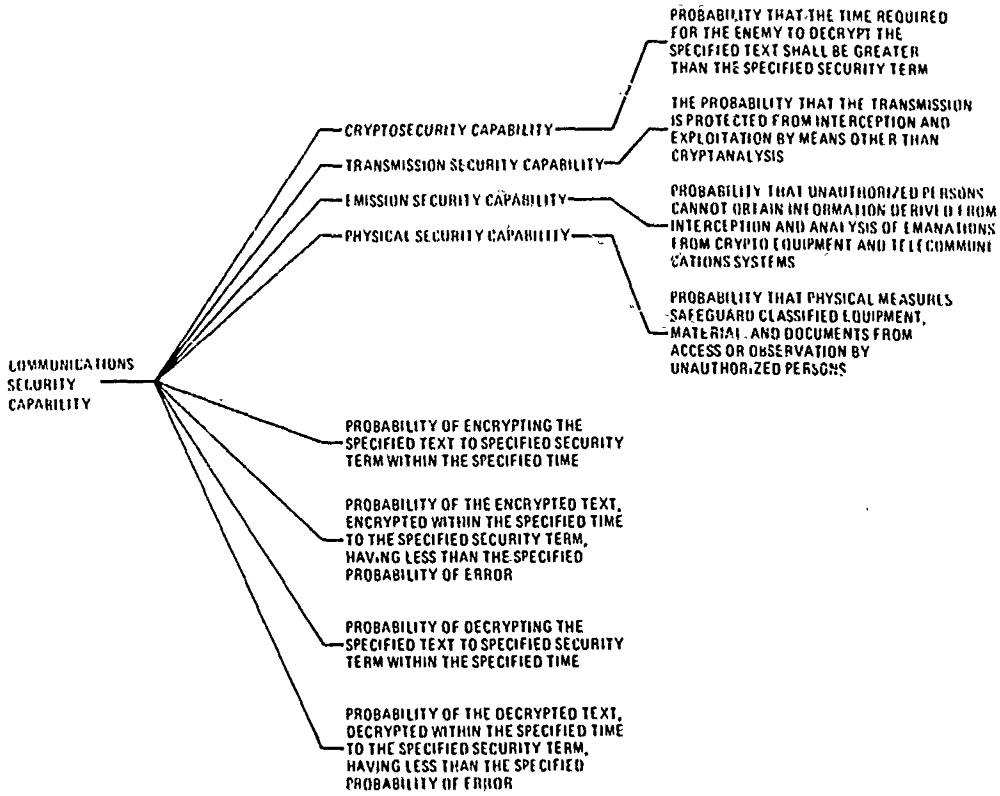


Figure 6-13. Communications security capability

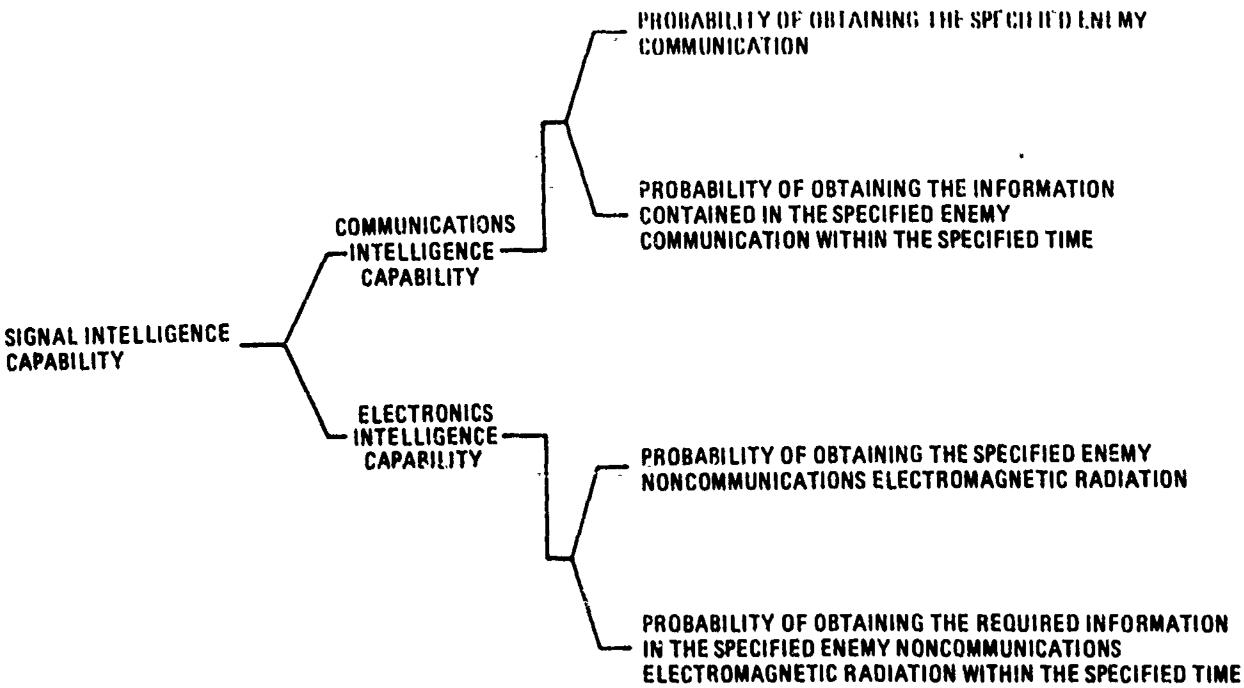


Figure 6-14. Signal intelligence capability

of evaluations (rather than test results), no attempt should be made to standardize the process by which they are derived. It is important to keep these evaluations flexible to allow for rapid changes.

Electronics Warfare Capabilities are covered separately under Defense Suppression in the following section as shown in Figure 6-15. Other connections between CE Systems and Defense Suppression will also be covered.

ELECTRONIC WARFARE CAPABILITY – SEE "DEFENSE SUPPRESSION"

Figure 6-15. Electronic warfare capability

7. MEASURES OF EFFECTIVENESS FOR DEFENSE SUPPRESSION SYSTEMS

As the term Defense Suppression System implies, the purpose for these systems is to decrease the capabilities of the enemy's defenses. From an Air Force point of view, Defense Suppression is accomplished by Air-to-Ground, Air-to-Air, or Electronic Warfare Systems. The first two of these types of systems will be addressed in the next section on Measures of Effectiveness of Aircraft Systems. In this section, only Electronic Warfare Systems will be addressed.

Since most Electronic Warfare (EW) Systems contain transmitters and/or receivers, their Availability and Dependability data elements are already covered in the section on Communications/Electronics Systems. The only EW Systems that are electronically passive are the chaff or flare dispensers and their Availabilities for mating to the aircraft are covered by the data element shown in Figure 6-16. Once the dispenser is mated to the aircraft, it becomes part of the Aircraft System and is covered by the data elements for Aircraft Systems.

CHAFF/FLARE DISPENSER AVAILABILITY	– PROBABILITY THAT THE CHAFF/FLARE DISPENSER PASSES THE PREMATING CHECK.

Figure 6-16. Chaff/Flare dispenser availability

The purpose of defense suppression is to decrease the enemy's defense capabilities. It is, therefore, natural and logical to measure the Capability of a Defense Suppression System by the decrease in the Capability of the EW targeted system. The only other aspects are how fast the system responds and what effects it has on friendly systems. With this in mind, and only addressing EW Systems, the data elements for Defense Suppression Capability are shown in Figure 6-17. As before, the tester must set the specified values called for in each of the data elements from alternate system's values, past experience, desired values, or system specifications.

Unlike some of the data elements in intelligence and security covered above, all of these data elements are measurable. The problem in EW is that the targeted

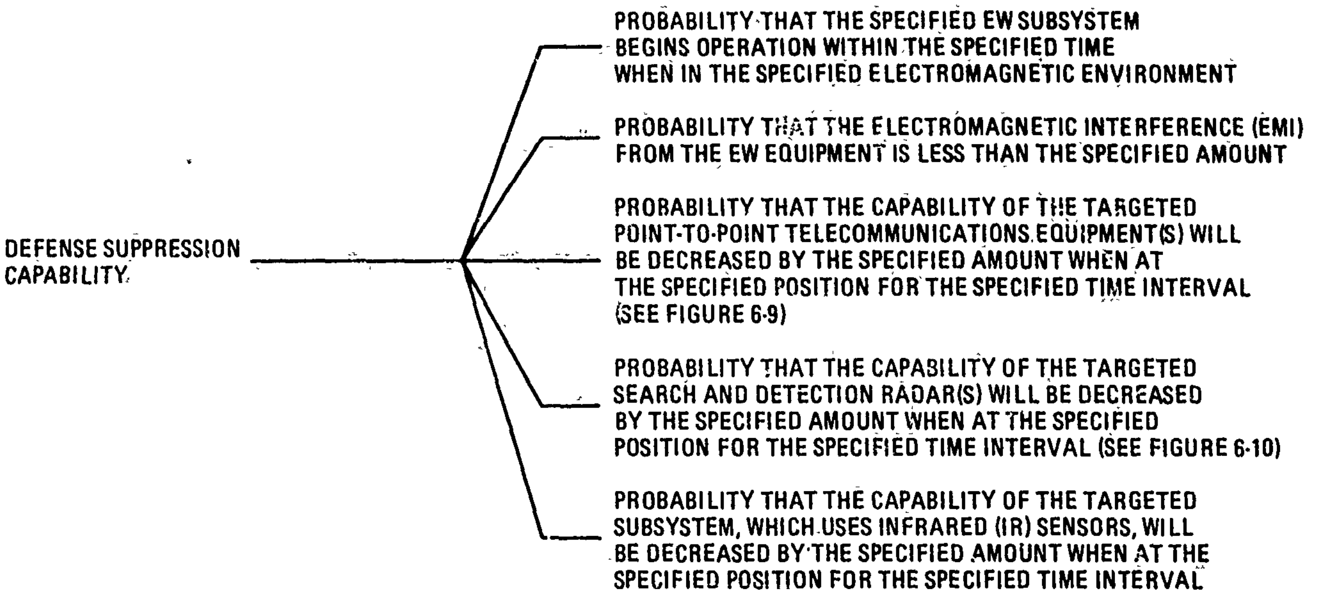


Figure 6-17. Defense suppression capability

system's capabilities are not always known with the highest confidence levels. The application of standard MOE's can be applied to EW Systems working against the threat simulation equipment. However, it must be realized that how well the threat is simulated can always be questioned.

The electromagnetic interference (EMI) referred to in the data elements should be measured for each of the other CE Systems that could be affected by the system being tested.

If the system being tested is actually a subsystem of an Aircraft System, the data elements in Figure 6-17 are used and only address the Aircraft System's EW Capabilities. This subject is discussed under Measures of Effectiveness for Aircraft Systems.

No attempt has been made to standardize the data elements that are inputs to the probabilities called for in Figure 6-17. The reason is that these inputs can be found under CE System's Capabilities. They are dependent upon the design and operation of each piece of EW equipment, the targeted equipment, and the test design. It is important that the measurements taken lead to a determination of the value of the data elements shown in Figure 6-17.

8. MEASURES OF EFFECTIVENESS FOR AIRCRAFT SYSTEMS

Since a system, by definition, is a self-sufficient unit for a mission, the term "Aircraft System", will include the cargo, rockets, bombs, missiles, pods, or whatever load the aircraft is carrying. For this reason, Aircraft Systems are very broad in their applications to missions. Table 6-2 lists Air Force missions and Aircraft Systems missions. It can be seen that there is not a one-to-one correlation between the two. For example, using an Aircraft System for air-to-air combat can be a part of counter air, close air support or combat air patrol Air Force missions. The MOE's used for Aircraft Systems are less redundant if they address the Aircraft System's mission rather than the Air Force mission. The scenario should cover the Air Force mission by statements such as, "an A7-D/AIM-7E Aircraft System, during combat air patrol, engages a MIG-19..."

There is a closer time tie between Availability (A), Dependability (D), Capability (C), and the sortie profile for Aircraft Systems than for Communications/Electronics Systems. For aircraft, Availability will address all operations executed up to the time the engines are to be started. Dependability will cover all operations executed from engine start to engine shut down including postflight aircrew and maintenance checks of the system. After the postflight checks, the Aircraft System is in the Availability portion of the cycle again. Capability addresses those periods of the sortie during which the aircraft missions shown in Table 6-2 are actually being executed. The data elements which address the Capability of the Aircraft System will help clarify exactly what is included in Capability.

The lists of data elements which address A, D, and C in the discussions which follow may appear to be rather extensive. However, it must be kept in mind that they include all Aircraft Systems. The lists include piston and jet engines; single and multiengined aircraft; bombers, fighters, helicopters, transports and trainers; various armaments (guns, rockets, missiles, bombs, guided bombs, etc.); and all other variations of Aircraft Systems. Since these lists must apply to so many systems, the tester must select those that apply to his system and when during the sortie, the data

TABLE 6-2. AIR FORCE VS AIRCRAFT MISSIONS

Aircraft System Missions	Corresponding Air Force Missions
1) Air-to-Air Engagements	1, 2, 5
2) Air-to-Ground Engagements	1, 2, 3, 4, 5
3) Search and Rescue/Recovery	5, 8
4) Airlift	2, 11
5) Command and Control	1, 5, 10, 13
6) Reconnaissance	3, 6, 7, 10, 15
7) Electronic Warfare	1, 3, 4, 5, 6
8) Airborne Atmospheric Sampling	15
9) Training	14
10) Airborne Test Bed	16
11) Refueling	9
12) Battlefield Illumination	2, 12
13) Demonstration Team	13

Air Force Missions	Corresponding Aircraft Missions
1) Counter Air	1, 2, 5, 7
2) Close Air Support	1, 2, 4, 12
3) Air Interdiction	2, 6, 7
4) Fire Suppression	2, 7
5) Combat Air Patrol	1, 2, 3, 5, 7
6) Electronic Warfare	6, 7
7) Reconnaissance	6
8) Search and Rescue/Recovery	3
9) Refueling	11
10) Forward Air Control	5, 6
11) Airlift	4
12) Battlefield Illumination	12
13) Command and Control	5
14) Training	9
15) Weather	6, 8
16) RDT&E	10
17) Demonstration Team	13

elements would normally be checked. Also the tester must interpret the data element statements from a general point of view. He should not expect them to always correlate, word for word, with what his system calls the information he is to gather. For example, his Aircraft System might measure engine speed in units other than RPM to which the data elements refer. However, if he keeps in mind that RPM is referring to engine speed, it makes no difference.

As always, the tester selects only those data elements that apply to his Aircraft System, scenario, and test objectives. If he is testing a subsystem, the tester only selects those data elements that apply to the subsystem under test and other subsystems with which it can interfere.

The measures of performance (A, D, and C) which make up the Measures of Effectiveness for an Aircraft System are shown in Figure 6-18. It is assumed that the tester has been given, or has stated, the scenario for the system. The scenario includes the system's mission. With this in mind the tester can proceed to those data elements that address the Aircraft System's Availability.

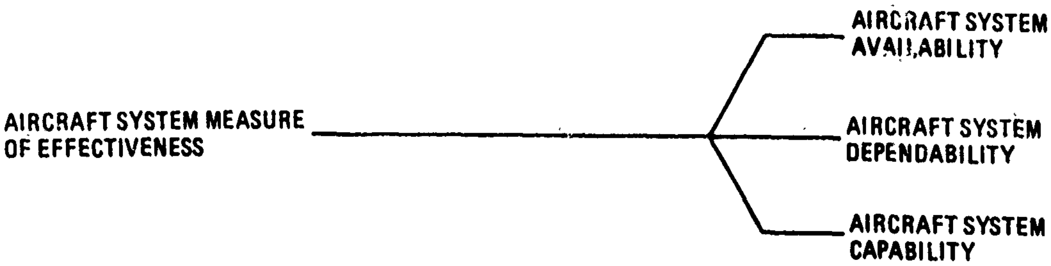


Figure 6-18. Aircraft System MOE

All Aircraft Systems are made up of the Aircraft, the Load, and the Aircrew Subsystems. The Aircraft Subsystem includes the AGE and ground crew. For an aircraft to be available, it must pass the aircrew and maintenance preflight checks as pointed out in Figure 6-19. The load must be available also. Loads are broken down into expendable and nonexpendable loads. Examples of each type are shown in Table 6-3. The details of the data elements for the load's Availability will be covered shortly. The aircrew's

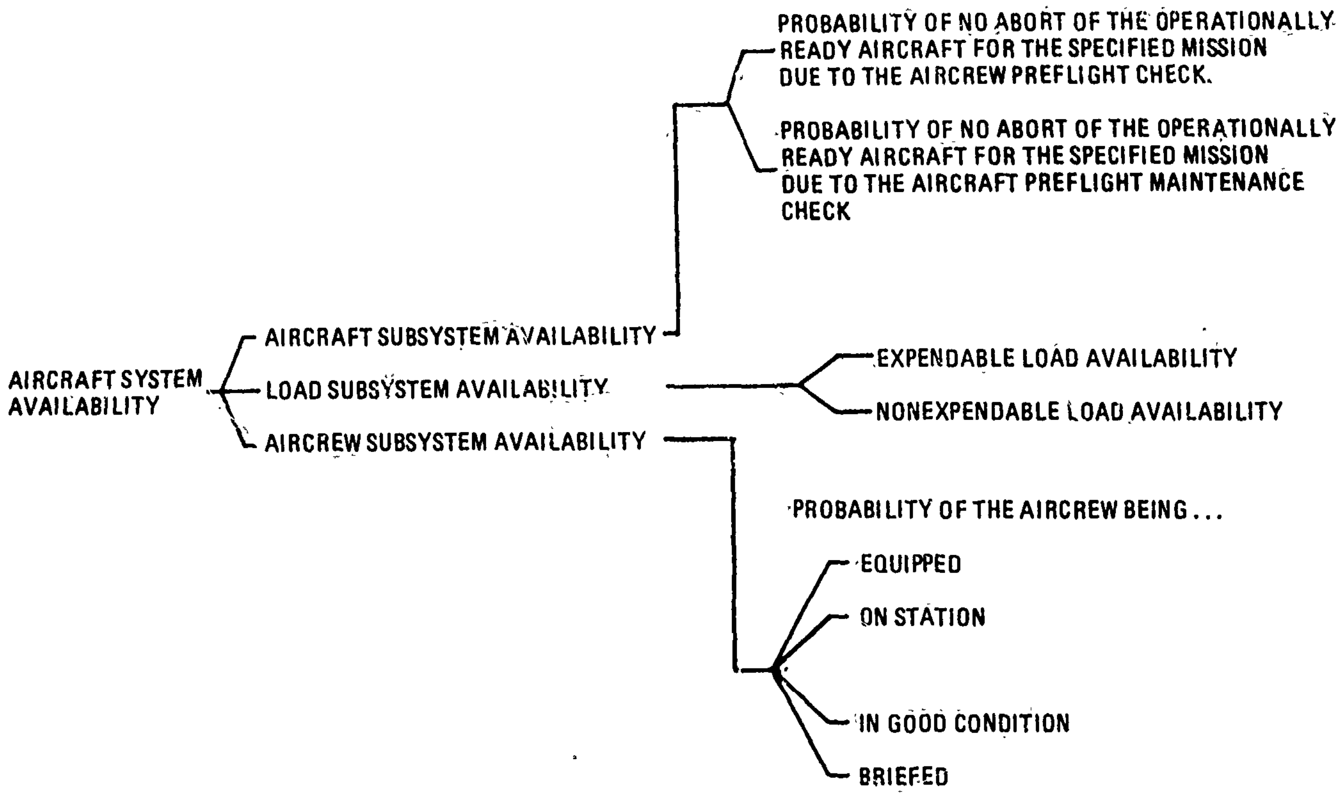


Figure 6-19. Aircraft system availability

Availability is covered by the data elements shown in Figure 6-19. The aircrew is considered equipped when they have the proper required personal equipment such as oxygen masks, helmets, earphones, microphones, etc. The other data elements are self explanatory.

There are several ways for the aircraft to be considered operationally ready which is the precondition of the aircraft called out in the data elements in Figure 6-19. The aircraft could have flown either an operational flight or a check flight. The data elements leading to the aircraft being operationally ready from either of these flights are shown in Figure 6-20. The tester will need to specify the length of time from the flight (operational or check) until the aircraft is to be operationally ready.

TABLE 6-3. TYPES OF AIRCRAFT LOADS

EXPENDABLE LOADS
1) AGM
2) AIM
3) ARM
4) Bombs
5) Guns
6) Decoys
7) RPV
8) CBU
9) Flares
10) SAR Floation Equipment
11) SAR Fire Fighting Equipment
12) Air Dropped Cargo
13) Air Dropped Troops
14) Guided Bombs
15) Fuel for Refueling
16) Film in Dispensers
SUBSYSTEMS AND NONEXPENDABLE LOADS
1) Command, Control and Communications (C^3) Equipment
2) Air Recovery Equipment
3) Cargo
4) Passengers/Troops
5) Sensors (Cameras, Electronic Radiation, Television, Infrared etc.)
6) Fuel
7) Data Collection/Recording Equipment
8) Lights
9) EW Equipment
10) Film for Sensors

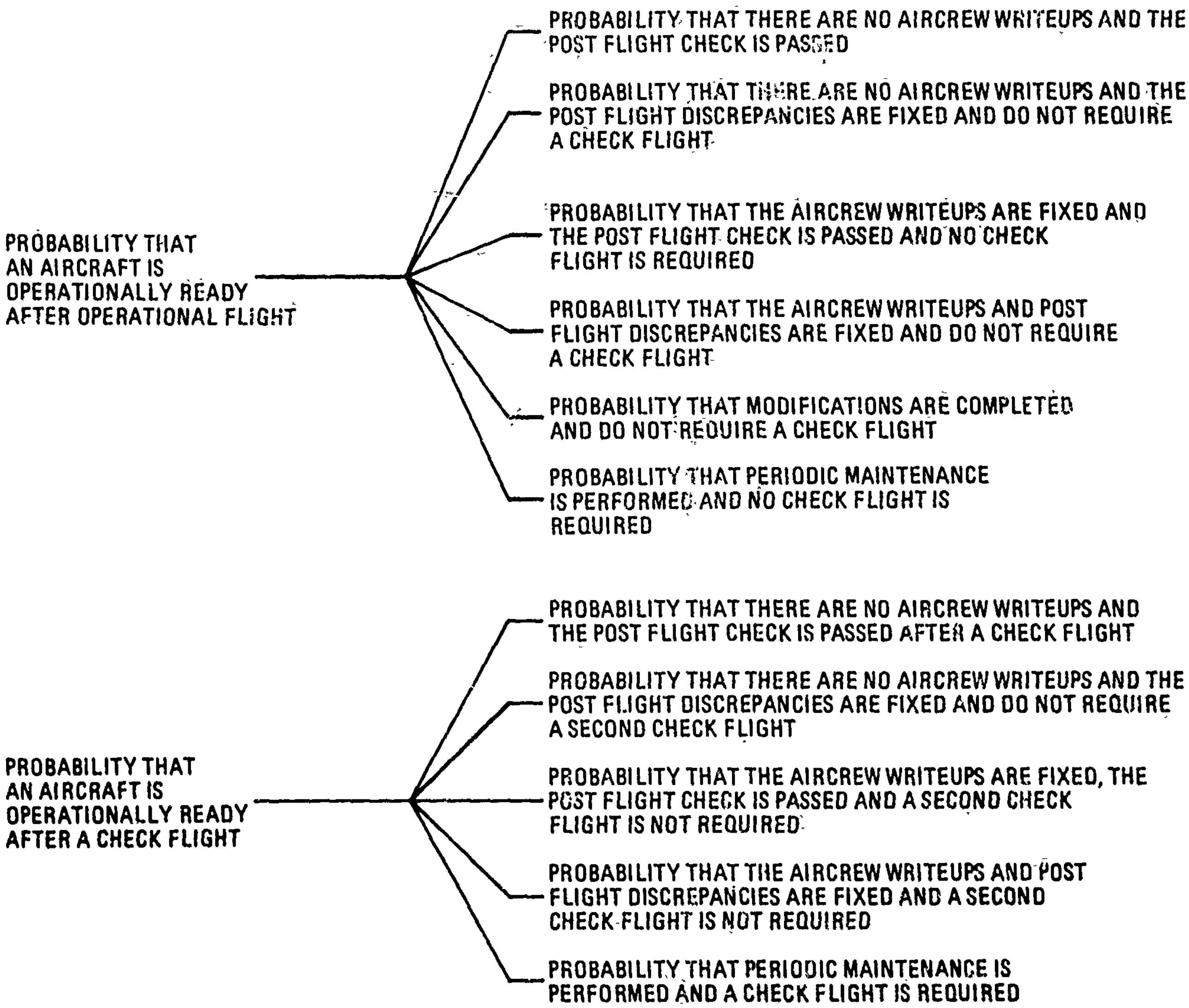


Figure 6-20. Probability that an aircraft is operationally ready after an operational flight or after a check flight.

Once the aircraft is operationally ready and it is called upon for a mission, it undergoes two checks. The first is the preflight maintenance check. The data elements for passing this test are shown in Figure 6-21. Again, these data elements apply to any aircraft so that the tester must select only those that apply to his system. The T.O. checklists are excellent guides for making the choice of data elements for all cases. The aircrew preflight check is the second check. Its data elements are shown in Figure 6-22. In

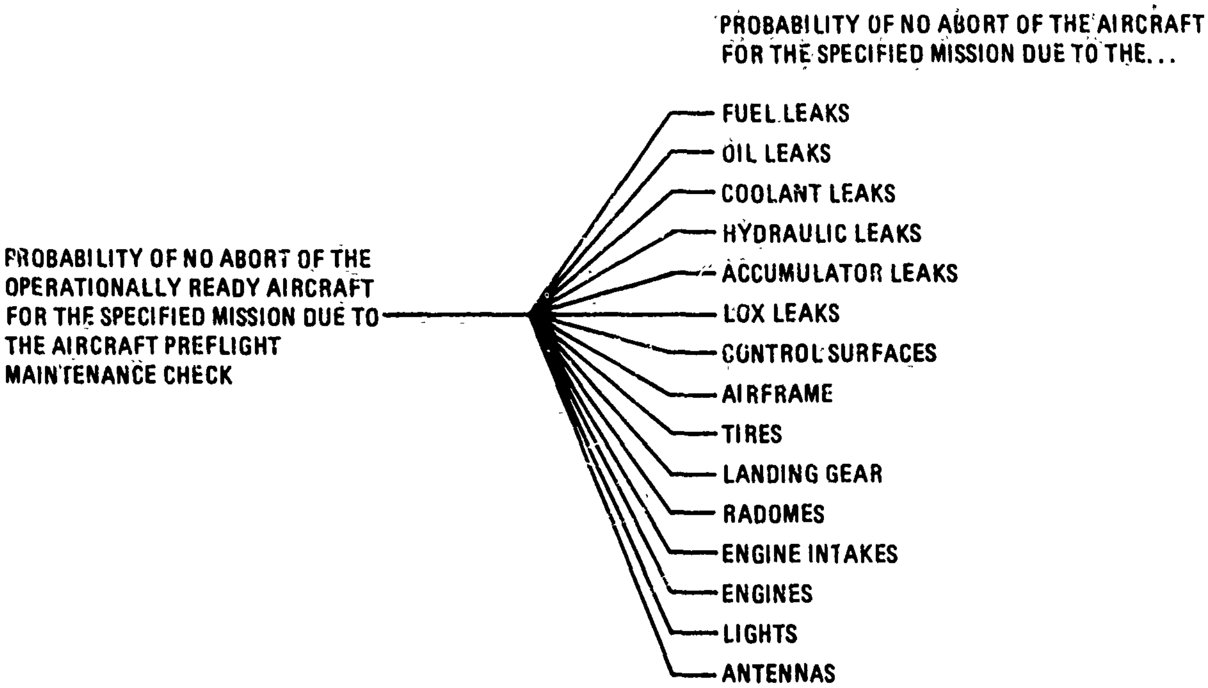


Figure 6-21. Probability of no abort due to preflight maintenance check

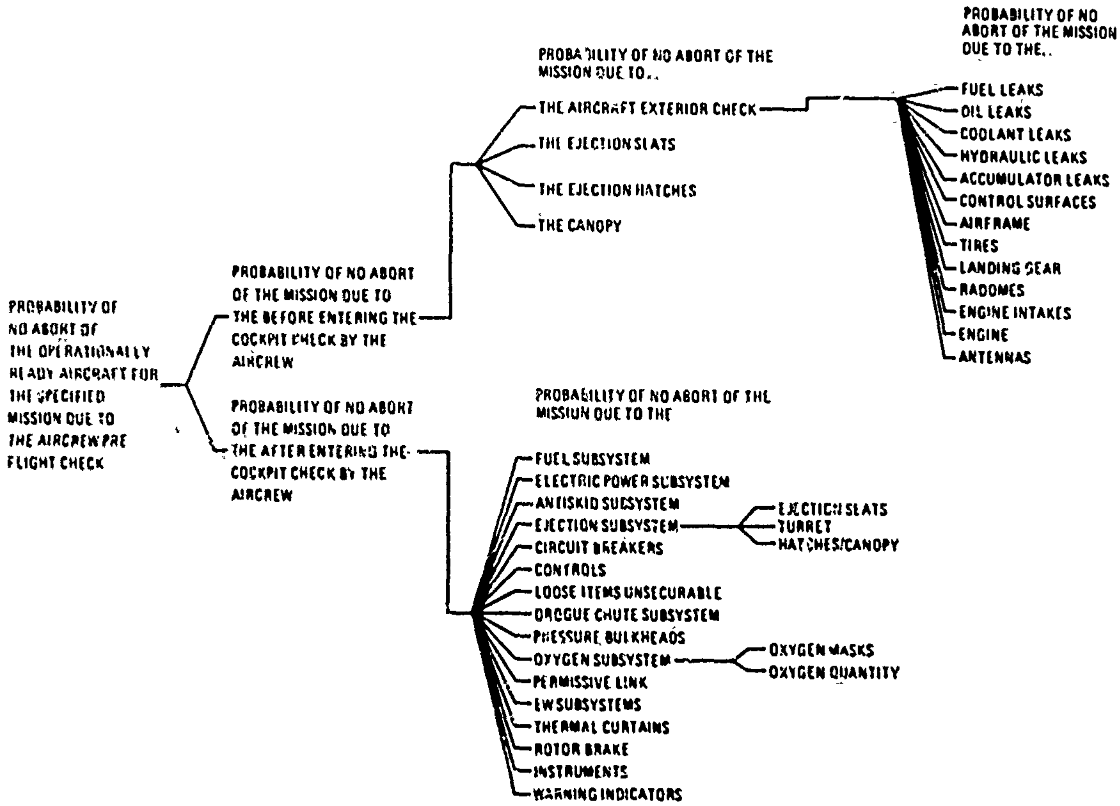


Figure 6-22. Probability of no abort due to the aircrew preflight check.

almost all cases, the aircrew preflight check is broken down into a "before entering the cockpit" and an "after entering the cockpit" check. Any discrepancies of items normally checked at this time, that do not lead to an abort of that particular mission for that aircraft, are not counted against the A of the aircraft but should be recorded by the tester. For example, if an antenna is broken but will not be used for the mission, the aircraft is still available for this mission but the data can be used to calculate the A of the system for another mission that would require the antenna. Reporting for A is done by exception. That is, if the maintenance and aircrew preflight checks are passed, the tester knows that many items have been checked and passed. If he knows that the aircraft was checked N times, then all items checked without any failures have passed N checks; those with one failure passed N-1 times, etc. If more than one discrepancy leading to an abort is discovered during the checks for a single sortie, then the system is counted as aborting once. However, each item checked that could have aborted the mission is counted as having a discrepancy.

The probability of an operationally ready load passing the preflight check of the load is covered by the data elements shown in Figure 6-23. Once the load passes the preflight check, it is Available. For the subsystems and nonexpendable loads that refer to Communications/Electronics (CE) Availability, the tester selects data elements from CE that apply to the system under test. For example, if the Aircraft System under test were AWACS, there would be extensive use of the CE data elements. No attempt should be made to separate the CE Subsystem from the Aircraft Subsystem as though they were independent systems. Even by the definition of a system they can, in no way, be considered two systems.

Once the aircraft has passed all of the preflight checks shown in Figures 6-19, 6-21, 6-22 and 6-23 the Aircraft System is available. It is recognized that the system may not be in perfect condition. It is in one of its states of operation where it is considered available for the mission. Over a number of trials, its average condition will be what may be expected, on the average, for its next mission. From an operational viewpoint, it really is of no interest what the system effectiveness happens to be for a perfectly operating

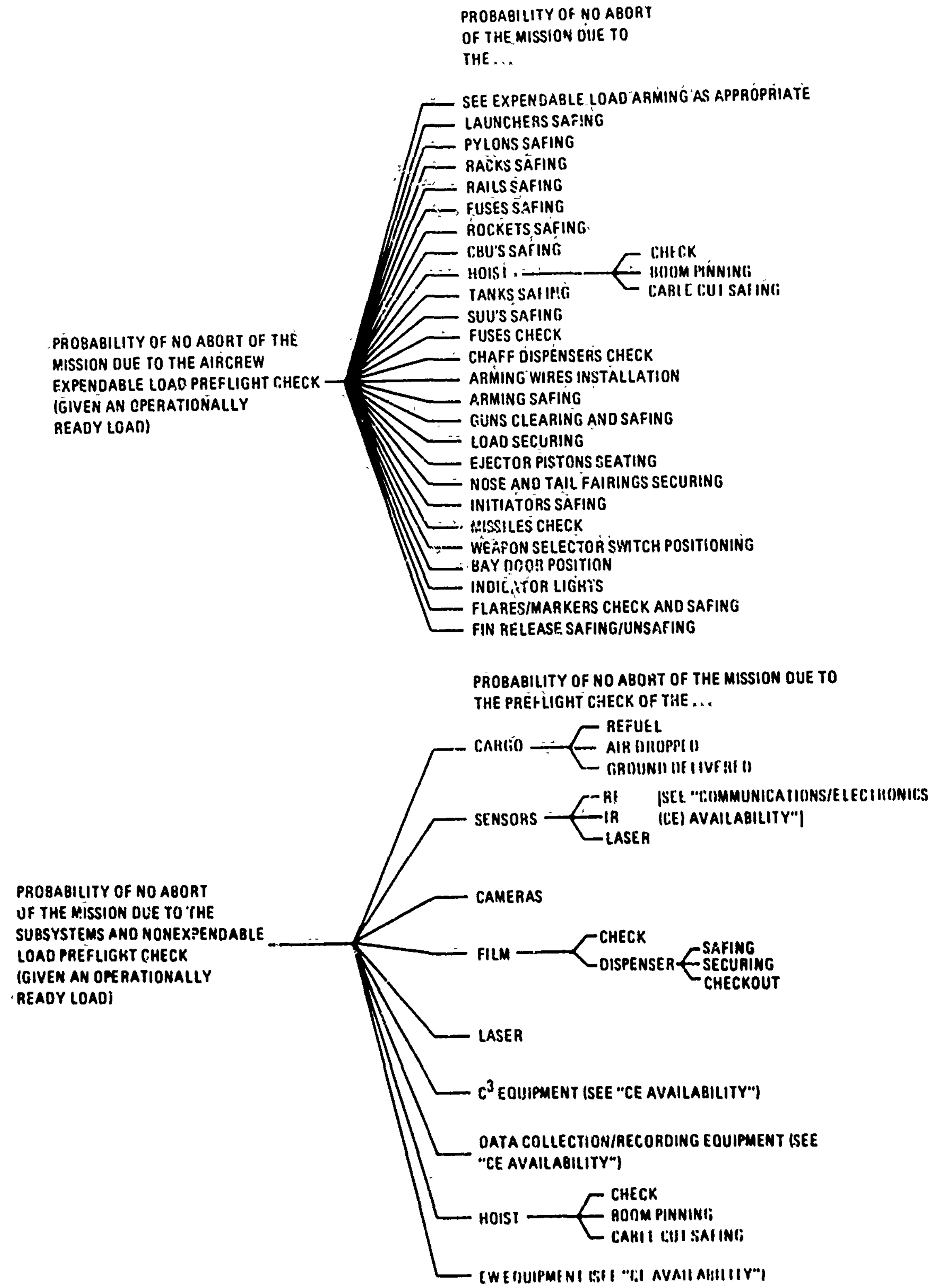


Figure 6-23. Probability of no abort due to preflight check and non-expendable load preflight check

system. This is true since, in a real operational situation, the probability of having a perfectly operating system approaches zero. What is operationally important is, "What is the system effectiveness that can be expected on the next mission?"

Given the Aircraft System's Availability, as derived above, the Aircraft System's Dependability can be addressed. The Dependability data elements are taken during different portions of the sortie as shown in Figure 6-24. Again, these portions are broken into sortie phases in the same manner as a T.O. checklist for the aircrew.

Engine start is used to describe all functions executed from the time the aircrew is in the aircraft and finished the checks covered by Availability until the Aircraft System starts taxing. During this time frame, checks are made of many of the aircraft subsystems as shown in Figure 6-25. When available, the T.O.'s are a good source of help in selecting the appropriate data elements from Figure 6-25. The terms used in the figures generally refer to functional uses rather than equipment. Any discrepancies that do not result in an abort of the missions, but result in a maintenance action, do not affect the Dependability of the system. It results in an effect on Availability and the evaluation of the system by Logistics Command. As with the Availability, this leads to operationally realistic system Dependability. It also leads to operationally

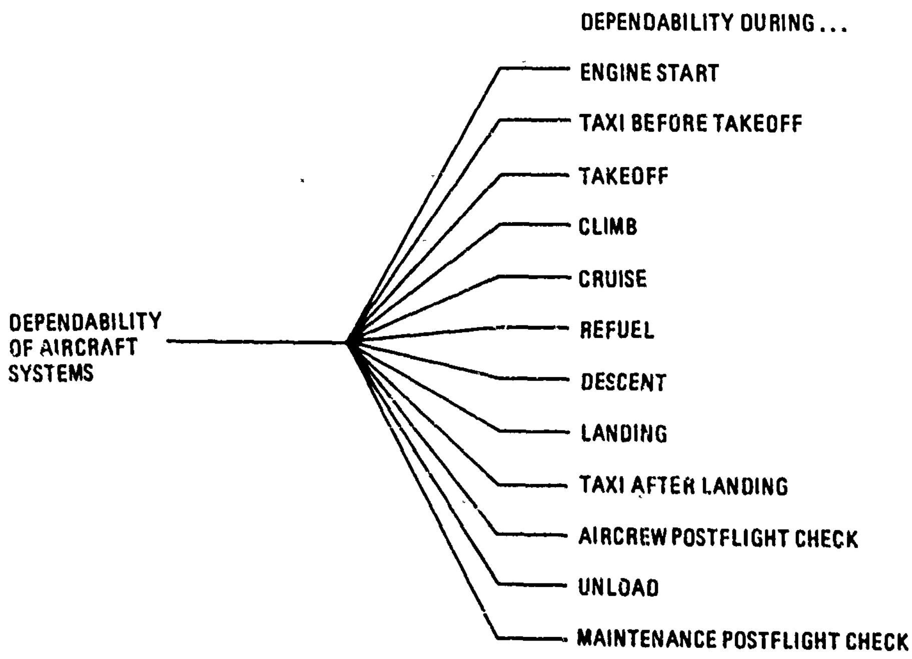


Figure 6-24. Dependability of Aircraft Systems

realistic Capability evaluations. The data element "Δ Pressure" means the pressure differential as measured in jet engines. The rest are self explanatory.

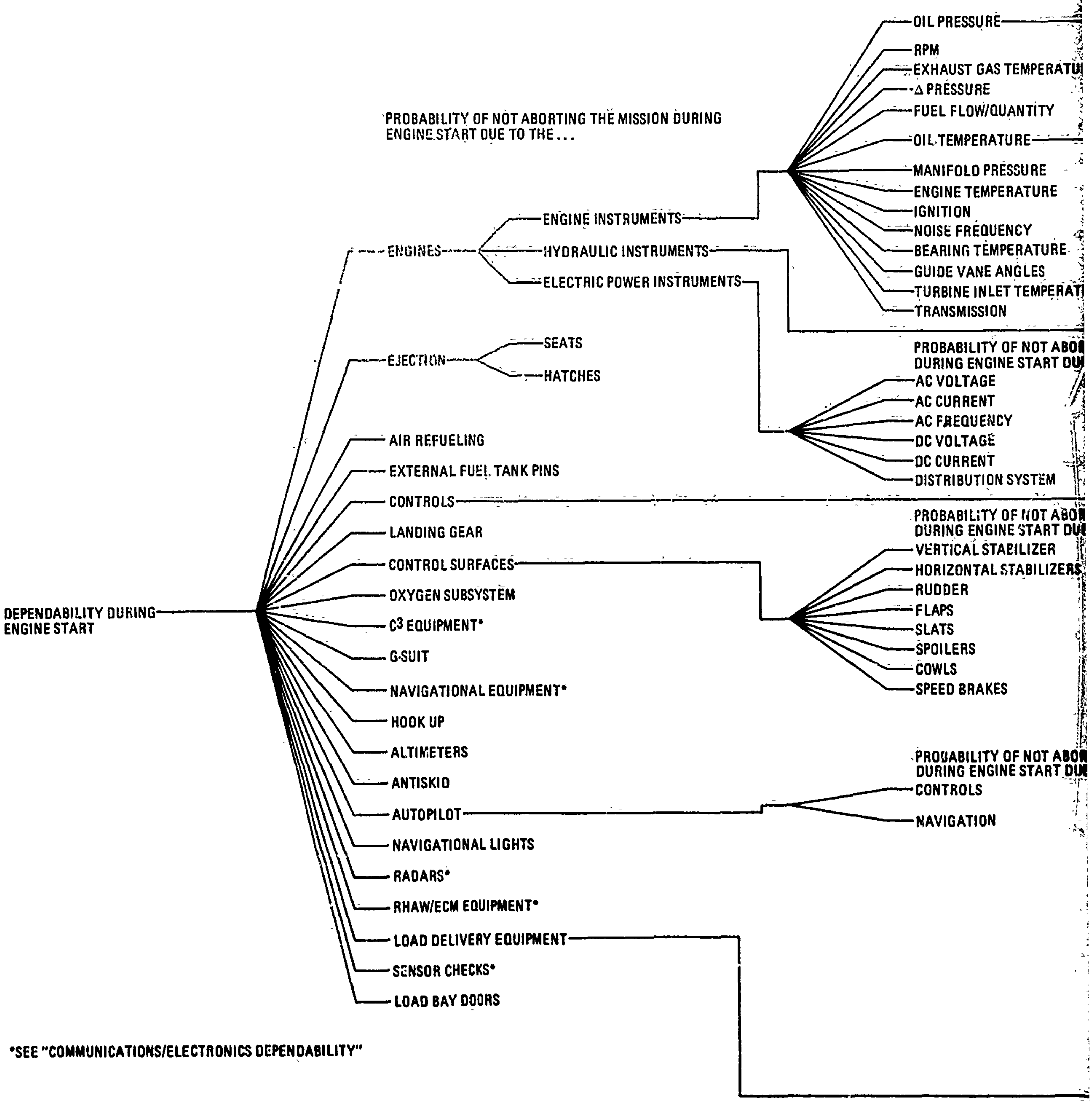
Given that the system passes the start engine checks of Figure 6-25, it is then ready for taxi before takeoff. Included in the taxi before takeoff actions are the arming of the expendable load and such other checks done during this time frame. These arming actions are accomplished in the designated arming area. The data elements leading to Dependability during taxi before takeoff are covered in Figure 6-26. As stated before, only those data elements that normally would be collected at this point in the mission will be selected by the tester.

The data elements that apply during takeoff are covered in Figure 6-27.

After takeoff, the next time frame is that taken up by climbing to altitude. The data elements covering this portion of the sortie are shown in Figure 6-28.

In most cases, the major portion of time during the sortie is spent in the cruise mode. Also, during this time frame many of the load delivery functions are checked. These facts, plus several others, lead to many checks being performed during cruise. Some of these checks may be performed more than once. The tester should select his appropriate data elements and state when they should be checked. Again, when T.O.'s exist, they can be used for guidance. The data elements for Dependability during cruise are shown in Figure 6-29. Here, many of the data elements can be further understood by refering to CE System's Dependability. The scenario should state the mode of operation of radars (normal or countermeasures mode). The data elements shown in Figure 6-29 are far more in number than would ever be required by any Aircraft System. They apply to all types of Aircraft Systems for all types of scenarios. The most elaborate Aircraft System would hardly use more than half of the data elements for a total system test. Depending upon the mission, the Capability may be executed during the Dependability time frame (e.g., Air-to-Air Engagement) or at the end of the Dependability time frame (e.g., Airlift of Passengers). Irrespective of when the Capability of the system occurs, it can be addressed separately since it addresses different aspects of the system than does the Dependability.

Probability of
during engine s



*SEE "COMMUNICATIONS/ELECTRONICS DEPENDABILITY"

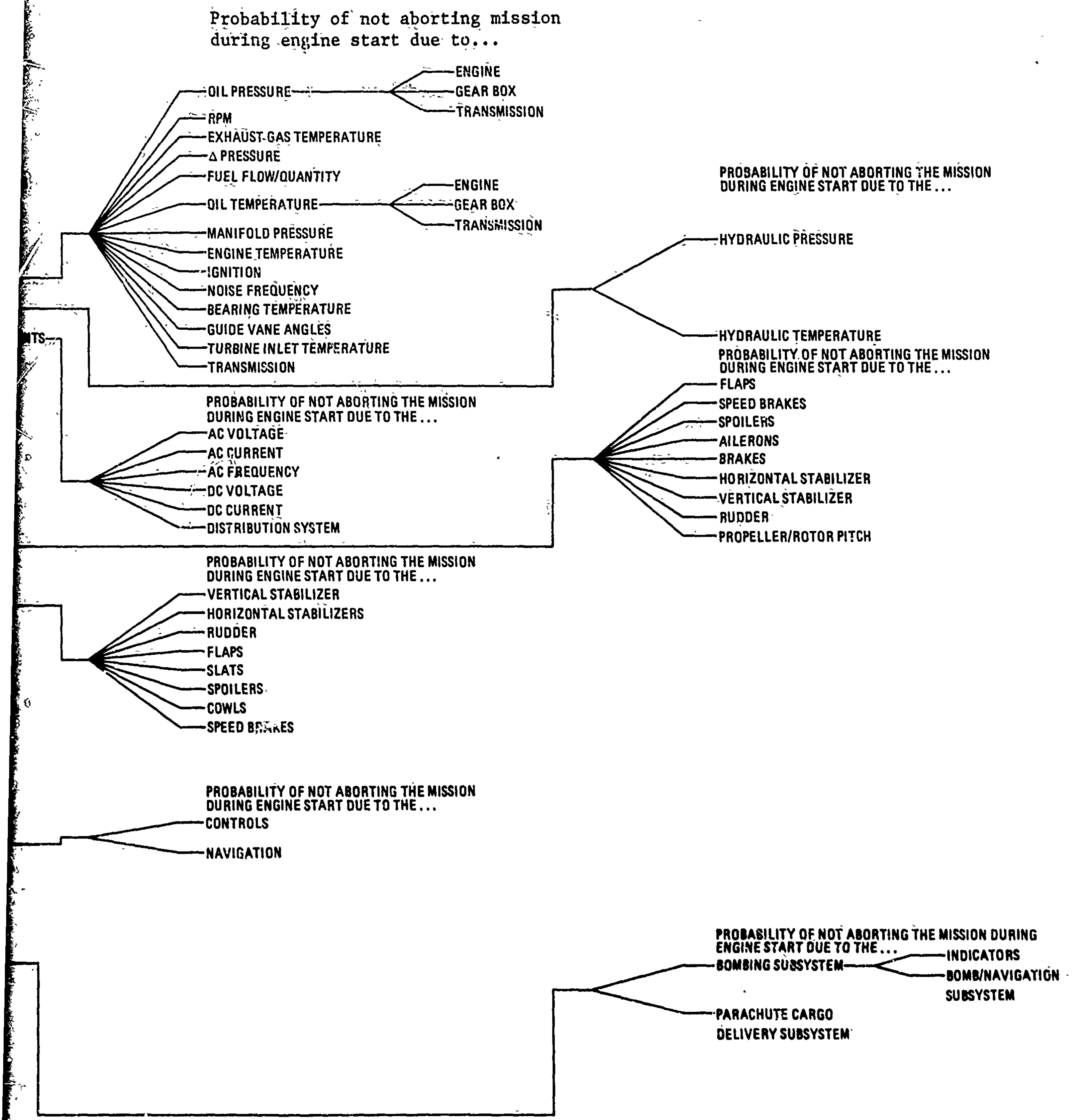
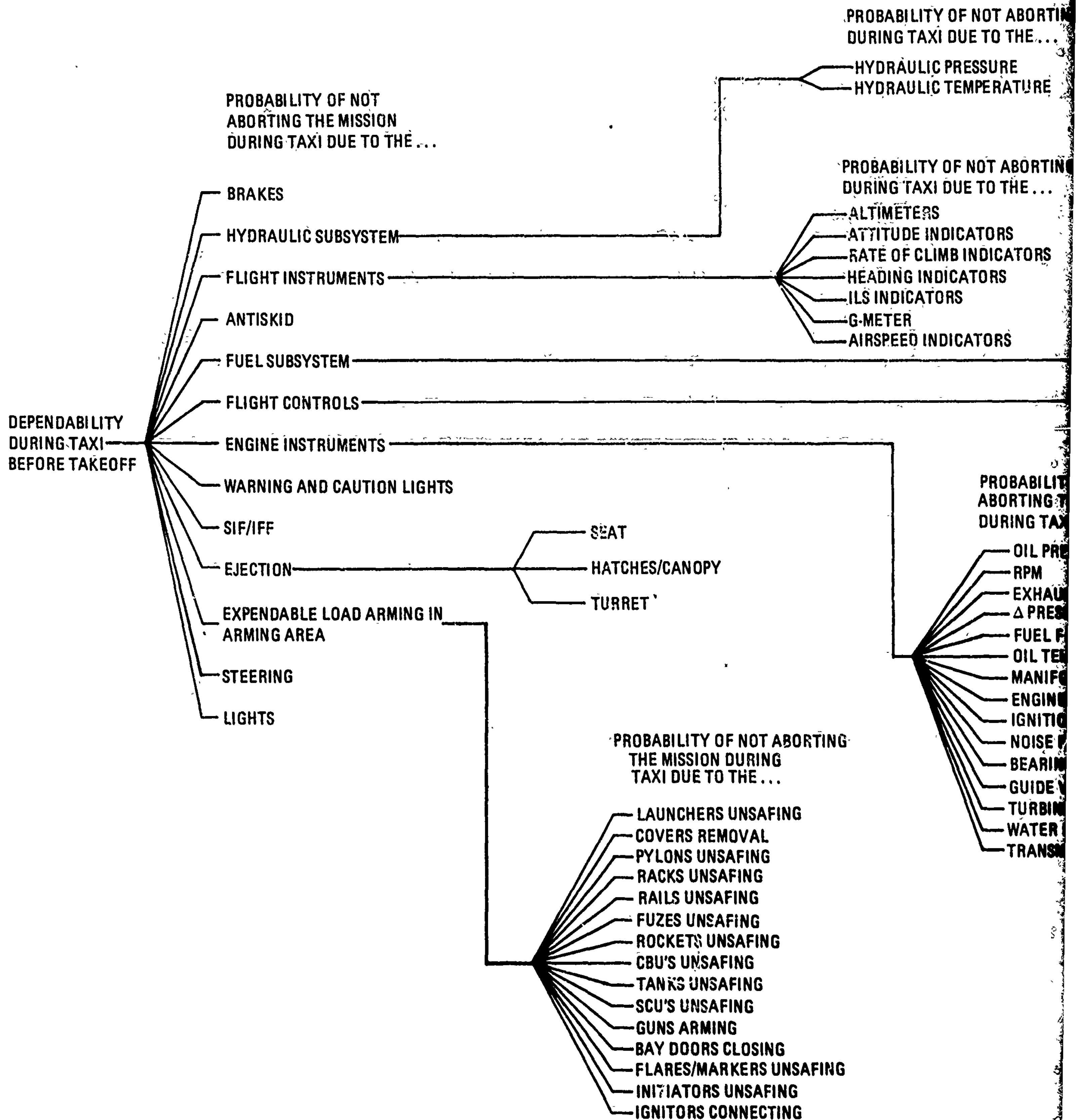


Figure 6-25. Dependability during engine start


DEPENDABILITY DURING TAXI BEFORE TAKEOFF
PROBABILITY OF NOT ABORTING THE MISSION DURING TAXI DUE TO THE . . .
BRAKES
HYDRAULIC SUBSYSTEM
FLIGHT INSTRUMENTS
ANTISKID
FUEL SUBSYSTEM
FLIGHT CONTROLS
ENGINE INSTRUMENTS
WARNING AND CAUTION LIGHTS
SIF/IFF
EJECTION
EXPENDABLE LOAD ARMING IN ARMING AREA
STEERING
LIGHTS
PROBABILITY OF NOT ABORTIN
DURING TAXI DUE TO THE . . .
HYDRAULIC PRESSURE
HYDRAULIC TEMPERATURE
PROBABILITY OF NOT ABORTIN
DURING TAXI DUE TO THE . . .
ALTIMETERS
ATTITUDE INDICATORS
RATE OF CLIMB INDICATORS
HEADING INDICATORS
ILS INDICATORS
G-METER
AIRSPEED INDICATORS
SEAT
HATCHES/CANOPY
TURRET
PROBABILIT
ABORTING T
DURING TA
OIL PR
RPM
EXHAU
Δ PRES
FUEL F
OIL TE
MANIF
ENGIN
IGNITI
NOISE
BEARI
GUIDE
TURBI
WATER
TRANS
PROBABILITY OF NOT ABORTING THE MISSION DURING TAXI DUE TO THE . . .
LAUNCHERS UNSAFING
COVERS REMOVAL
PYLONS UNSAFING
RACKS UNSAFING
RAILS UNSAFING
FUZES UNSAFING
ROCKETS UNSAFING
CBU'S UNSAFING
TANKS UNSAFING
SCU'S UNSAFING
GUNS ARMING
BAY DOORS CLOSING
FLARES/MARKERS UNSAFING
INITIATORS UNSAFING
IGNITORS CONNECTING

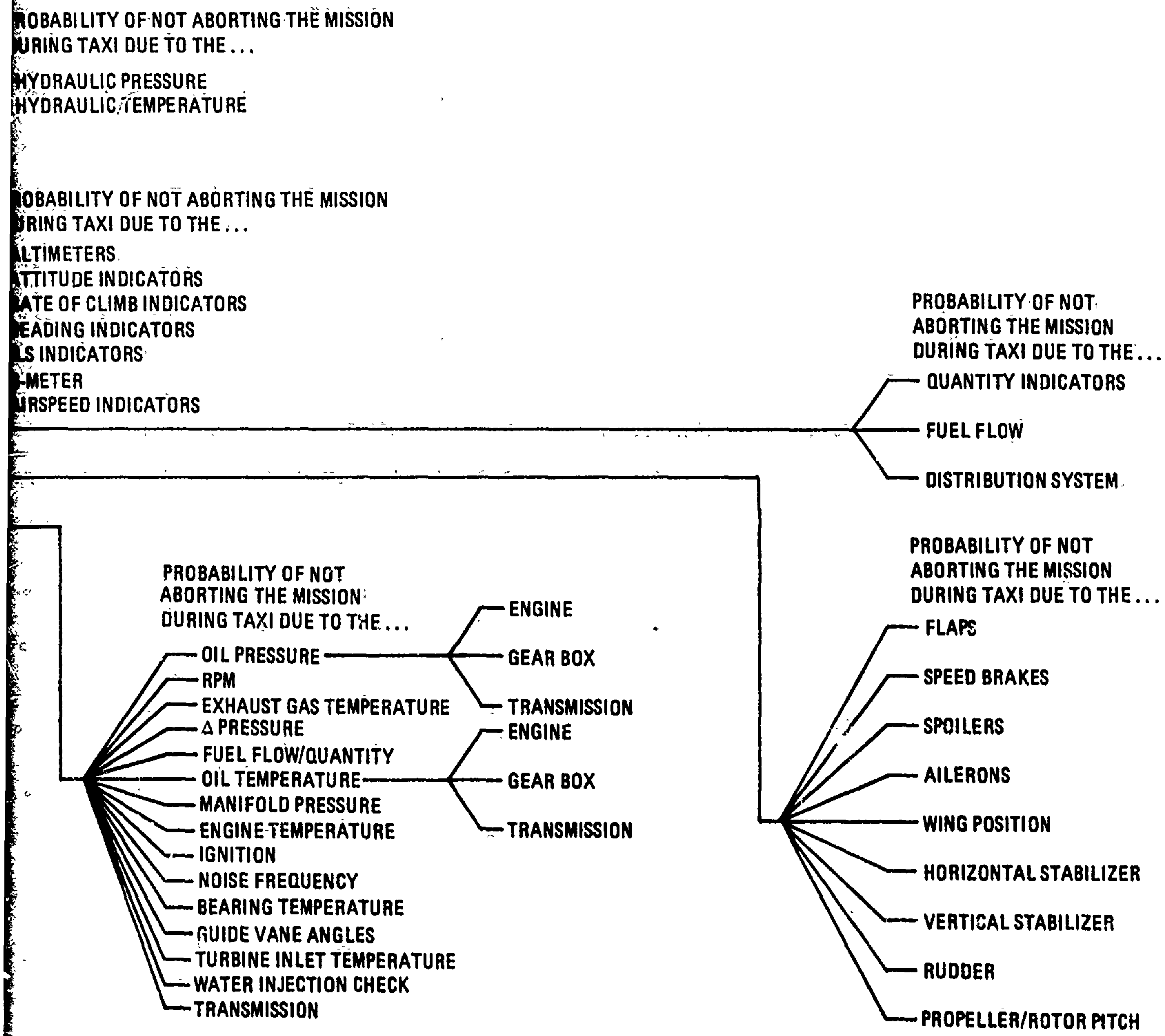


Figure 6-26. Dependability during taxi

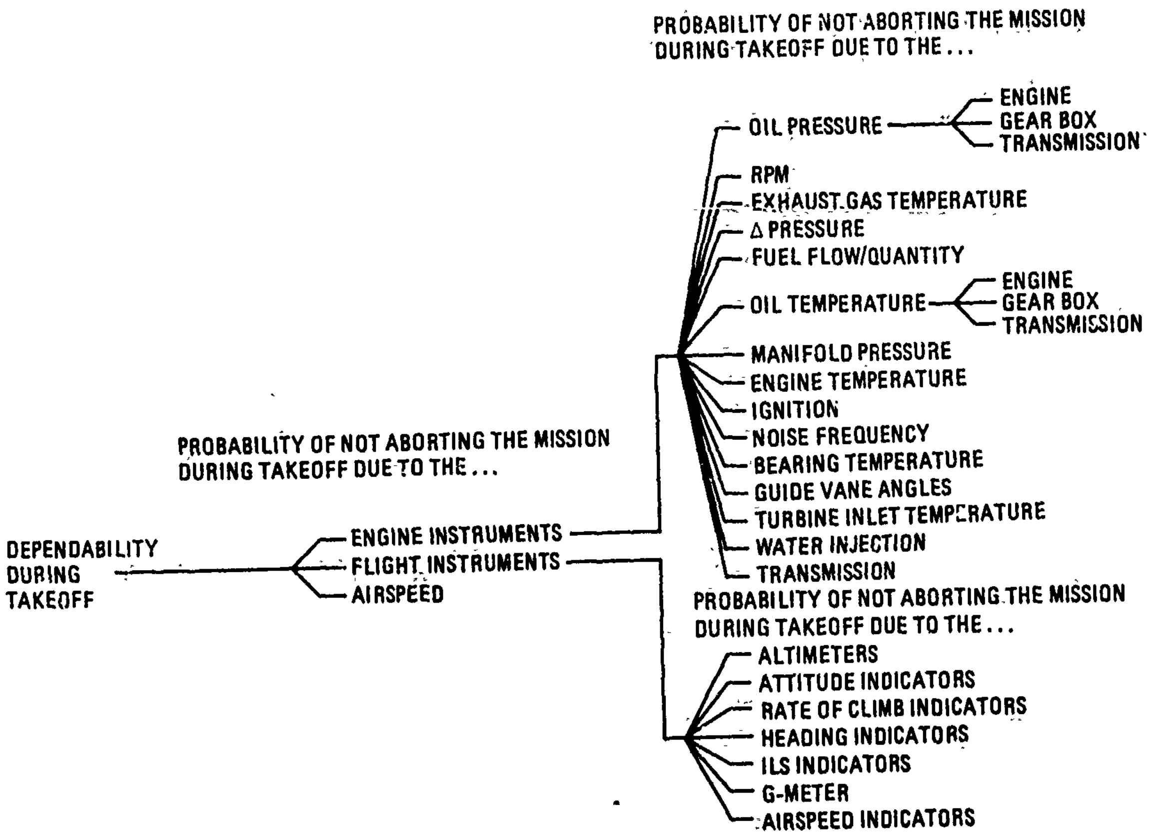


Figure 6-27. Dependability during takeoff

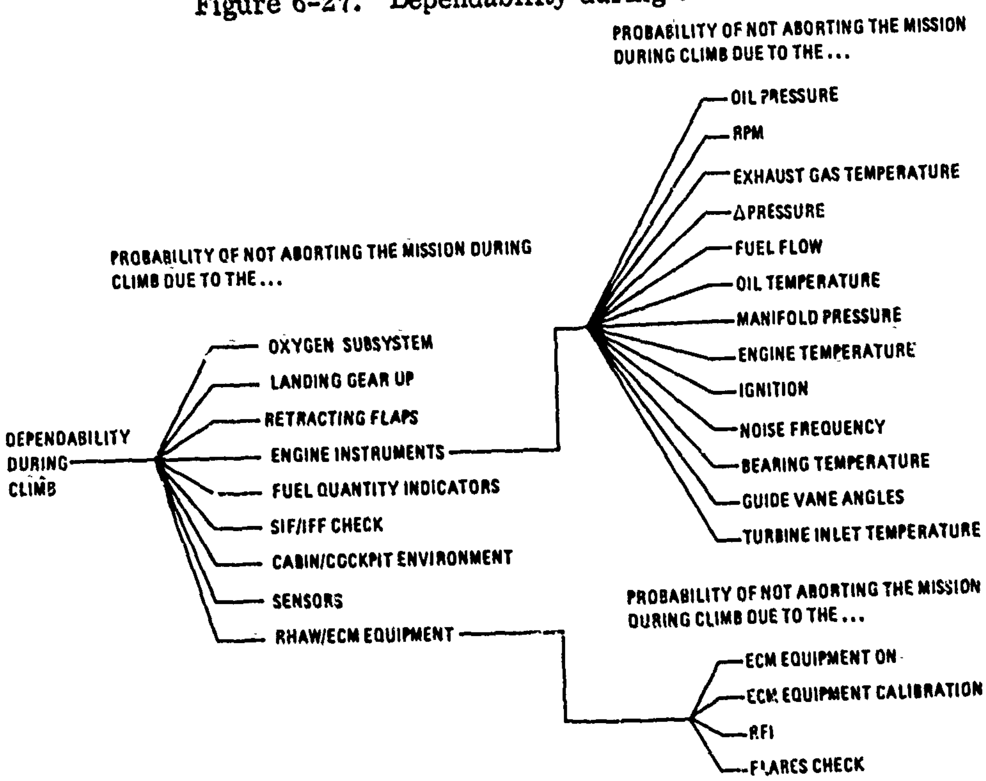
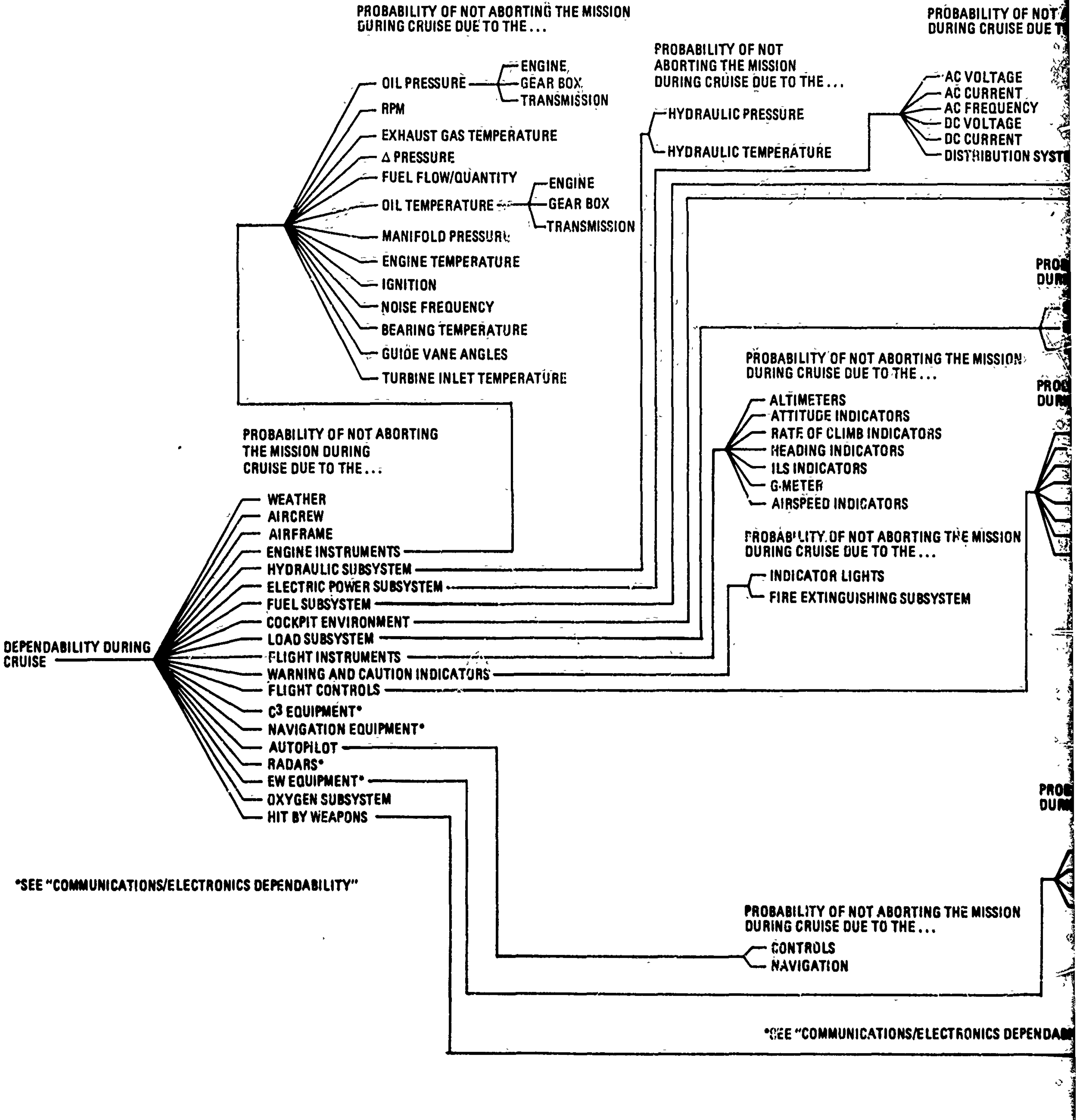


Figure 6-28. Dependability during climb



DEPENDABILITY DURING CRUISE
PROBABILITY OF NOT ABORTING THE MISSION DURING CRUISE DUE TO THE . . .
WEATHER
AIRCREW
AIRFRAME
ENGINE INSTRUMENTS
HYDRAULIC SUBSYSTEM
ELECTRIC POWER SUBSYSTEM
FUEL SUBSYSTEM
COCKPIT ENVIRONMENT
LOAD SUBSYSTEM
FLIGHT INSTRUMENTS
WARNING AND CAUTION INDICATORS
FLIGHT CONTROLS
C3 EQUIPMENT*
NAVIGATION EQUIPMENT*
AUTOPILOT
RADARS*
EW EQUIPMENT*
OXYGEN SUBSYSTEM
HIT BY WEAPONS
*SEE "COMMUNICATIONS/ELECTRONICS DEPENDABILITY"
PROBABILITY OF NOT ABORTING THE MISSION DURING CRUISE DUE TO THE . . .
OIL PRESSURE
ENGINE
GEAR BOX
TRANSMISSION
RPM
EXHAUST GAS TEMPERATURE
Δ PRESSURE
FUEL FLOW/QUANTITY
OIL TEMPERATURE
ENGINE
GEAR BOX
TRANSMISSION
MANIFOLD PRESSURE
ENGINE TEMPERATURE
IGNITION
NOISE FREQUENCY
BEARING TEMPERATURE
GUIDE VANE ANGLES
TURBINE INLET TEMPERATURE
PROBABILITY OF NOT ABORTING THE MISSION DURING CRUISE DUE TO THE . . .
HYDRAULIC PRESSURE
HYDRAULIC TEMPERATURE
PROBABILITY OF NOT
DURING CRUISE DUE T
AC VOLTAGE
AC CURRENT
AC FREQUENCY
DC VOLTAGE
DC CURRENT
DISTRIBUTION SYST
PROBABILITY OF NOT ABORTING THE MISSION DURING CRUISE DUE TO THE . . .
ALTIMETERS
ATTITUDE INDICATORS
RATE OF CLIMB INDICATORS
HEADING INDICATORS
ILS INDICATORS
G-METER
AIRSPEED INDICATORS
PROBABILITY OF NOT ABORTING THE MISSION DURING CRUISE DUE TO THE . . .
INDICATOR LIGHTS
FIRE EXTINGUISHING SUBSYSTEM
PROBABILITY OF NOT ABORTING THE MISSION DURING CRUISE DUE TO THE . . .
CONTROLS
NAVIGATION
*SEE "COMMUNICATIONS/ELECTRONICS DEPENDA

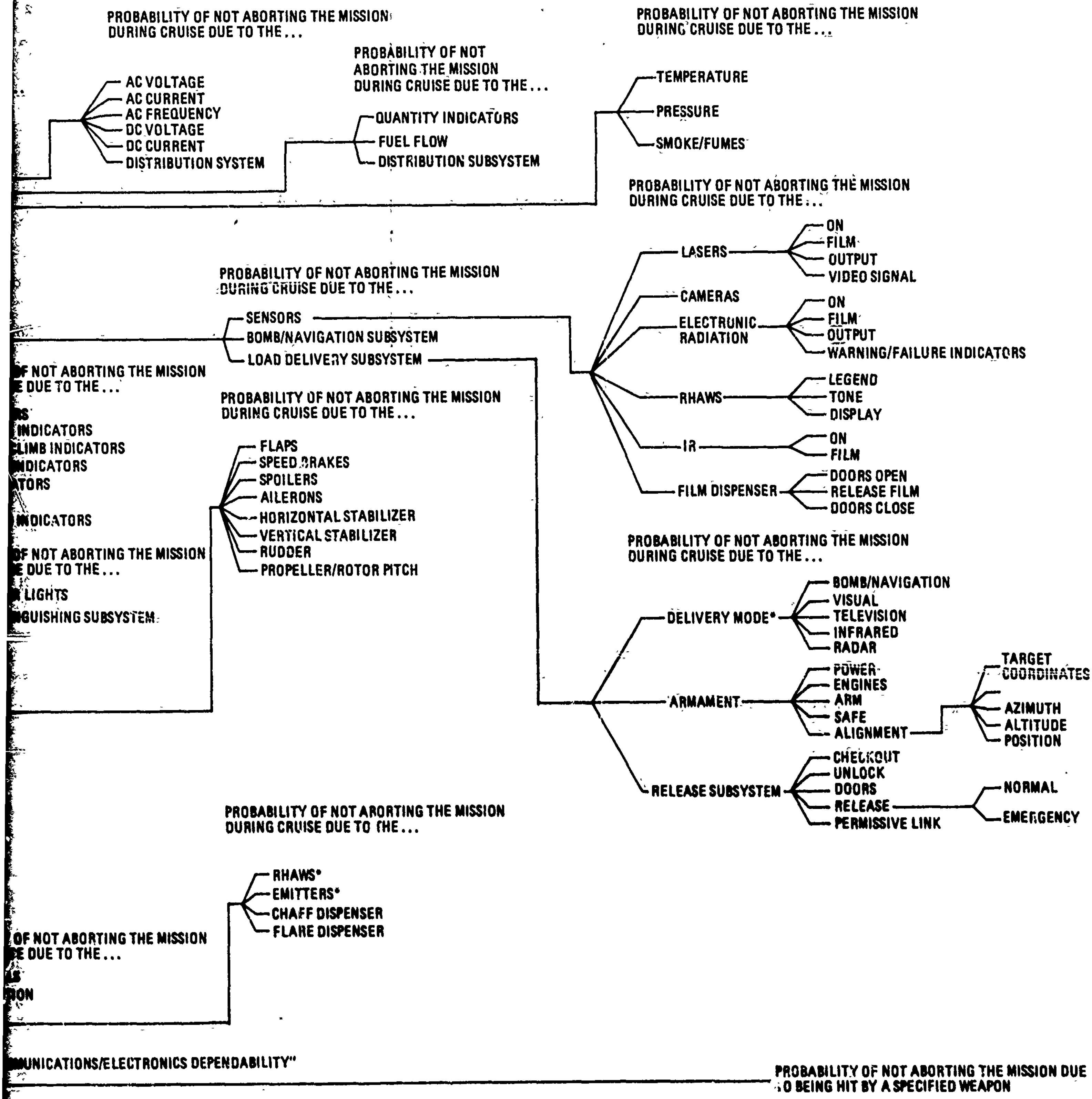


Figure 6-29. Dependability during cruise

If the scenario calls for air to air refueling during the sortie, then the data elements shown in Figure 6-30 are used. These data elements address the Dependability of the aircraft being refueled. If the Aircraft System being tested has a mission of refueling other aircraft, then the Capability for Refueling is to be addressed. This Capability will be addressed later with other Aircraft Systems' Capabilities. The meaning of the data elements in Figure 6-30 are self evident.

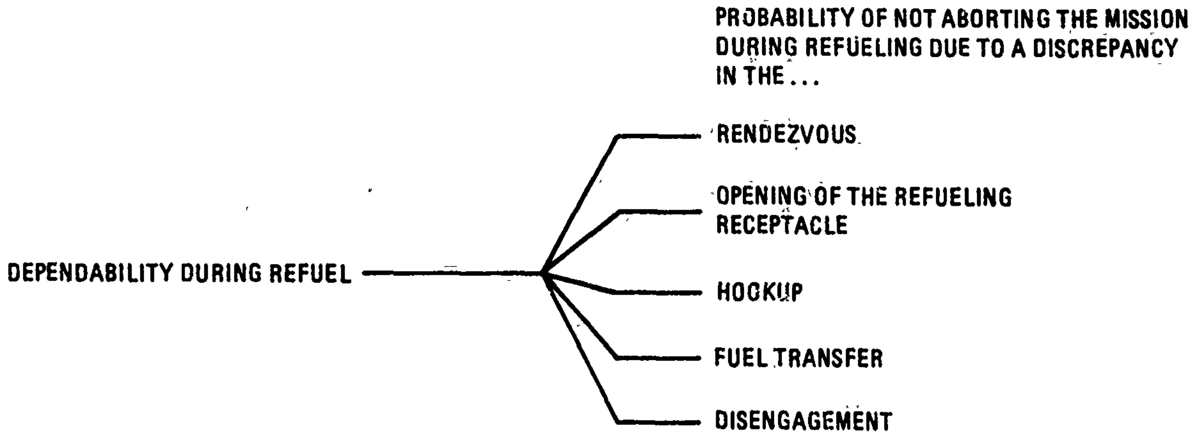


Figure 6-30. Dependability during refuel

At the end of the cruise portion of the sortie, the descent portion starts. Malfunctions occurring at this point seldom lead to an abort of the mission. They do affect the Availability for the next mission and contribute to calculations of the Dependability for other portions of the sortie. That is, they lead to calculations of mean time between failures (MTBF) which affects the Dependability during other portions of the sortie. The data elements for Dependability during descent are shown in Figure 6-31.

After descent, the next sortie portion is landing. Dependability during landing is addressed by the data elements shown in Figure 6-32.

Taxi after landing data elements are shown in Figure 6-33. These data elements give the Dependability during this time frame.

If the Aircraft System has expendable loads that require safing, the Dependability of the expendable load safing in the designated arming area is addressed by the data elements shown in Figure 6-34. If the Aircraft System has no expendable load requiring safing, then Figure 6-34 is ignored.

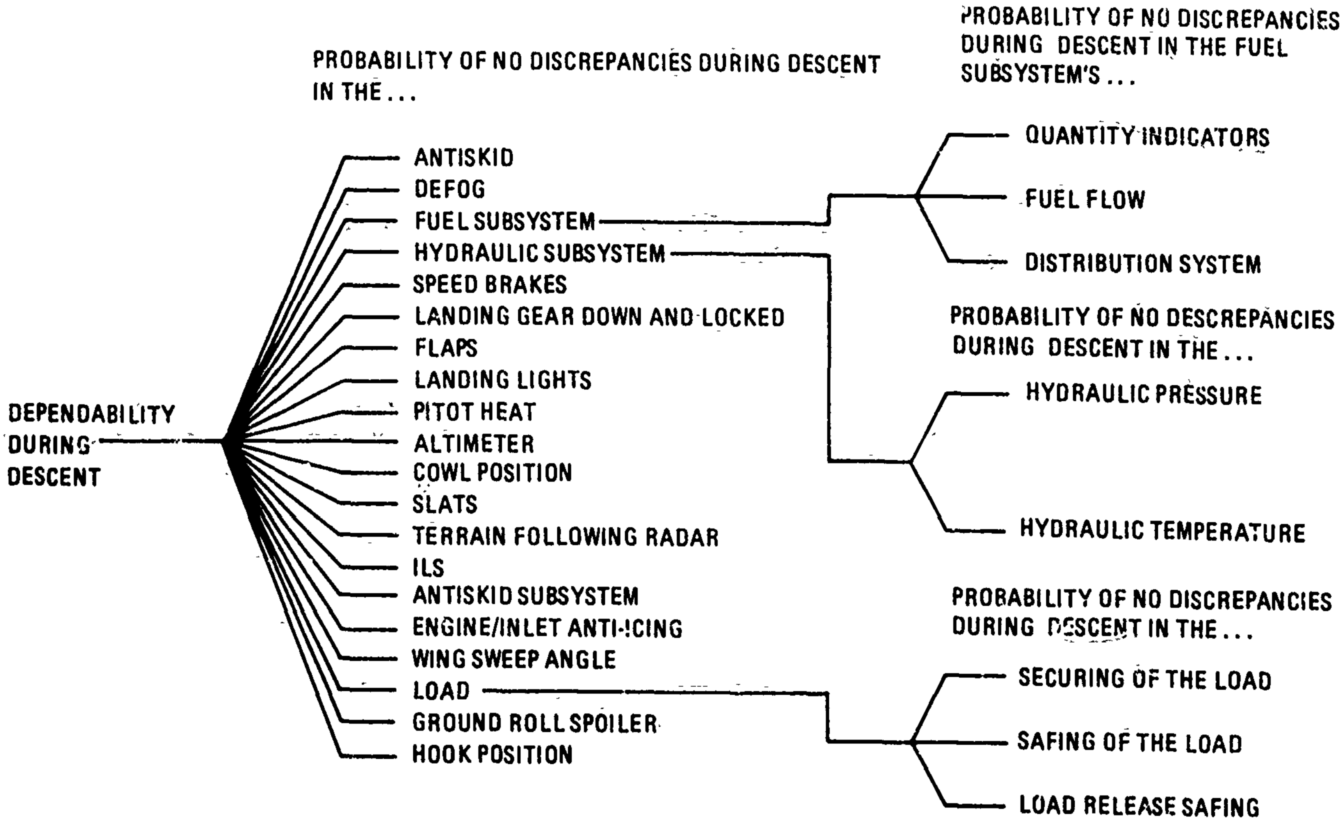


Figure 6-31. Dependability during descent

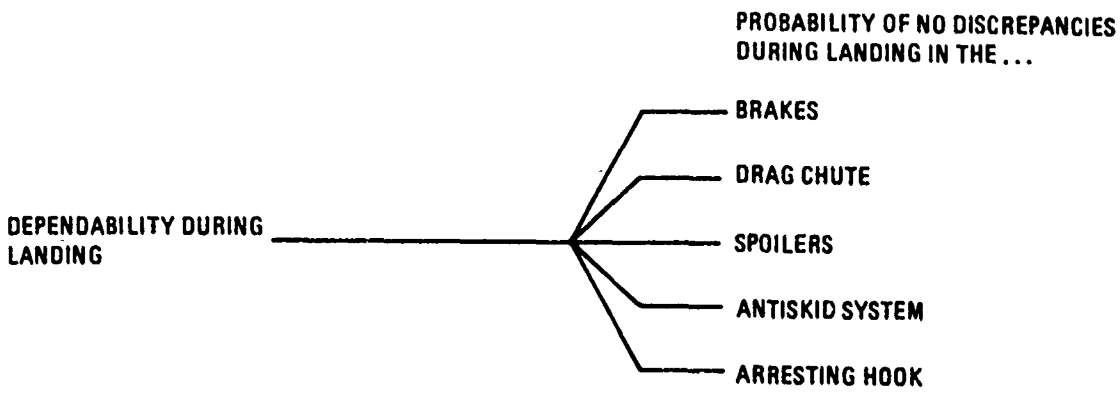


Figure 6-32. Dependability during landing

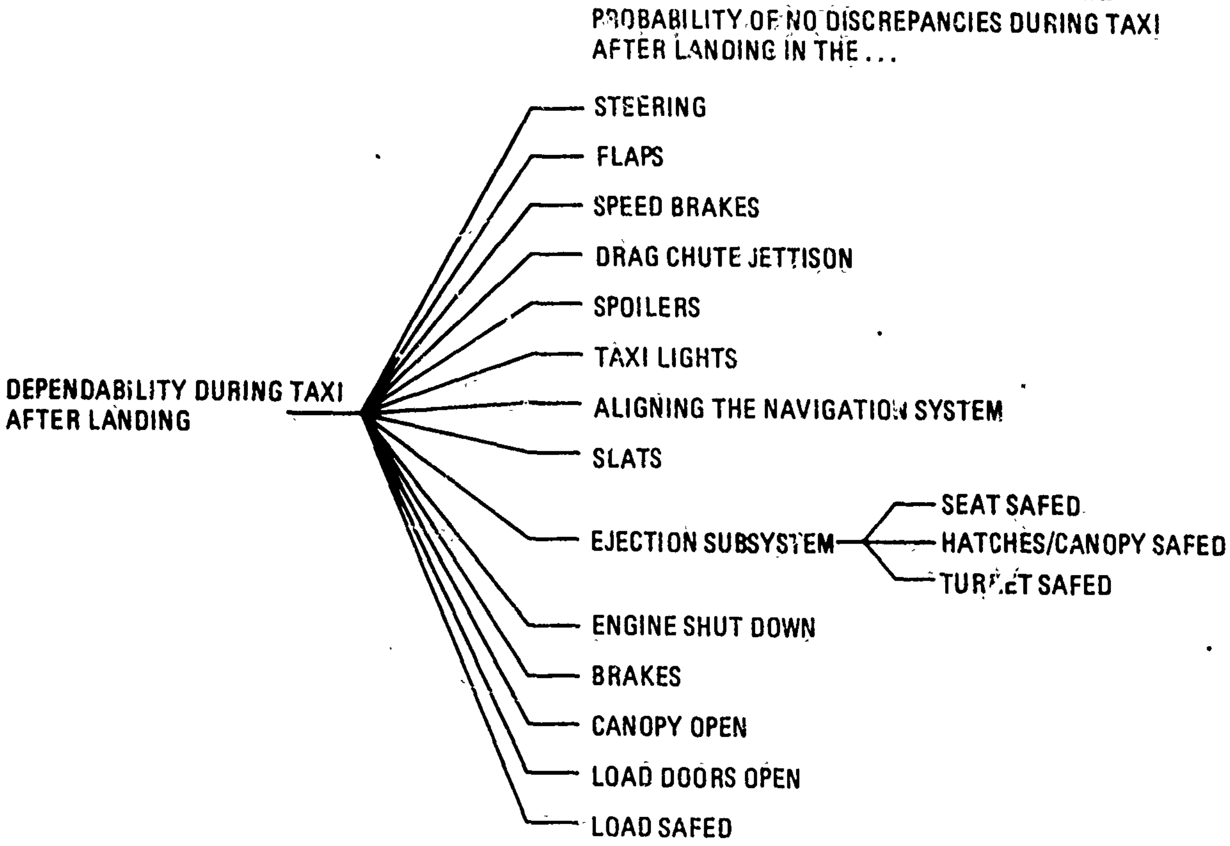


Figure 6-33. Dependability during taxi after landing

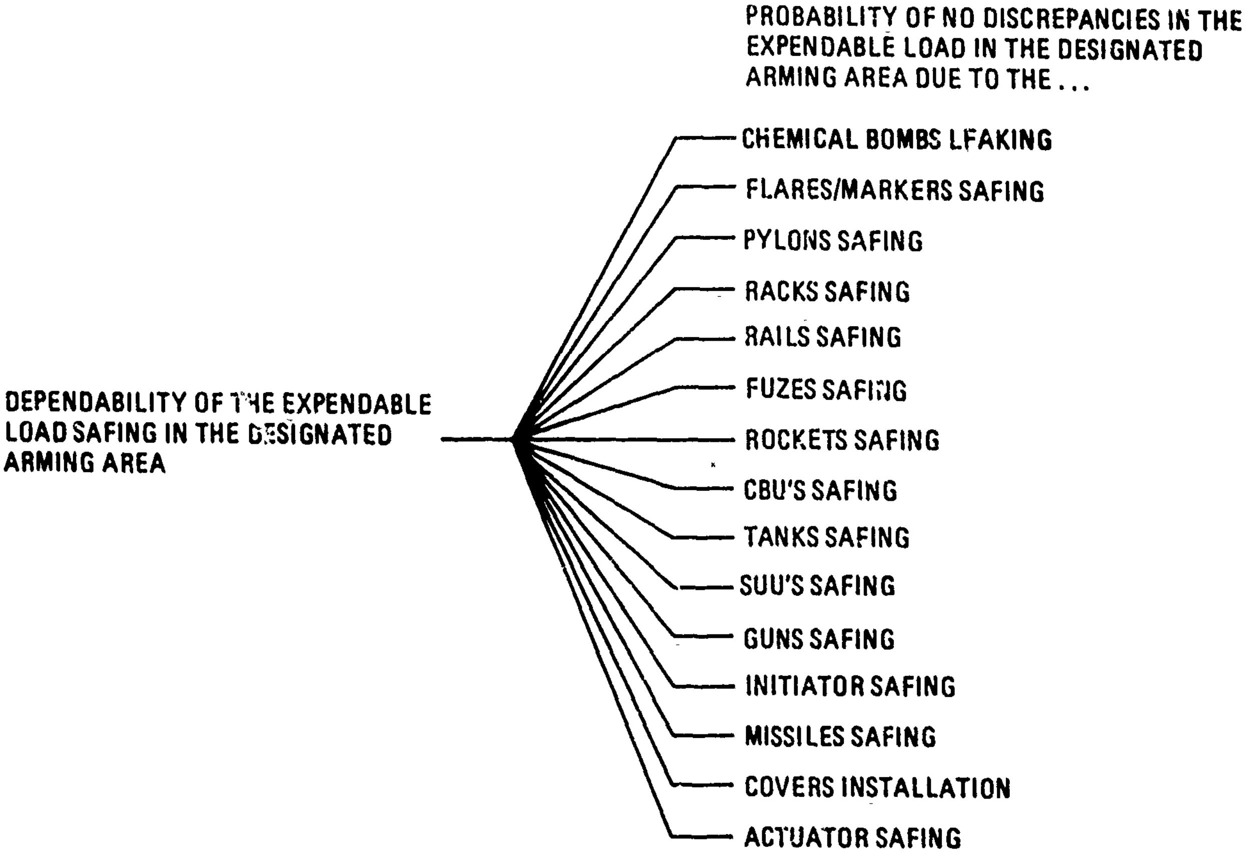


Figure 6-34. Dependability of expendable load safing in designated arming area

The Dependability during the aircrew postflight check is addressed by the data elements shown in Figure 6-35.

The Dependability during the postflight maintenance check is obtained from the data elements shown in Figure 6-36. At this point in time, the Aircraft System's Availability is addressed again.

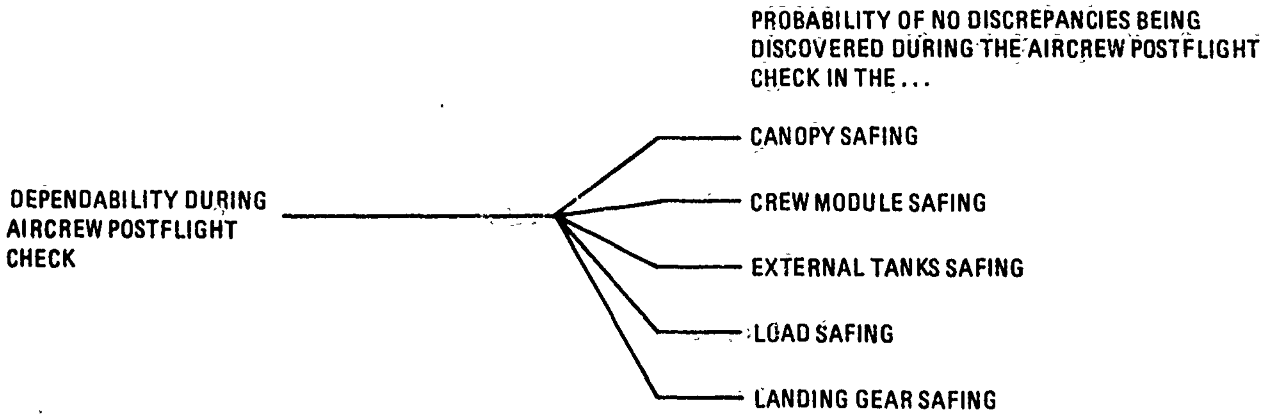


Figure 6-35. Dependability during aircrew postflight check

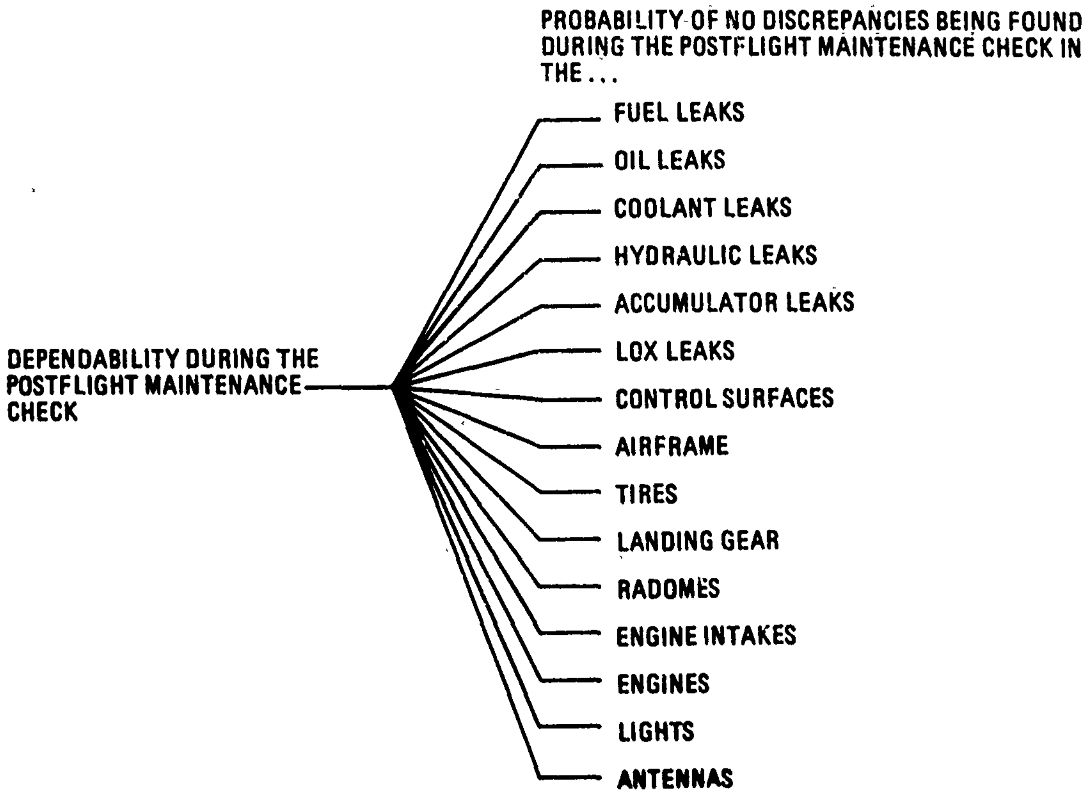


Figure 6-36. Dependability during postflight maintenance check

The product of all of the above Dependabilities lead to the Dependability of the Aircraft System for the particular scenario. Other scenarios can lead to different Dependabilities. This is obvious if, for example, one considers that one scenario may require refueling and another may not. There are other less obvious ways in which the scenario can affect the Dependability of the system. For example, just as sortie rate affects Availability, it can also affect Dependability by overworking the system (which includes people and equipment).

Having now determined the Availability and Dependability of the Aircraft System, the only part of the MOE left is the Capability. The data elements for A and D are fairly mission independent for a given set of items that make up the Aircraft System. Capability, on the other hand, addresses the Aircraft System's specific mission. As shown in Table 6-2, Aircraft Systems have certain missions. A given Aircraft System will have a certain Capability for each of these missions. Figure 6-37 shows how the Capability of an Aircraft System is stated as that system's Capability for a specific aircraft mission.

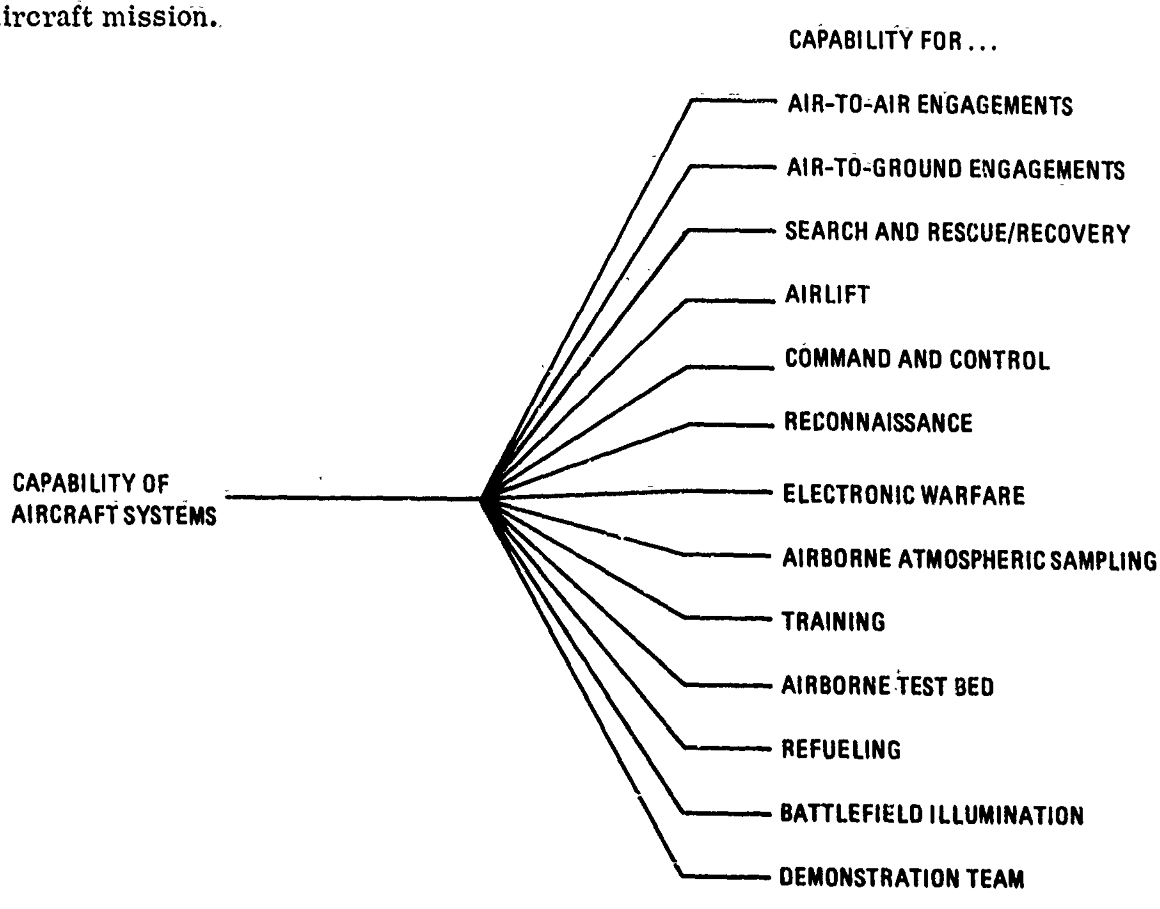


Figure 6-37. Capability of Aircraft Systems

The data elements that make up an Aircraft System's Capability for Air-to-Air Engagements are shown in Figure 6-38. These data elements apply to all types of aircraft and munitions. Therefore, the tester selects only those that apply to his Aircraft System and neglects the others. He should state which data elements he is using. For example, his scenario may state that his test system is assumed to be firing first. If that is the case, he ignores the first data element, the "Probability of Firing First". The data elements that refer to a "specified time" require the tester to specify the value of this time. Sources of information which can guide him in this selection are his past experience, an alternate system's time requirements, or the specifications of the system being tested. Take, as an example, that an alternate system loses lock-on after 1 minute, on the average, and requires 10 seconds, on the average, to reacquire lock-on. Then the tester might want the probability of the new system maintaining lock-on for more than 75 seconds (25% more than the alternate) and the probability

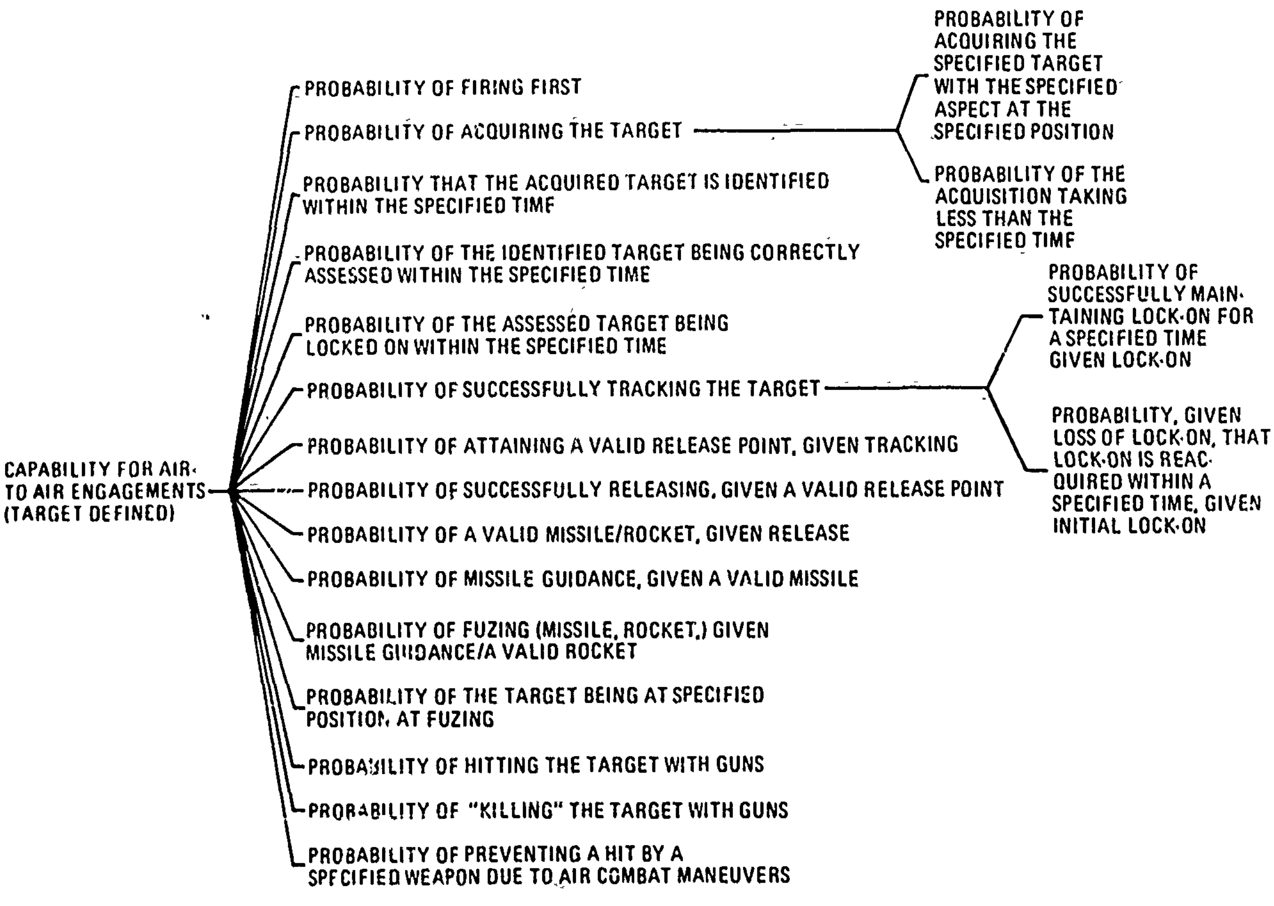


Figure 6-38. Capability for air-to-air engagements

of reacquiring lock-on in 10 seconds (at least as good as the alternate). The data element "the probability of the target being at a specified position at fuzing" needs some elaboration. Position means the location, in three dimensions, of an item (in this case, the target) with the origin of the coordinates taken as the system under test (in this case, the missile or rocket). The reason for measuring and reporting this probability rather than the probability of kill is that the vulnerability of the target may vary with time due to increased information, quality of the estimates, or target modifications. The data can be used for inputs to these evaluations as time goes on and no retesting would be required. If the measurements are used in the current vulnerability estimates and only the results of this evaluation are recorded, then new vulnerability estimates would require a retest of the system to reobtain the data previously obtained. It should also be pointed out that there are several levels of kill, the number depending upon who is making the evaluation and for what type weapon. One of the listings of effects of the target being at a specified position at fuzing is shown in Figure 6-39. If the tester also reports the probability of kill, he should state what is meant by "kill". For this reason, the probability of "killing" the target with guns, as shown in Figure 6-38, has quotation marks about the word, kill. The important data element for guns is, "the probability of hitting the target with guns". Many of the data elements in Figure 6-38 are common to whatever the weapon subsystem is in the Aircraft System. These data elements are those that address such things as target acquisition, identification, etc. The last of

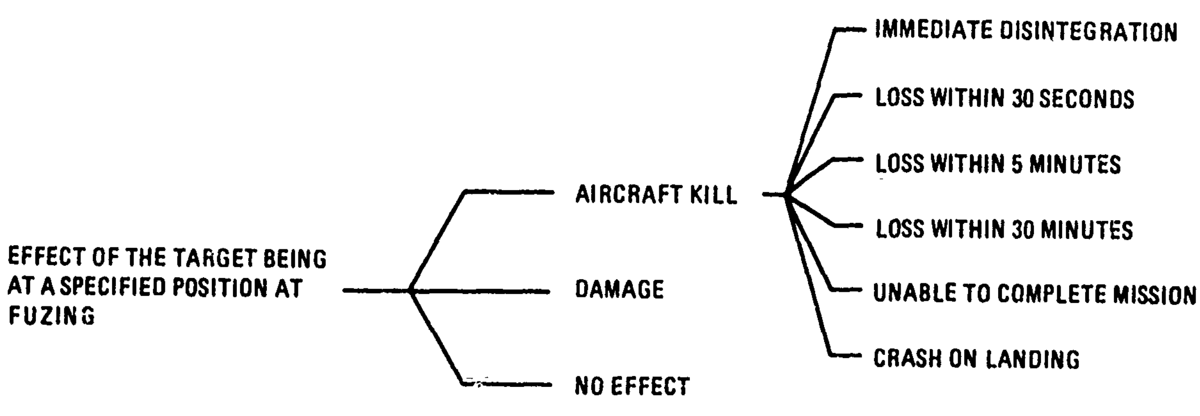


Figure 6-39. Effect of the target being at a specified position at fuzing

the data elements in Figure 6-38 address the capability of the Aircraft System to prevent damage or kill to itself by air combat maneuvers when the system is under attack by a specified weapon. Other forms of defensive capabilities are covered elsewhere, such as EW being covered under Defense Suppression. The product of all of the appropriate data elements in Figure 6-38 give the Capability for Air-to-Air Engagements for the given scenario.

The next Aircraft System mission to be covered is Air-to-Ground Engagements. Typical munitions for this mission are shown in Figure 6-40. The fact that these munitions are so varied does not lead to a significantly expanded list of data elements. These munitions are useful against a number of targets which are shown in Figure 6-41. The specific target(s) for which the Aircraft System is to be evaluated is (are) covered in the scenario. The tester should explicitly state the type of target and as much other detail as is required to clearly explain what Capability is being measured. For example, if the target is a tank, what kind is it, is it camouflaged, is it moving, etc.? These

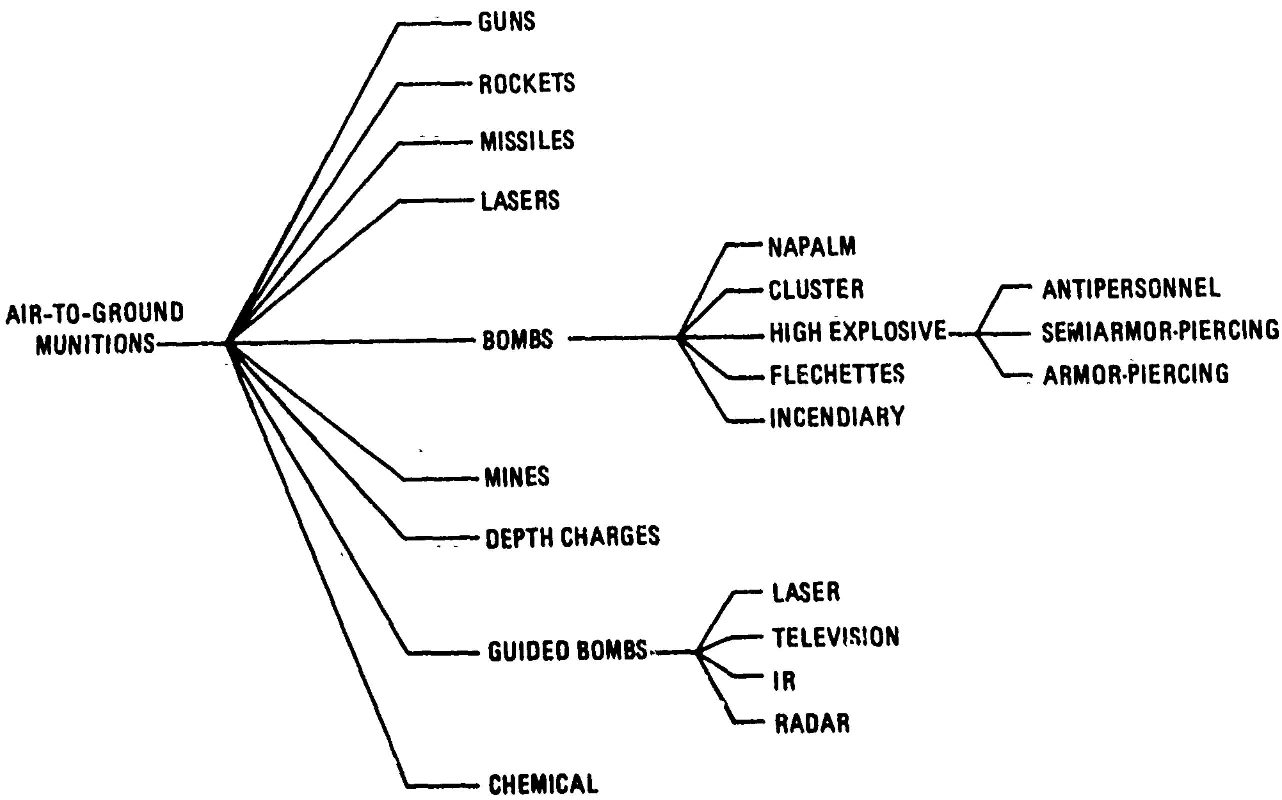


Figure 6-40. Air-to-ground munitions

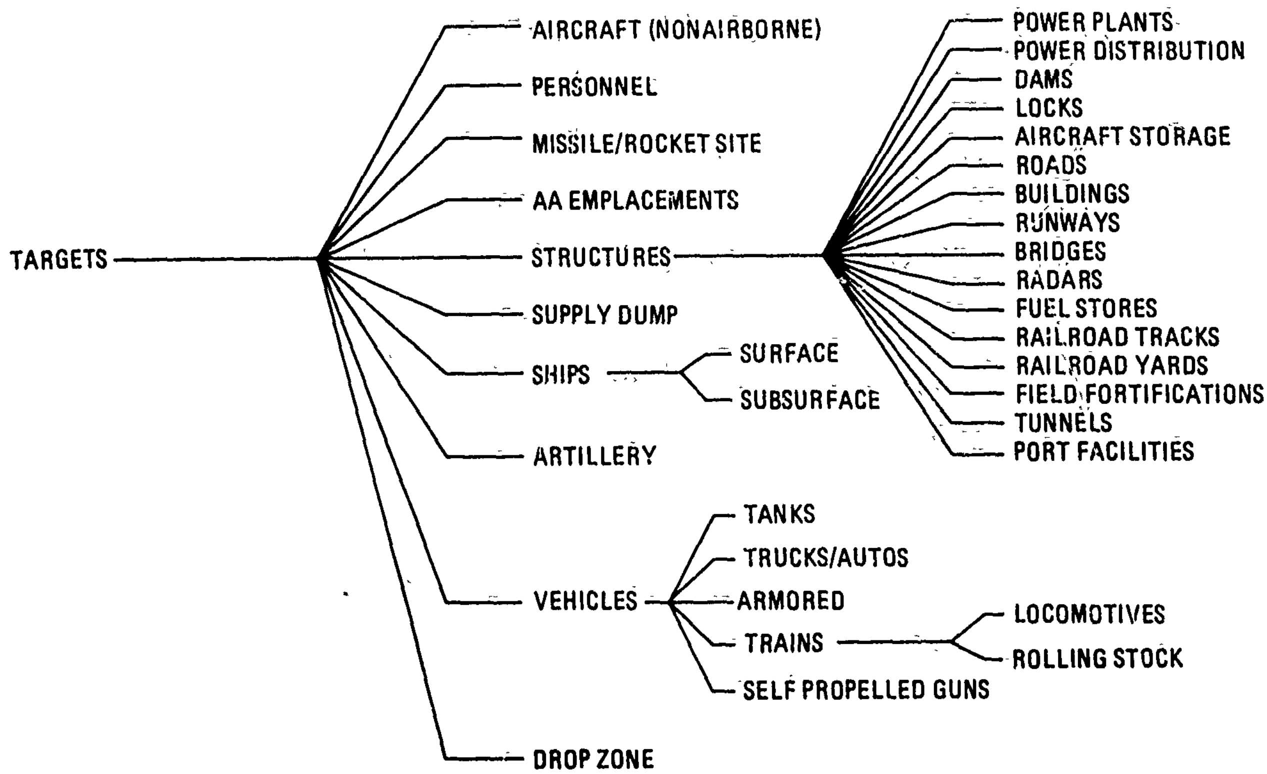


Figure 6-41. Targets for air-to-ground munitions

details should be covered in the scenario and should be restated in the Test Report. The data elements which are used to measure Capability for Air-to-Ground Engagements are shown in Figure 6-42. The similarity with the data elements for Air-to-Air Engagements is obvious. The reason for this is that the actions that are performed, like acquiring the target, are the same. The difference, of course, is how well the system can perform these actions under the different circumstances, targets, etc. As before, the comments about the specific data elements apply (such as the target position at fusing). The effect of the target being at a specified position at fusing can best be determined by refering to the Joint Munitions Effectiveness Manual entitled Weapon Effectiveness, Selection and Requirements (TH61A1-1-1-1) Chapter 3. These effects are shown in Table 6-4 which is taken from the referenced document. It states, for example, that there are three levels of kill for tanks. This fact, and the time dependence of vulnerability assessments of foreign equipment make the reporting of "kill probability"

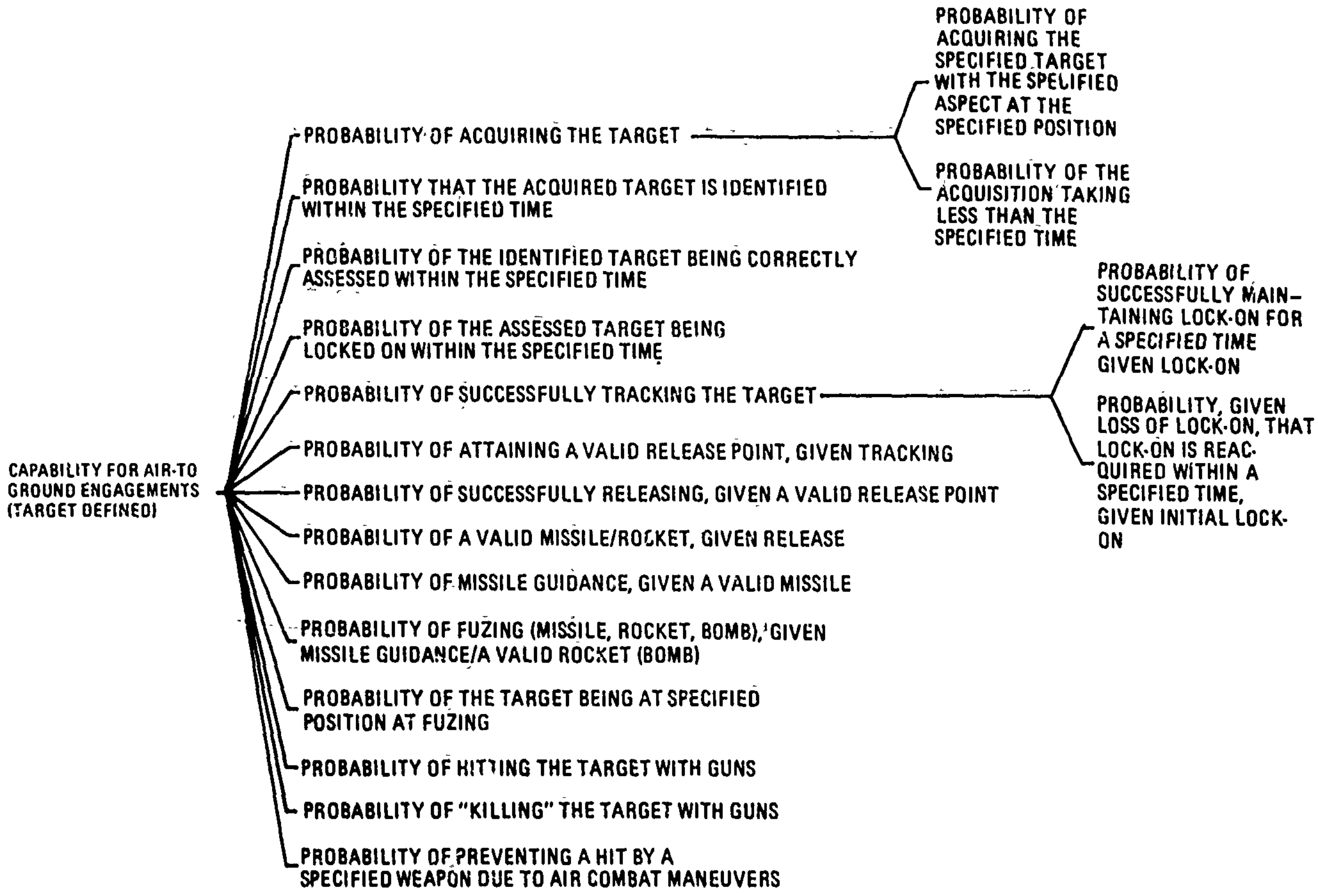


Figure 6-42. Capability for air-to-ground engagement

alone not a very useful parameter. If the tester wishes to report, in addition to the data elements, a "kill probability" he should define what is meant by "kill". For example, if the target was a tank, does the tester mean M-kill, F-kill, K-kill, all three, or some other combination of them. The product of all of the appropriate data elements in Figure 6-42 gives the Capability for Air-to-Ground Engagements. It should be noted that the probability statements that combine to give this Capability are all conditional probabilities. That is, they say given what should have happened before, what is the probability of the next action. For example, "the probability of the acquired target being identified within the specified time" requires that the target was first acquired. For a further discussion of conditional probabilities, see Appendix G.

TABLE 6-4. TARGET DAMAGE CRITERIA (Sheet 1 of 4)

Target	Damage Criteria	
1. Personnel	Defense 30 sec Assault 30 sec Assault 5 min Supply 12 hr	The casualty criteria for personnel are dependent upon the tactical situation. A casualty results when an individual's capacity to perform his military duty is reduced by a specified degree in a specified time. For damage by fragments, three tactical roles with related time periods have been defined. The times are the periods between injury and incapacitation; if the soldier is wounded by one or more fragments so that he is unable to perform a useful military function within his tactical role and he becomes incapacitated within the time specified, he is considered a casualty.
2. Armored Vehicles		
a. Tanks	Mobility, M-kill	Damaged so that the tank is uncontrollable within five minutes and is not reparable by the crew on the battlefield.
	Firepower, F-kill	Defeat of the main armament either because the crew has been rendered incapable of operating it or because the armament or its associated equipment has been rendered inoperative and not reparable by the crew on the battlefield.
	Catastrophic, K-kill	Damaged beyond repair or to the extent that repair is not economically feasible.
b. Armored Personnel Carrier	Mobility, M-kill	Damaged to the extent that the vehicle is uncontrollable within five minutes and is not reparable by the crew on the battlefield.
	Catastrophic, K-kill	Damaged beyond repair or to the extent that repair is not economically feasible.
	Personnel, P-kill	Incapacitation of transported personnel. The P-kill is a calculated ratio of the number of personnel incapacitated to the total number of personnel being transported in the personnel compartment of the APC.
c. Self-propelled AA Gun	(Same as for tanks)	(Same as for tanks)
3. Field Artillery		
a. Artillery	Firepower, F-kill	Damage that prevents the accurate delivery of munitions on intended targets or damage that cannot be repaired in the field.
b. Rocket and Launcher	Firepower, F-kill	Sufficient damage to all rockets, launcher tubes, and/or launcher system to make the weapon or system inoperable until it receives major repair.
	Mobility, M-kill	Damage to the carrier vehicle that immobilizes the vehicle within 5 minutes.
4. Missile Sites		
a. Surface-to-Surface	Firepower, F-kill	Damage that prevents the missile from being launched and is not reparable in the field.
b. Surface-to-Air	Firepower, F-kill	Minimum downtime of four hours.

TABLE 6-4. TARGET DAMAGE CRITERIA (Continued) (Sheet 2 of 4)

Target	Damage Criteria	
5. Radar Installations	Fire control, FC-kill	Damage sufficient to prevent the radar from tracking targets or directing fire. The damage will cause a two-hour downtime under expected operational conditions.
6. Airfields		
a. Parked Aircraft	Prevent takeoff (PTO)	Damage such that the aircraft is unable to generate sufficient power to take off, or the pilot is unable to control the aircraft.
	Castrophic, K-kill	Irreparable damage to the aircraft that renders it unfit for any purpose except cannibalization and scrap.
b. Runways	Interdiction	Surface cratering sufficient to prevent aircraft from taking off or landing; no undamaged part of the runway is long enough or wide enough for use as a take-off strip.
c. Hangars	Complete K-kill	Collapse of the hangar by loss of at least one-half of the roof-supporting beams on one side or destruction of all the hangar shops by demolition of walls or ceiling.
7. Field Fortifications	Complete K-kill	Breaching of the walls or ceiling causing a complete breakdown of structural integrity and possible filling of the interior by dirt, scabbing, and debris.
8. Supply Dumps		
a. Stacked Ammunition		Complete destruction of the stack.
b. POL Storage Tanks		Destruction of the tank, ignition of contents resulting in sustained fire, or leakage from the tank leaving less than 25 percent of the original contents.
c. POL Drums		Destruction of, or setting fire to, 50 percent of the drums within the unit.
9. Land Transportation		
a. Roads		Road damaged sufficiently to prevent passage of vehicular traffic for a specified period of time.
b. Trucks	Mobility M-kill	Damage that immobilizes a moving vehicle within a given time. This type of kill is usually conditional, based on provisions for repairing the damage.
	Category A Category B Category C	Vehicle is immobilized within 5 minutes. Vehicle is immobilized within 20 minutes. Vehicle is immobilized within 40 minutes.
	Interdiction, I-kill	Damage that prevents the vehicle from being used for transportation of supplies to the fighting front. I-kill is subdivided as follows:
	Expedient 0.5, 1.5-, and 8-hour Repair	Expedient Repair indicates that only those components absolutely necessary to operation are repaired. Repairs may be temporary in nature; e.g., holes in liquid systems may be plugged rather than soldered or welded.

TABLE 6-4. TARGET DAMAGE CRITERIA (Continued) (Sheet 3 of 4)

Target	Damage Criteria	
	Thorough 0.5-, 1-5-, and 8-hour Repair	Thorough Repair indicates that all components which contribute significantly to performance, safety, and efficiency of operation are repaired. Repairs are permanent in nature. The repair times indicated mean that at least one component requiring the stated repair time has been damaged. Additional components may have been damaged so that total repair time may be significantly greater than stated. The damage level associated with an 8-hour Expedient Repair damage criterion is significantly greater than that associated with an 8-hour Thorough Repair.
	Catastrophic K-kill	Damage that renders the vehicle unfit for any purpose except salvage.
c. Railroads		Severe damage to, or undermining of a section of track so that a train cannot pass until the damaged section of track is repaired or replaced.
d. Rolling Stock		Derailment, destruction of the superstructure, perforation of tank cars, or other damage to render the rolling stock temporarily unusable.
e. Locomotives		
(1) Moving	S2	Train stops within 5 minutes and is not derailed; stops within 1 mile.
(2) Stationary	R2	Requires more than 15 days to repair.
	R3	Requires 4 to 15 days to repair.
f. Tunnels		Breaching of ceiling or walls, causing complete breakdown of structural integrity and filling by dirt, water, or debris so as to render the tunnel impassable.
10. Bridges	S2	Damaged to the extent that it can no longer support normal traffic.
	S3	Damaged to the extent that it causes a span to drop.
11. Water Transportation		
a. Locks		Deformation or other damage to the gates or caissons so they cannot be moved or cannot retain water in the lock. Pumps and gate-opening machinery damaged, rendering the lock useless. Levees, dikes, or dams must be breached.
b. Port Facilities		Blocking port-entrance channels and wharves by sinking vessels; damage to the port rail- or road-clearance facilities; damage or destruction of port warehouses and transit sheds or damage to loading cranes on wharves.
c. Merchant Ships		Sinking, 100-percent seaworthiness kill.
d. Junks		Sinking kill based on the buoyancy/cargo relationship. Requires structural damage in critical portion of the hull, uncontrolled leaks in all compartments, flooding within 20 minutes, loss of maneuverability, inability to reach shallow water, and loss by sinking if the junk is motorized or carying high-density cargo.

TABLE 6-4. TARGET DAMAGE CRITERIA (Continued) (Sheet 4 of 4)

Target	Damage Criteria
e. Sampans	Same as Junks.
f. Barges and Small Craft	Sinking, 100-percent seaworthiness kill.
12. Soviet Naval Vessles	
a. Destroyer Escort	Sinking or degradation of mission effectiveness.*
b. Destroyer	Sinking or degradation of mission effectiveness.*
c. Helicopter Carrier	Sinking or degradation of mission effectiveness.*
d. Light Cruiser	Sinking or degradation of mission effectiveness.*
e. Patrol Boat	Sinking or degradation of mission effectiveness.*
f. Landing Ship	Sinking, 100-percent seaworthiness kill.
g. Guided-Missile-Destroyer Leader	Sinking or degradation of mission effectiveness.*
h. Guided-Missile Destroyer	Sinking or degradation of mission effectiveness.*
i. Minesweeper	Sinking or degradation of mission effectiveness.*
13. Buildings	Structural damage of a specified level (percent of roof or floor area). Structural damage is damage to principal load-carrying members (trusses, beams, columns, load-bearing walls) requiring replacement or special support during repair. Fifty-percent structural damage implies that half of the total floor space has been rendered unusable.
14. Powerplants	Damage to critical components to the extent that the plant is no longer capable of generating or transmitting the specified percentage of its output capacity for at least 60 days.
15. Dams	
a. Earthen Dams	Breaching of the dam to the extent that erosion significantly reduces the storage capacity of the reservoir.
b. Concrete Dams	Breaching or displacing the dam to release impounded water; no reliance on erosion to deepen breach.
16. Industrial Targets	See discussion for individual target elements.
17. Transformer Substations	Damage to transformers or circuit breakers so as to achieve either a severe (1-month) or moderate (36-hour) downtime.

*Incapacitation of at least one weapon system and/or appreciable reduction of maximum speed.

The Capability for Search and Rescue/Recovery (SAR) missions is covered by the data elements shown in Figure 6-43. The specified area referred to in the data elements can be near a base, a battle zone, a combat zone, or an area of open sea or land. The area must be specified by the tester. The use of the terms "suppressing a configuration" refers to suppressing a fire which often occurs after a crash of an aircraft. This term was used to prevent confusing it with "fire suppression" which is used to refer to decreasing the enemy's fire. This latter ac ion is often required during a SAR mission in a battle or combat zone. The study of fire suppression (enemy fire from weapons) is done by taking the results of the Capability for Air-to-Ground Engagement and evaluating how this reflects itself on the enemy's fire. It is not possible to do testing of this effect in peacetime as it requires knowledge about the enemy's tactics, equipment, training, psychology of the troops, etc. Since MOE's refer to measures, this subject, and those others that can only be done by studies, are not covered in this manual.

Airlift is another mission for Aircraft Systems. Figure 6-44 shows the different Airlift missions and the Capability statements for each one. Aircraft ferrying missions are covered by a cargo Airlift mission where the specified cargo is the aircraft itself. Here again, the data elements require broad interpretation so that the total data element list is of reasonable length. The tester should always give an explanation of what he means, explicitly, by the data elements if there is any possibility of confusion.

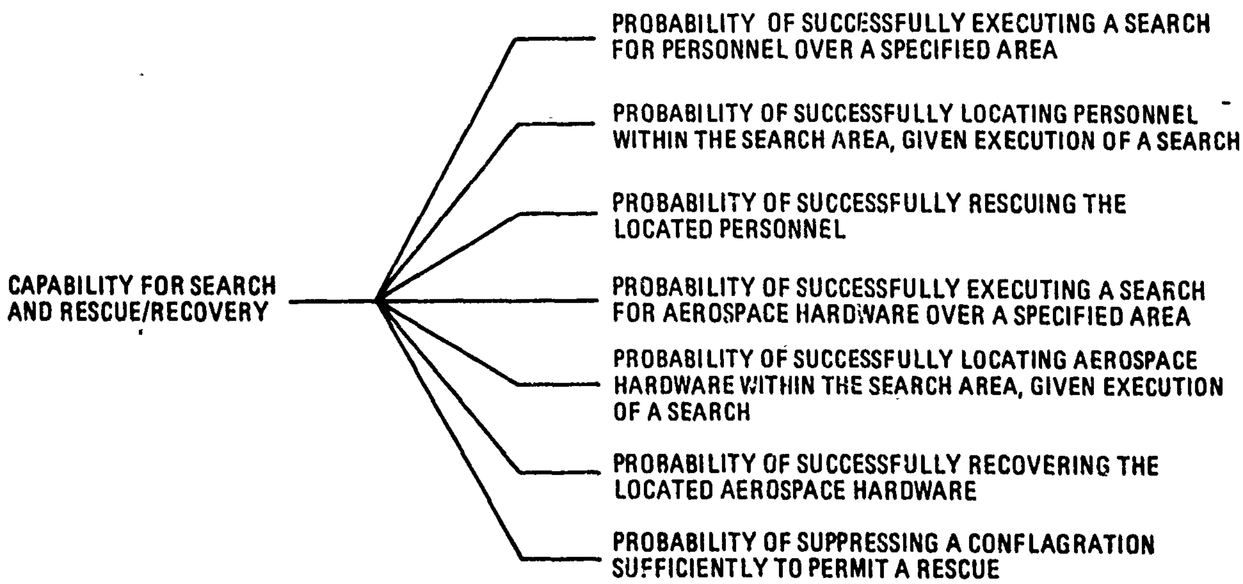


Figure 6-43. Capability for search and rescue recovery

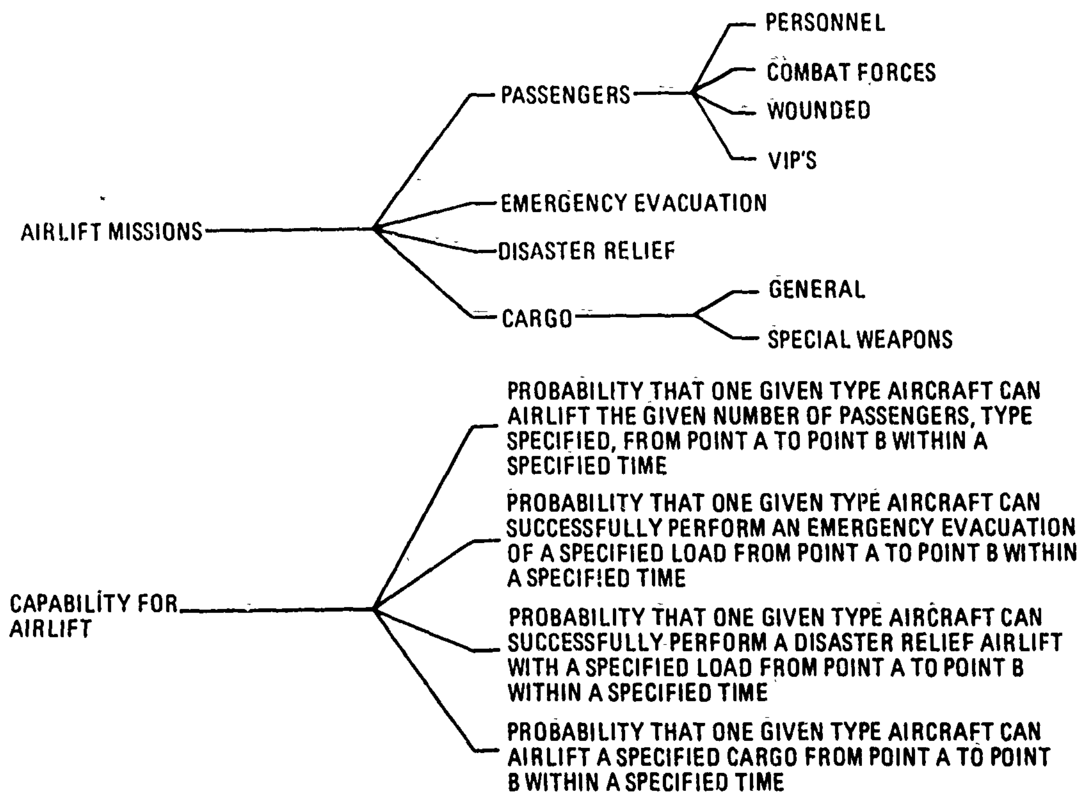


Figure 6-44. Airlift missions and Capability

Typical Command and Control missions which utilize Aircraft Systems are shown in Figure 6-45. The data elements and Capability statements which apply to this aircraft mission can be found in the Communications/Electronics section. The tester selects those data elements that apply to the methods of operation and mission of his test Aircraft System. See, in particular, the data elements in Figures 6-9 and 6-10 for Point-to-Point Telecommunications Capabilities and Search and Detection Radar Capabilities.

Another important aircraft mission is Reconnaissance. Figure 6-46 shows the different types of equipment used in reconnaissance, the different Reconnaissance missions and the Capability statements for each of the missions. The data elements that make up each of the Capability statements depend upon the type of equipment used. In general, these data elements are covered in the section on Communications/Electronics (CE) Capability. As always, only those data elements needed (and applicable) to the system and scenario under test need be selected. For visual data gathering, the data elements are the same as some of the CE Capability data elements. For example, the "probability of target detection at a specified position with a specified time".

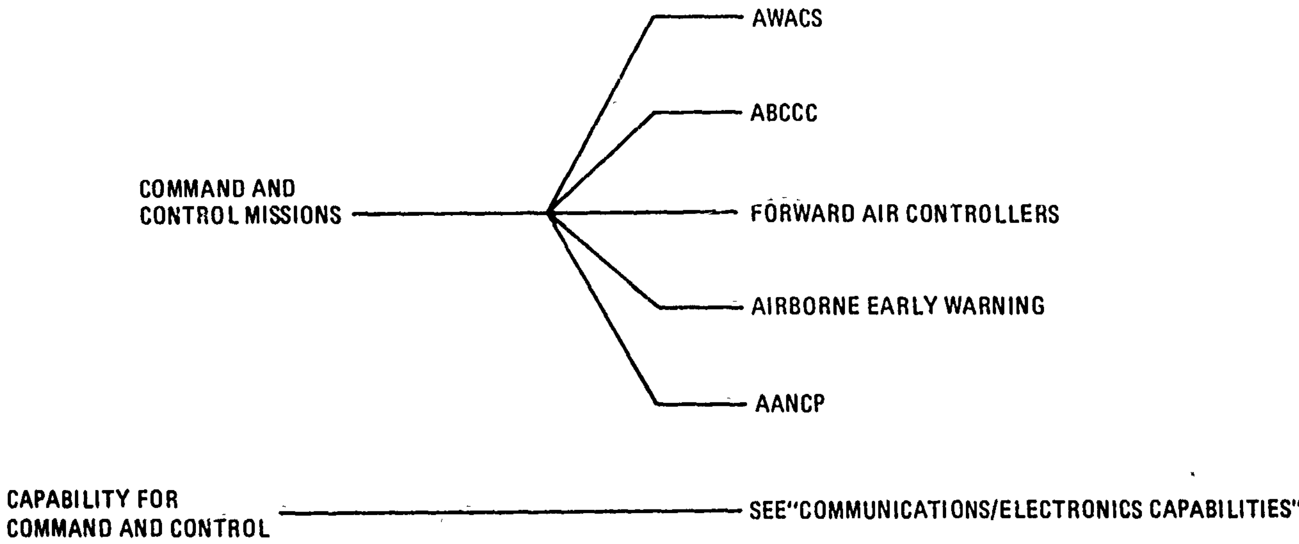


Figure 6-45. Command and control missions and capability

The Capability for Electronic Warfare (EW) is shown in Figure 6-47. The Capability for Signal Intelligence is covered by the data elements for Signal Intelligence in the CE section. The data elements for Electronic Countermeasures Capability are covered by those for Defense Suppression Capabilities in the Defense Suppression Section. The Capability for Electronic Counter-countermeasures are obtained by measuring the appropriate system's Capability when it is in a specified electromagnetic environment. This environment should be specified in the scenario. According to AFM11-1 Vol. III, Electronic Countermeasures is "the division of Electronic Warfare involving actions taken to prevent or reduce an enemy's effective use of the electromagnetic spectrum. It includes:

1. Jamming. The deliberate radiation, reradiation, or reflection of electromagnetic energy with the object of impairing the use of electronic devices, equipment, or systems being used by an enemy.
2. Deception. The deliberate radiation, reradiation, alteration, absorption, or reflection of electromagnetic energy in a manner intended to mislead an enemy in the interpretation or use of information received by his electronic systems. It includes:
   (a) Imitative. Introducing radiations into enemy channels which imitate his own emissions.

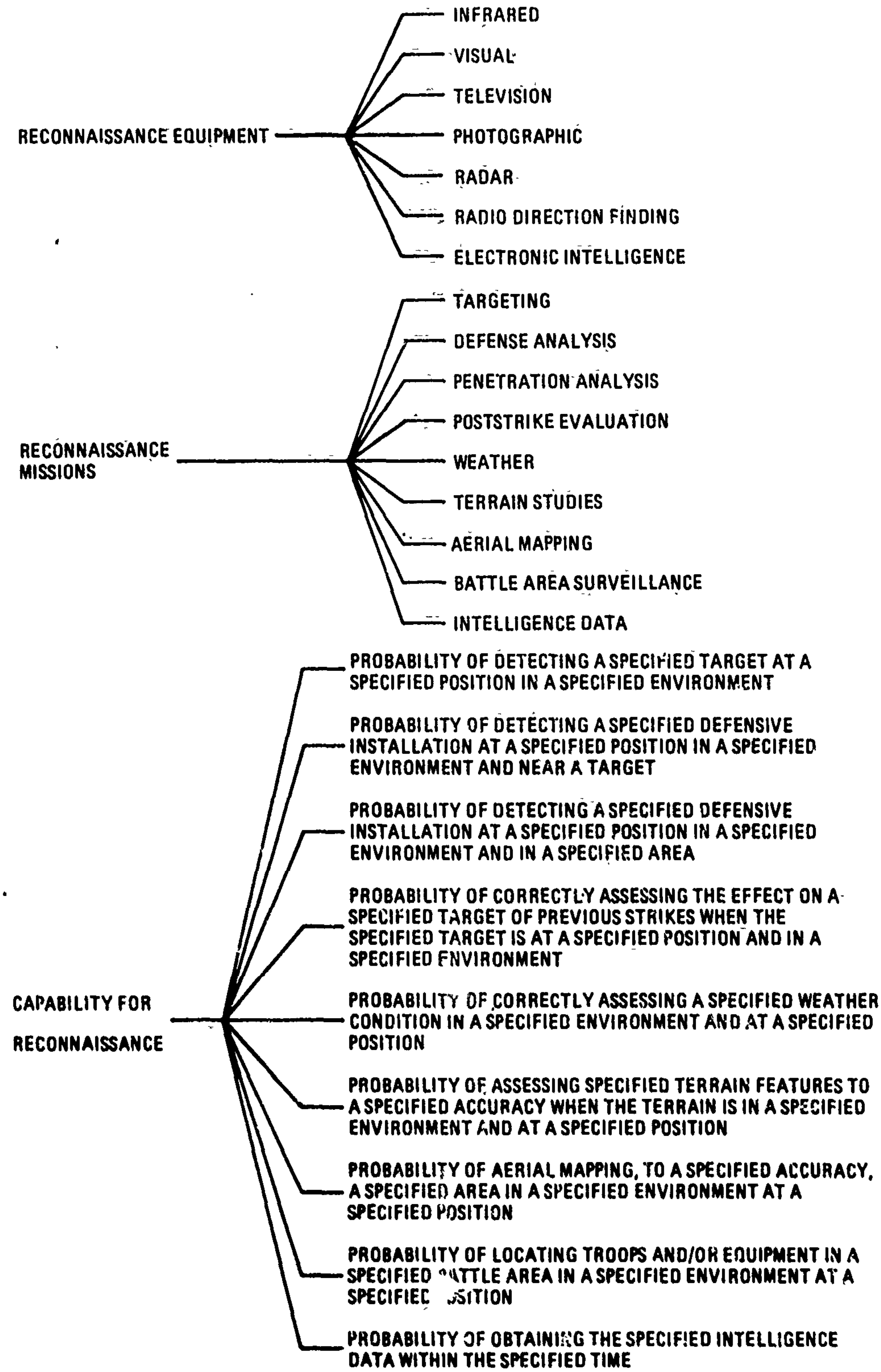


Figure 6-46. Reconnaissance

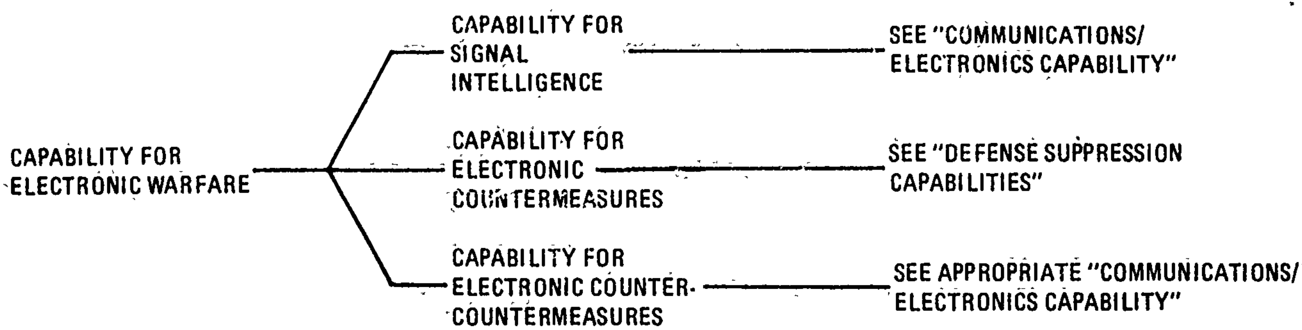


Figure 6-47. Capability for electronic warfare

(b) Manipulative. The alteration or simulation of friendly electromagnetic radiations to accomplish deception."

The meaning of Electronic Counter-countermeasures, in the same reference, is "the division of electronic warfare involving actions taken to insure friendly effective use of the electromagnetic spectrum despite the enemy's use of electronic warfare".

Airborne Atmospheric Sampling Capabilities are obtained from the data elements shown in Figure 6-48. The data elements that are the inputs to the probability statements in Figure 6-48 depend on many things such as sample particle size, gas sampling, total sample size, sampling altitude, sample distribution in the atmosphere, etc. The precise inputs to be used are left to the tester. However, these inputs should result in the evaluation of the appropriate probability specified in Figure 6-48.

Training of personnel, while a very important aspect of every mission, seldom has OT&E done on it. The probability statements shown in Figure 6-49 are measures of the Capability for Training. For the Training mission, the scenario would be the course of instruction given the specified personnel. By specified personnel is meant those that have the required qualifications for the particular specialty. The words "being able to perform assignments as ..." means that the specified individual can perform all assignments expected of personnel at that level of training and experience. Equipment repairman refers to any type of equipment such as, autos, aircraft engines, electronics, weapons, etc. Equipment operators are all those not specifically covered by the other titles, e.g., vehicle drivers.

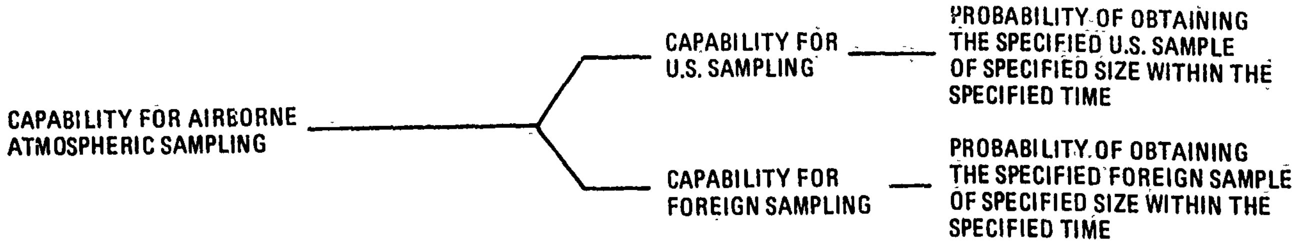


Figure 6-48. Capability for airborne atmospheric sampling

The Capabilities for Airborne Test Beds are shown in Figure 6-50. Since test bed aircraft are highly specialized and highly flexible (in that, normally, they can make many measurements with little modification), it is difficult to get much lower in detail than the data elements shown in Figure 6-50. The tester must select data elements that can yield the probabilities in Figure 6-50.

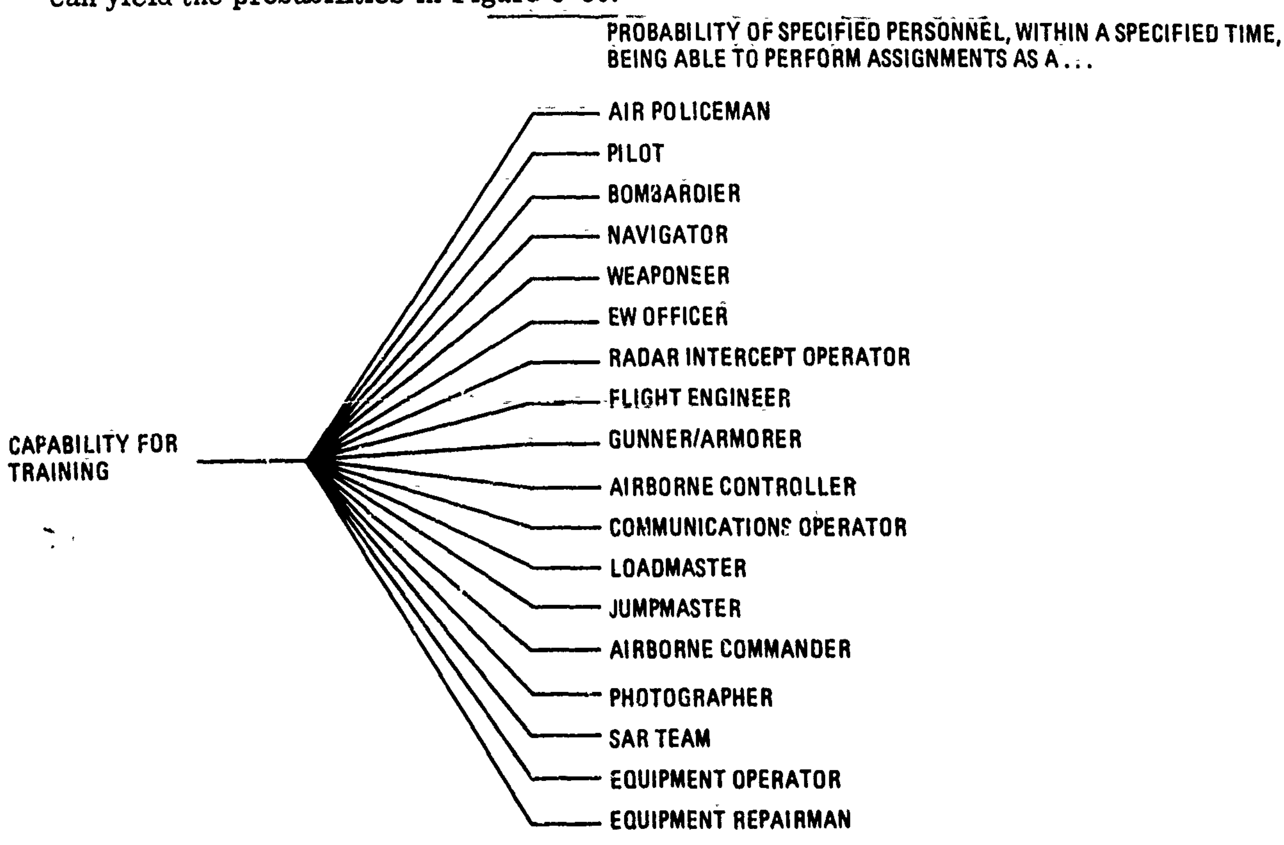


Figure 6-49. Capability for training

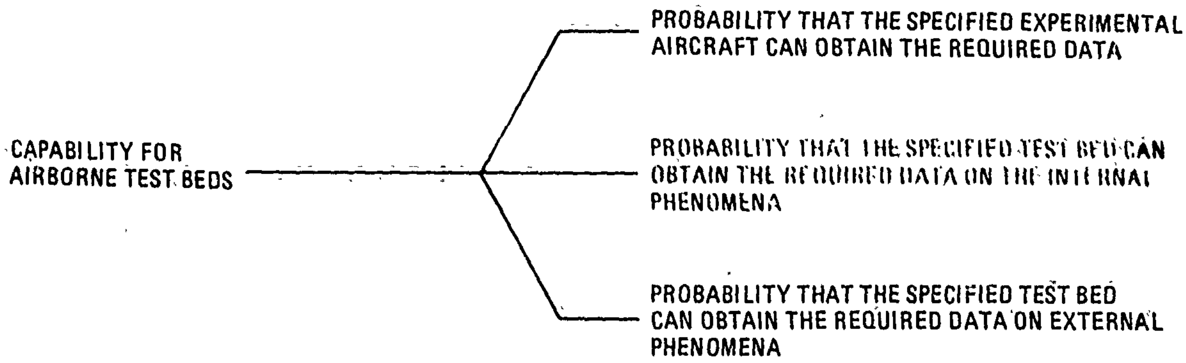


Figure 6-50. Capability for airborne test beds

As would be expected, the Capability for Refueling refers to the aircraft whose mission is to refuel other aircraft. The aircraft that receives the fuel is covered by the Dependability for refueling. The data elements for Capability for Refueling are shown in Figure 6-51. The first of these statements refers to the Aircraft System's Capability of carrying the required amount of fuel to the designated rendezvous point or points. The other data elements are self-explanatory. The product of the data elements gives the Capability for Refueling.

The Battlefield Illumination mission is covered in the Capability statement and data elements shown in Figure 6-52. Again, the product of the data elements gives the Capability for Battlefield Illumination.

Since there is no operational mission for the Demonstration Team, no Capability statement can be measured in OT&E. The purpose of the Demonstration Team is public relations; OT&E is normally not done on public relations.

In all cases where actions are of a sequential nature and these actions lead to the execution of the Capability of the system, the probability statements have been made conditional probabilities. That is, they state the probability for accomplishing action B, given that action A was accomplished. The reason for this is that the product of the two probabilities (probability of accomplishing A and probability of accomplishing B given A was accomplished) is just the probability of accomplishing A and then accomplishing B. For a further discussion of conditional probabilities, see Appendix G.

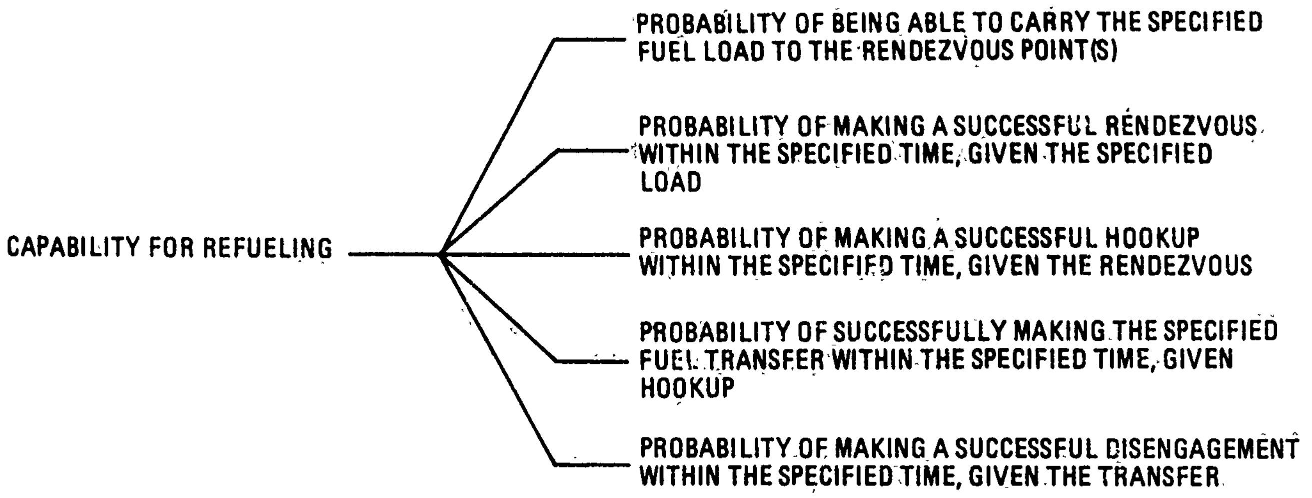


Figure 6-51. Capability for Refueling

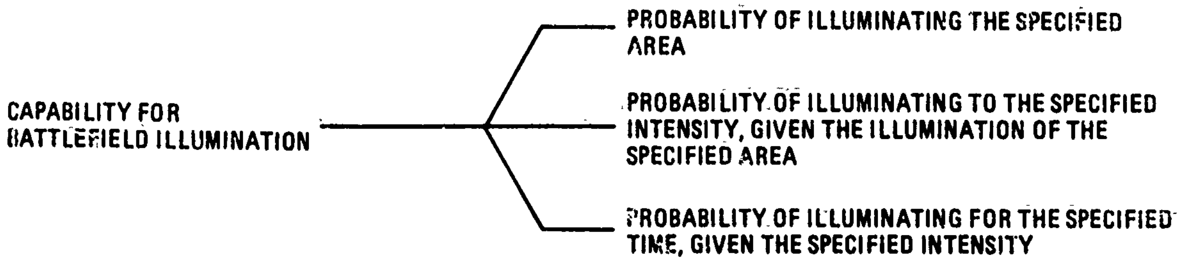


Figure 6-52. Capability for battlefield illumination

9. MEASURES OF EFFECTIVENESS FOR MISSILES

Before a discussion of Measures of Effectiveness for Missiles can be undertaken, it must be made clear what is meant by the term "Missile." Weapons that are explosive devices can be guided or unguided and have powered or nonpowered flight. Thus, there can be four different weapons that are explosive devices. As used here, they are:

a. A bomb weapon is an unguided explosive device with nonpowered flight.

b. A "smart" bomb weapon is a guided explosive device with nonpowered flight.

c. A rocket weapon is an unguided explosive device with powered flight.

d. A missile weapon is a guided explosive device with powered flight (often called a guided missile).

The MOE's to be developed herein apply only to weapons that are guided explosive devices with powered flight. There will be a further restriction on the missiles to be included here in that only nonnuclear weapons are to be considered.

With these restrictions, the only missiles to be discussed are those Missile Subsystems that are part of an Aircraft System.

The expendable load Availability was one of the data elements for an Aircraft Systems Availability as shown in Figure 6-19. Missiles are one of the possible expendable loads. Before the missile is considered for mating with the aircraft, there are several types of checks that the missile may undergo. One of these checks is called the missile container inspection check. The data elements which are involved in this check are shown in Figure 6-53. This check is performed on containers which have a missile in them.

Another check for which there may be a requirement, is called the missile visual check. The data elements which are involved in passing this check are shown in Figure 6-54. Again, the tester selects only those data elements which apply to the Missile Subsystem that is part of the Aircraft System under test. A good source of guidance as to which data elements to select are the T. O.'s that apply to the Missile Subsystem.

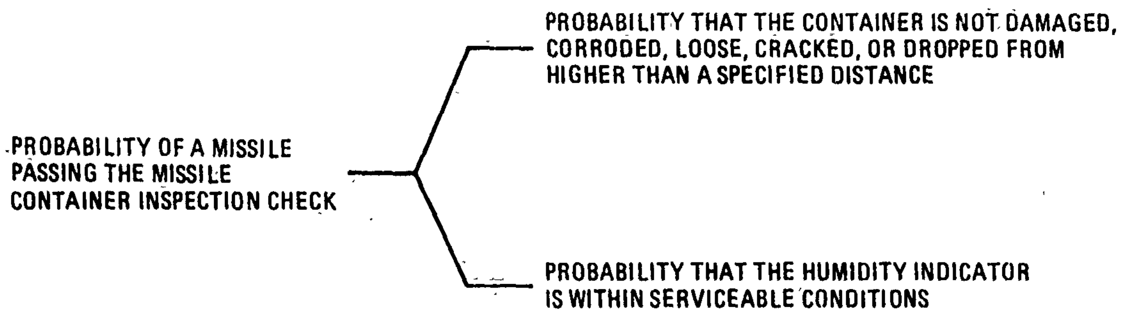


Figure 6-53. Probability of Missile Passing Missile Container Inspection Check

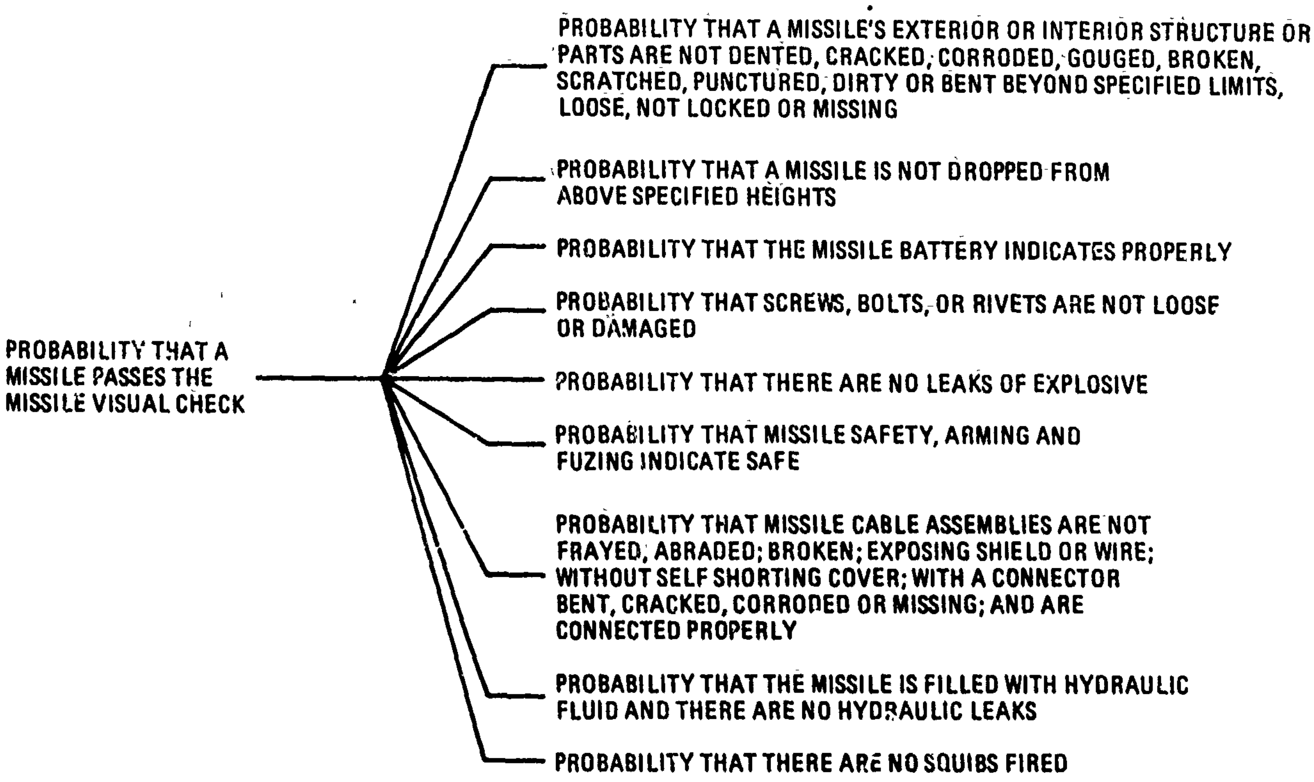


Figure 6-54. Probability that missile passes visual check

A more detailed check of a missile is called the complete missile functional test. The probability that a missile passes this functional test is covered by the data elements shown in Figure 6-55. Since, by definition, the test set is part of the Missile Subsystem, the first of these data elements addresses the probability of the test set passing its self check. The other two probabilities are conditional

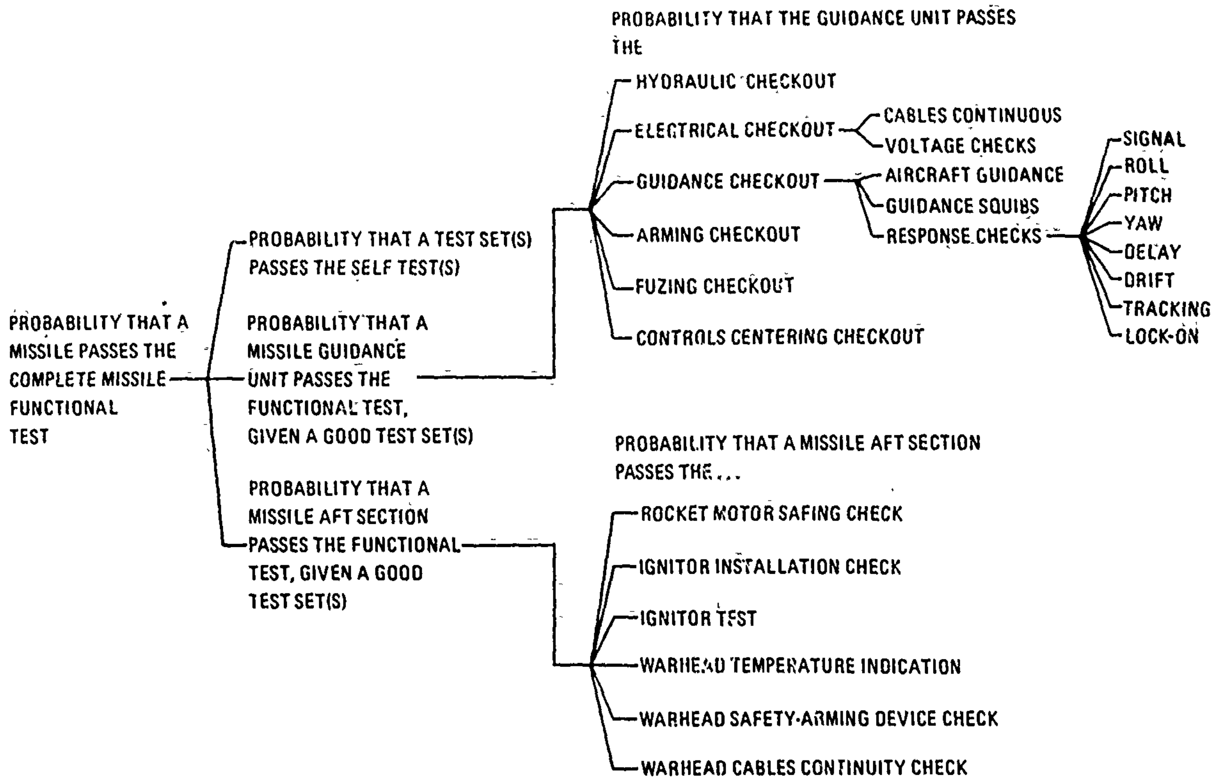


Figure 6-55. Probability that missile passes functional test.

probabilities in that they require the test set(s) to pass the self test(s). The other two tests are independent of each other since they check different sections of the missile, one is the guidance unit and the other, called the aft section, is the rest of the missile. The data elements that give the probability that the guidance unit passes the functional test are also shown in Figure 6-55. Similarly, Figure 6-55 shows the data elements for the missile aft section passing the functional test.

The checks covered by the data elements in Figure 6-53, 6-54 and 6-55 cover all of the checks done to have a missile available for a mission. However, certain missiles are handled and carried in what are called "missile clusters." A missile cluster is a group of one or more missiles of the same type that are loaded onto a launcher; the missile(s) and launcher are then mated to the aircraft as a single unit. These missile clusters also have a set of checks that they must pass before they can be considered available for mating to the aircraft.

The first of the checks of missile clusters is covered by Figure 6-56. The check is called the missile cluster visual inspection. The data element referring to each missile passing the missile visual check is covered by the data elements in Figure 6-54. The other data elements are self-explanatory. Again, the T.O.'s for the missile cluster are a good guide as to which of the data elements in Figure 6-56 are to be used and their exact meaning. As always, data elements must be interpreted in general terms so that they can apply to as many items of the same type (missiles, in this case) as possible.

The other missile cluster check to be covered is called the complete missile cluster functional test. The data elements that give the probability of the missile cluster passing this test are shown in Figure 6-57. Again, four of the probabilities are conditional probabilities requiring that the test set(s) be good (passes its self test). The probability that the launcher passes the complete missile cluster functional test depends upon its electronics unit passing the test as each missile is checked. In other words, in general, there is not a separate check of just the launcher electronics unit when the

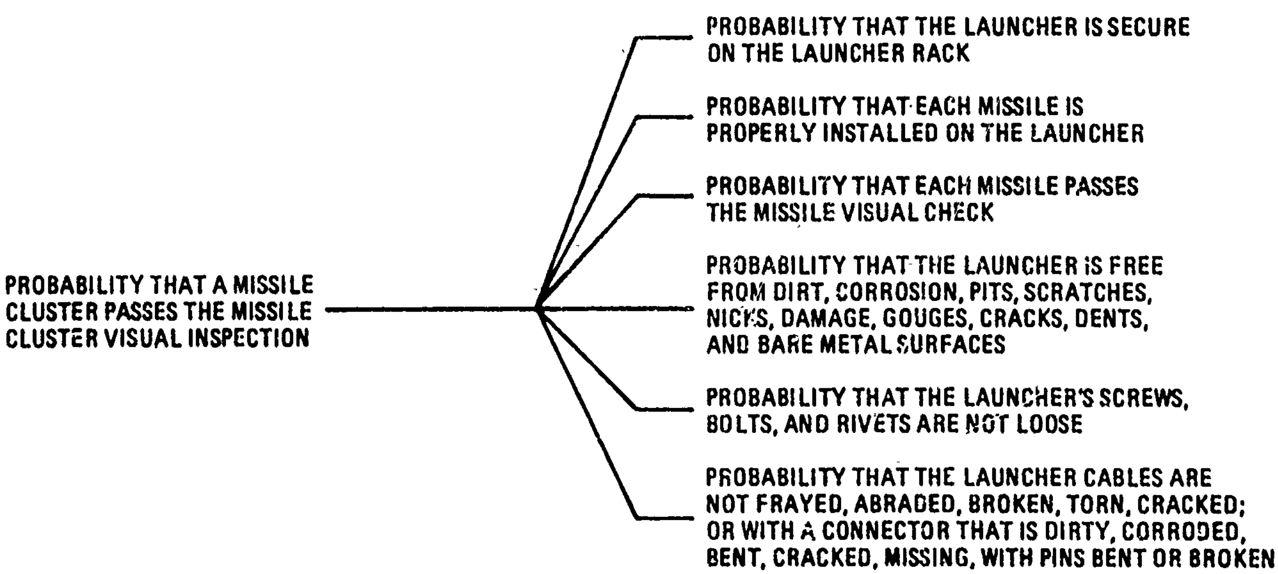


Figure 6-56. Probability that missile cluster passes visual inspection

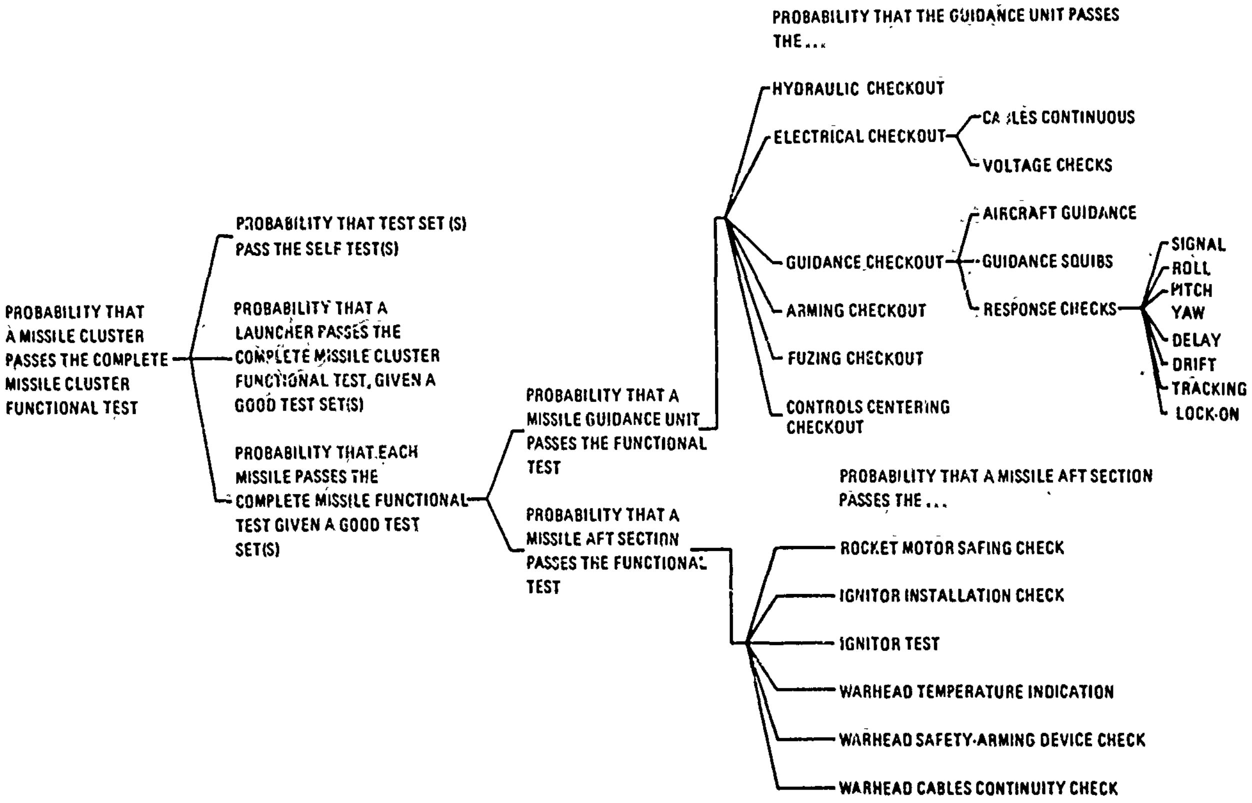


Figure 6-57. Probability that missile cluster passes functional test

complete missile cluster functional test is performed. The other data elements for each missile's functional test are the same as those shown in Figure 6-55.

With the different types of test defined, it is now possible to address the Availability of a Missile or a Missile Cluster Subsystem. First, a missile's Availability will depend upon from where the missile originates. Different checks are made depending upon what is the origin of the missile. The missile could be coming from storage, the pre-loading area, the flightline, receiving, missile maintenance, or periodic inspection. The type of test(s) performed for each of these sources and the appropriate probability statements are shown in Figure 6-58. The appropriate figure is referenced after each data element statement. The statement relative to a missile returned from the flight-line passing the missile container check, applies to missiles that are returned in their container. All of the other statements are self-explanatory.

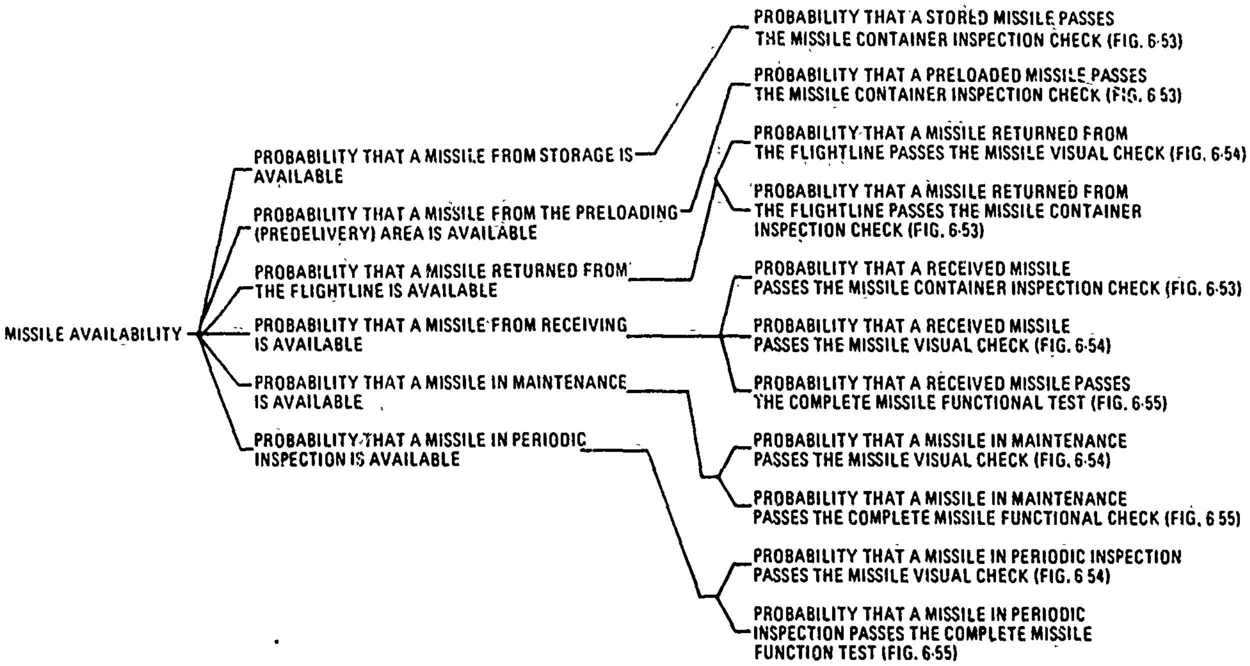


Figure 6-58. Missile Availability

The Availability of a missile cluster is addressed by the data elements shown in Figure 6-59. The missile cluster has fewer sources of supply than the missile since the missile cluster can be from the preloading area, the flightline, maintenance, or periodic inspection. Again, each of the appropriate tests are referenced in the data element probability statement.

All failures of any consequence during any of the different tests shown in Figures 6-53 through 6-57 result in the return of the missile(s) or launchers to the depot for repair. Hence, no attempt will be made to go into greater details in the data elements. Also data elements are not given below the level shown because the depot is not part of the operating command. The MOE's are to be addressed at the user level which is the operating command.

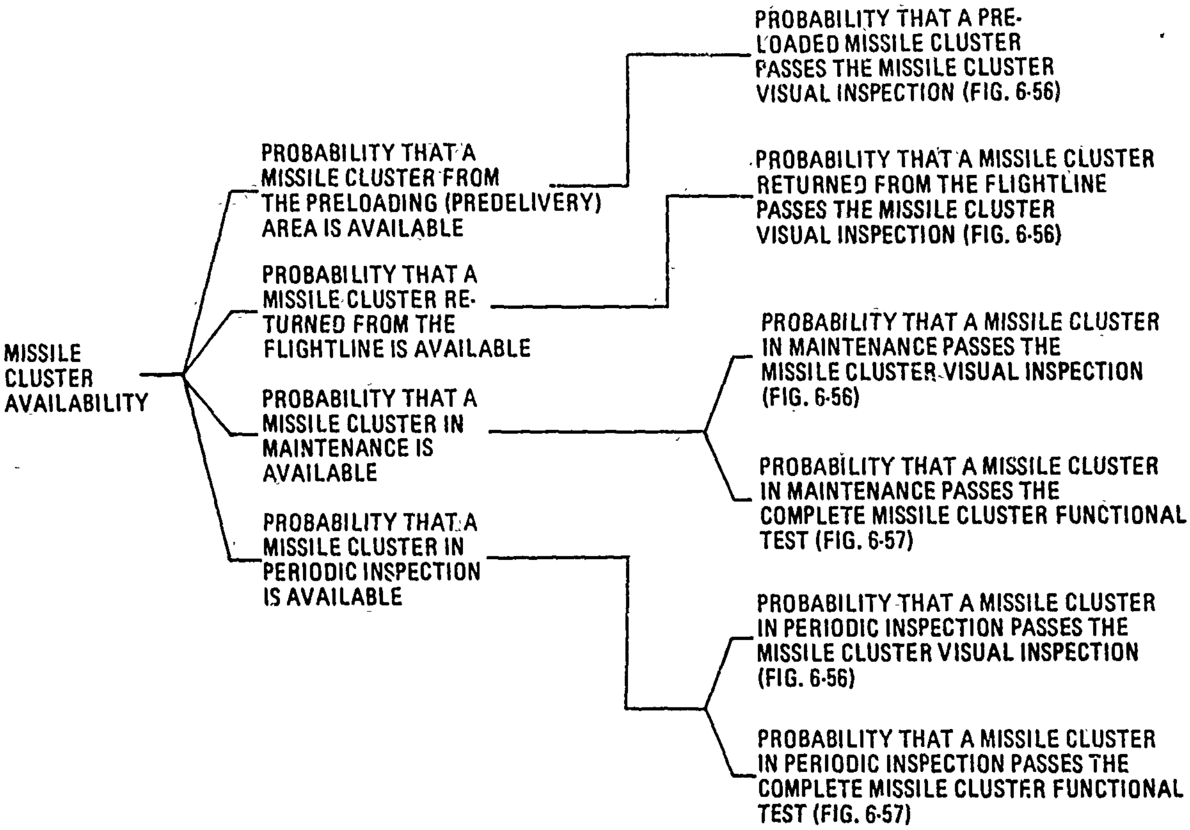


Figure 6-59. Missile cluster Availability

Once the missile or missile cluster is available, it is mated to the aircraft and becomes part of the Aircraft System. At that point, the data elements that address the aircrew preflight check are covered in Aircraft System's Availability.

The Missile or Missile Cluster Subsystem's Dependability is addressed in Aircraft System's Dependability as shown in Figure 6-60. The reason is that the missile or missile cluster is a subsystem in the Aircraft System. The same is true for the missile or missile cluster subsystem Capability as shown in Figure 6-61. If they are air-to-air missiles, the Capability is covered by Aircraft System's air-to-air capability (Figure 6-38). If the missiles are air-to-ground missiles, then see Aircraft System's air-to-ground capability (Figure 6-42).

MISSILE DEPENDABILITY – SEE "AIRCRAFT SYSTEM DEPENDABILITY"

MISSILE CLUSTER DEPENDABILITY – SEE "AIRCRAFT SYSTEM DEPENDABILITY"

Figure 6-60. Missile Dependability missile cluster

MISSILE CAPABILITY – SEE "AIRCRAFT SYSTEM CAPABILITY"

MISSILE CLUSTER CAPABILITY – SEE "AIRCRAFT SYSTEM CAPABILITY"

Figure 6-61. Missile Capability missile cluster

10. SYSTEM ATTRIBUTES OTHER THAN EFFECTIVENESS

There are many system attributes other than system effectiveness that are of interest to the operational commanders. These other attributes of a system are not measured directly. Only their effect on A, D, and/or C (i. e., system effectiveness) will be measured by MOE's. The results of reliability (the probability that a system, subsystem, or equipment will perform a required function under specified conditions, without failure, for a specified period of time - AFM 11-I Vol. I) is nothing more than the data elements that make up the measure of performance, Dependability. Maintainability (a characteristic of design and installation which is expressed as the probability that an item will conform to specified conditions within a given period of time when maintenance action is performed in accordance with prescribed procedures and resources - AFM 11-1 Vol. I) is reflected in the item's Availability given that some maintenance action is required. Interoperability (the ability of systems, units or forces to provide services to and accept services from other systems, units or forces and to use the services so exchanged to enable them to operate effectively together - AFM 11-1 Vol. I) is addressed whenever the definition of the test includes items with which the test item must interoperate. Survivability (the capability of a system to withstand a man-made hostile environment without suffering an abortive impairment of its ability to accomplish its designated mission. - AFM 11-1 Vol. I) is addressed by evaluating the MOE of an item with and without the man-made environment. The ratio of the MOE with the man-made environment to the MOE without it is a good index of survivability.

Compatibility can affect the MOE of an item but is not measured by the MOE.

Needed system modifications can be determined, strictly from a system effectiveness viewpoint, by determining which data elements lead to low values of A, D, and/or C.

Other aspects of a system (e. g. doctrine, organization, operational techniques, tactics, and training of operator and maintenance personnel) can be examined only by how these various aspects affect the system effectiveness. It is in no way a measure of these aspects when they are varied and the effect on the MOE is noted. The effect on the MOE is only one of the many important features of these aspects. For example, one consideration in deciding between two tactics should be the system effectiveness for each tactic. This can be determined by the MOE using each tactic. Other aspects of the tactics must also be examined. Some of these are the vulnerability of the system using each tactic, the ease of using each, the training required for each, the affect on interoperability with other systems, etc.

It cannot be stressed too highly that MOE's only measure system effectiveness as defined presently by approved DOD documents. Other attributes of systems must be evaluated by some other means.

11. EXAMPLE OF MOE APPLICATION

The procedures that the tester must follow to use the MOE's and measures of performance (A, D, and C) as shown in this manual can best be summarized by an illustrative and hypothetical example utilizing the MOE check list shown at the end of this chapter.

The example to be used will lead to a brief and simplified test. It will be assumed that data gathered previously has indicated that there is trouble with a particular assembly. For this reason, a new assembly has been developed and DT&E has been conducted. It is now appropriate to do some IOT&E on the new assembly. The data to be collected will be used to evaluate the new assembly and to calculate what the system's A, D, C and MOE would be if the new assembly were used to replace the old one.

First, the tester must explicitly define the item to be tested. For the example, the test item is a new assembly used to measure the hydraulic pressure in an F-24D/ AGM-105A Aircraft System. The assembly uses a sensor on the output side of the hydraulic pump in the F-24D. The AGM-105A does not use the aircraft hydraulic power set. The output of this sensor is a voltage that is amplified and displayed in the same location as the old hydraulic pressure indicator. The test objective stated in the Test Directive is to determine the effect of the new sensor on the system's MOE for an Air-to-Ground Engagement against tanks in a tropical environment. Knowing this objective, the tester must now select the level of the test between multi-system, system, subsystem and set as shown in the check list. He must also select what type system is involved. In this case, the test item belongs to an Aircraft System. Looking at Figure 6-19, it is obvious that the test item only affects the Aircraft Subsystem (the F-24D). In fact, it can only affect part of the aircraft, therefore the test should be conducted at the "set" level. It is also clear from the method of operation of the test item that it only has an interaction with the hydraulic and electrical power sets of the Aircraft Subsystem.

The tester must now define the scenario for his test. The mission has been given as an Air-to-Ground Engagement against tanks. The environment was given as tropical.

The tester selects a test location in a tropical environment and plans typical flight profiles for Air-to-Ground Engagements against tanks. The Aircraft System's Capability for an Air-to-Ground Engagement is determined by the data elements in Figure 6-42. Examination of these data elements indicates that the test item has no effect on the system's Capability. Thus, the tester does not need to gather data that leads to Capability. Other scenario considerations listed in the check list are not applicable to this test item.

The tester has now defined his test item, the items with which it interoperates, the scenario, and the limits of his data elements in that only the Aircraft Subsystem's Availability and Dependability are of concern.

The tester now selects all data elements for the System's Availability that apply to the test item or items with which it interoperates. That is, all data elements that the test item can affect. The results of this selection are shown in Figure 6-62. The Figure number after each data element is the figure from which the data element was taken. The letter (A_1, A_2, etc.) is the symbol to be used for the data element. All other data elements in Figures 6-19, 6-20, 6-21, 6-22, and 6-23 do not apply in this example. The values of the data elements that are applicable are obtained from the number of aborts caused by the observations indicated in Figure 6-62.

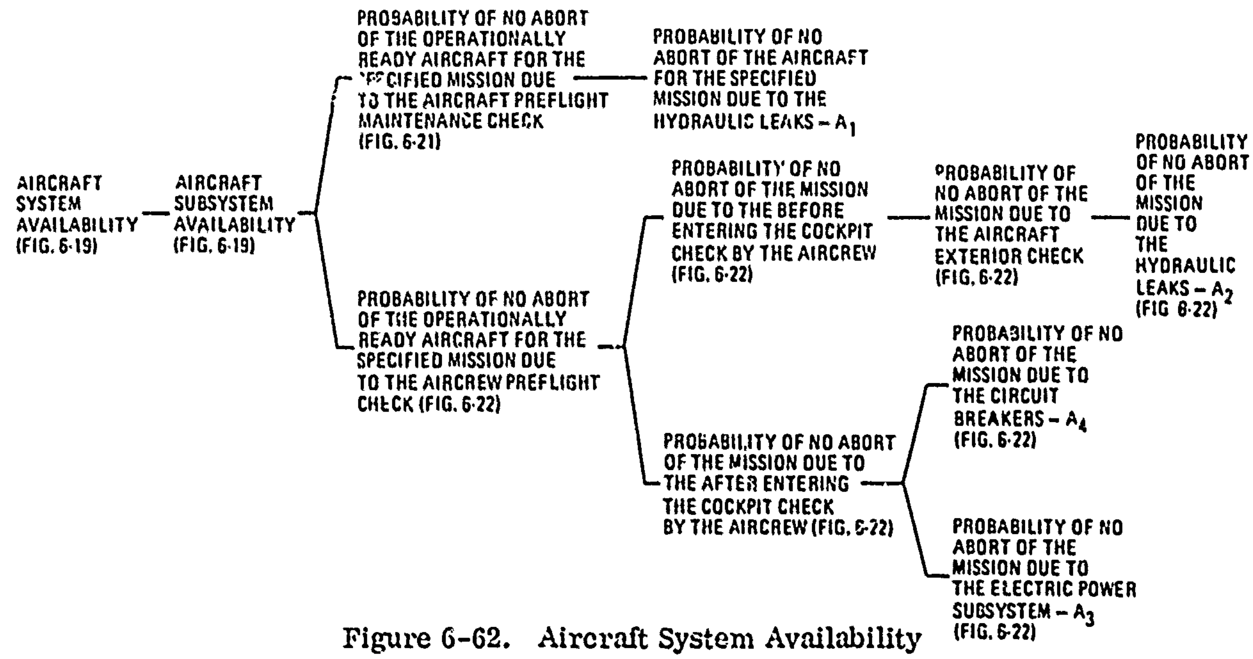


Figure 6-62. Aircraft System Availability

The tester now goes through the same process for the Aircraft System's Dependability. Of the Dependability time frames shown in Figure 6-24, it turns out that the hydraulic and electrical power subsystems are only checked during engine start (Figure 6-25), taxi before takeoff (Figure 6-26), cruise (Figure 6-29) descent (Figure 6-31), and postflight maintenance check (Figure 6-36). Thus the data elements for Dependability during takeoff (Figure 6-27) climb (Figure 6-28), refuel (Figure 6-30), land (Figure 6-32), taxi after landing (Figure 6-33), expendable load safing (Figure 6-34), and the aircrew postflight check (Figure 6-35) are independant of the test item and are, therefore, not measured.

Selecting the applicable data elements from the appropriate figures leads to the indicated data elements shown in Figure 6-63. All others in Figures 6-25, 6-26, 6-29, 6-31, and 6-36 are not applicable to the test item of this example. The tester will evaluate the data elements from the number of aborts attributable to each of the data elements. The letter (D_1, D_2, . . ., D_{21}) symbols listed after each data element will be used to refer to specific data elements.

It will be assumed that the tester will measure each of these data elements for the new assembly and that he knows their value from past experience with the old assembly, i.e., from data available from the Air Force standardized maintenance data collection system.

He must be assured that the scenarios were the same for the data for the old assembly. If they are not the same, or if he is not sure they are the same, he must test the old system also.

During the test, the tester notes the number of aborts attributable to each data element and the number of attempts; a method of obtaining this information must be incorporated into the test design.

Assume that the results of such a test design are as shown in Table 6-5 which also shows the values of the 95% confidence limits of each of the data elements. The effect of the new assembly on the Aircraft System's Availability and Dependability is shown

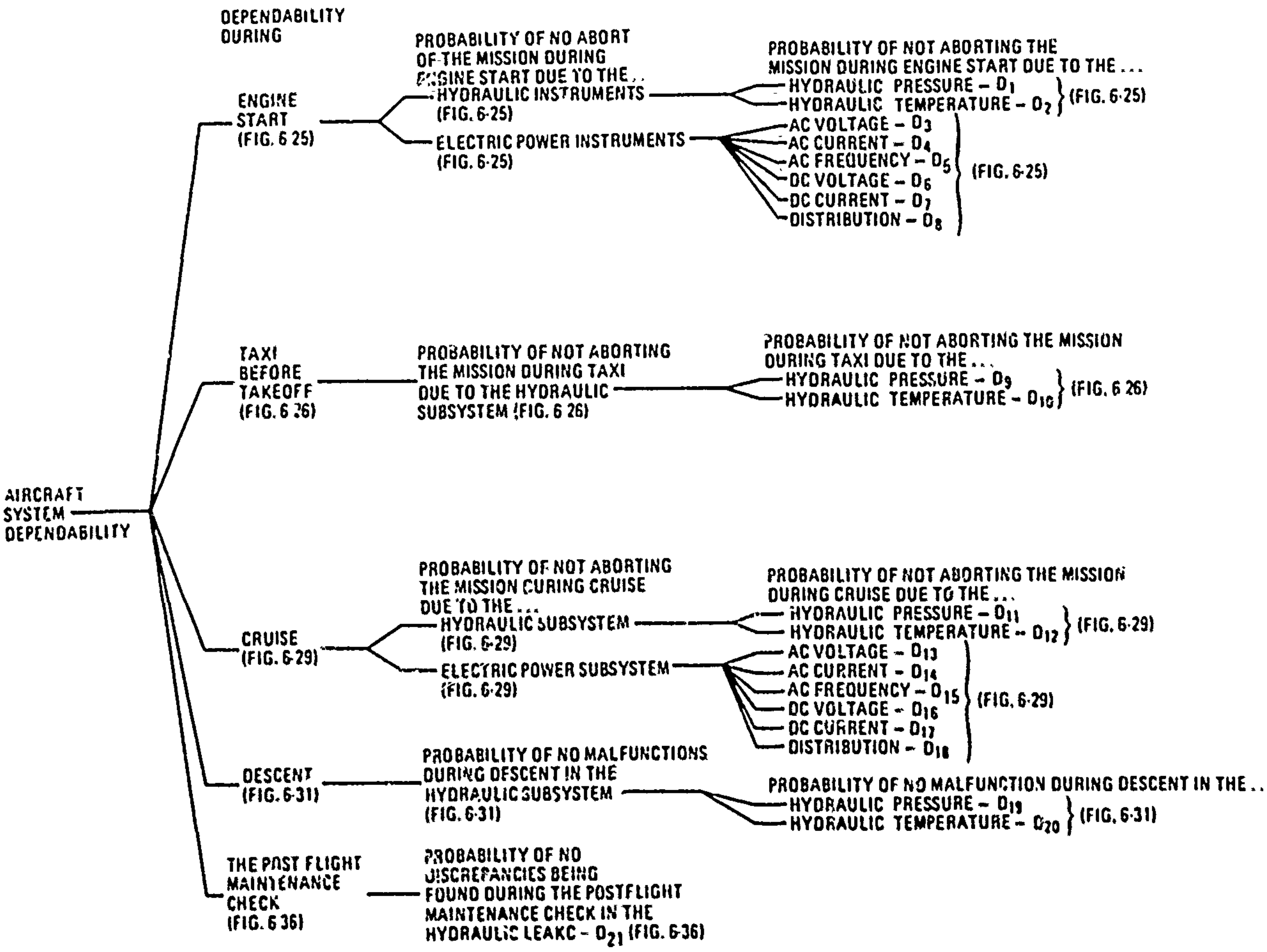


Figure 6-63. Aircraft System Dependability

TABLE 6-5. TEST RESULTS

Data Element		Old Assembly			New Assembly		
			95% Confidence Limits			95% Confidence Limits	
		Value	Upper	Lower	Value	Upper	Lower
Availability	1	.897	.903	.891	.997	.999	.993
	2	.997	.998	.996	1.000	1.000	.999
	3	1.000	1.000	.9999	1.000	1.000	.999
	4	.999	.9995	.998	.998	.9998	.994
$\prod_{i=1}^{4} A_i$		.893	.899	.887	.995	.998	.990
Dependability	1	.932	.937	.927	.992	.997	.986
	2	.999	.999	.998	.999	1.000	.996
	3	.999	.999	.998	1.000	1.000	.999
	4	1.000	1.000	.9999	1.000	1.000	.999
	5	1.000	1.000	.9999	1.000	1.000	.999
	6	.993	.995	.991	1.000	1.000	.999
	7	.999	.999	.998	.999	1.000	.996
	8	.996	.997	.995	.998	.9998	.994
	9	.987	.989	.985	.997	.999	.993
	10	1.000	1.000	.9999	.999	1.000	.996
	11	.872	.878	.865	.996	.999	.991
	12	.991	.993	.989	1.000	1.000	.999
	13	.994	.995	.992	1.000	1.000	.999
	14	.998	.999	.997	1.000	1.000	.999
	15	1.000	1.000	.9999	1.000	1.000	.999
	16	.995	.996	.994	.998	.9998	.994
	17	.999	.999	.998	1.000	1.000	.999
	18	.993	.995	.991	.990	.995	.983
	19	.989	.991	.987	.993	.997	.987
	20	1.000	1.00	.9999	.999	1.000	.996
	21	.872	.878	.865	.995	.998	.990
$\prod_{i=1}^{21} D_i$		.662	.671	.653	.956	.968	.942

$$\text{Improvement in System Availability} = \prod_{i=1}^{4} \frac{A_i\ (\text{new})}{A_i\ (\text{old})} = 1.114 \begin{smallmatrix} +.008 \\ -.009 \end{smallmatrix}$$

$$\text{Improvement in System Dependability} = \prod_{i=1}^{20} \frac{D_i\ (\text{new})}{D_i\ (\text{old})} = 1.444 \begin{smallmatrix} +.027 \\ -.029 \end{smallmatrix}$$

Where $\prod_{i=1}^{n}$ means that the product of the n items which are subscripted i (that is, (N_i) equals 1.

at the bottom of the Table. It was recognized that the System's Capability was unaffected by the new assembly. Let us assume that the Capability of the System for the stated mission has been measured previously and found to be 0. 830. Also assume that the Availability and Dependability had also been previously measured and found to be $A_{old} = 0.768$ and $D_{old} = 0.597$. It can be seen from the numbers in Table 6-5 for the contribution to the Availability and Dependability using the old assembly that this assembly was a major cause of the low A and D of the system. Using the new assembly the A should be 0. 768 x 1. 114 or $A_{new} = 0.856 {}^{+.006}_{-.007}$ and D should be 0. 597 x 1. 444 or $D_{new} = 0.862 {}^{+.016}_{-.017}$. Since the MOE = A x D x C, the old $MOE_{old} = 0.381$ and using the new assembly $MOE_{new} = 0.612 {}^{+.012}_{-.013}$. In each case the value after the + and - are the approximate 95% confidence limits and not the standard deviation.

From this hypothetical example, several conclusions are apparent. One is that using the MOE's in this manual for simple tests is easy and straightforward. In fact, for large, complicated tests, the only difference is that the list of data elements becomes longer. The second conclusion is that all testers will always measure the same things when they have the same test item and scenario. Third, by reporting the values of the data elements, System A, System D, System C and System MOE for the specific scenario, the decision-maker can see the strong and weak points of the test item. Also, by comparing the test item to an alternate item, a point-by-point evaluation can be made at all decision levels. Lastly, the data elements for A and D are easy to obtain and are presently gathered from the Air Force standard maintenance data system.

12. CONFIDENCE LIMITS

Calculating the approximate confidence limits of a data element that results from the measurements of a single factor is rather straightforward (See., for example, ref. 12 and 13)., the calculation of an approximate confidence limit for a data set is rather difficult (See ref. 1-11) and the aid of the operations analyst should be sought.

Some References on Confidence Limits –

1. G. P. Steck, Upper Confidence Limits for the Failure Probability of Complex Networks, Sandia Laboratories, SC-4133 (TR), December 1957.
2. R. D. Halbgewachs, R. Mueller, F. W. Muller, Classical Upper Confidence Limits for the Failure Probability of Systems, Sandia Laboratories, SLA-73-0563, September 1973.
3. Epstein, B. "Estimation from Life Test Data" Technometrics Volume 2, No. 4, November 1960.
4. Springer, M. D. & Thompson, W. E. "Bayesian Confidence Limits for the Product of N Bionomial Parameters" Biometriks, Vol 53, Parts 3 and 4, 1966.
5. Springer, M. D. & Thompson, W. E. "Bayesian Confidence Limits for the Reliability of Cascade Exponential Subsystems", IEEE Transactions on Reliability, Vol R-16, September 1967.
6. Springer, M. D. & Thompson, W. E. "Bayesian Confidence Limits for Reliability of Redundant Systems when Tests are Terminated at First Failure", Technometrics, Vol 10 February, 1968.
7. Springer, M. D. & Thompson, W. E. "Bayesian Confidence Limits for System Reliability" Annals of Assurance Science, 1969.
8. Buehler, R. J. "Confidence Intervals for the Product of two Binomial Parameters" Journal of the American Statistical Association, December 1967.
9. Schick, G. "Bayesian Concepts for Reliability and Confidence" 1968 Annals of Assurance Sciences.

10. Rosenblatt, J. R. "Confidence Limits for the Reliability of Complete Systems" Statistical Theory of Reliability, University of Wisconsin Press 1963.

11. Madansky, A. "Approximate Confidence Limits for the Reliability of Series and Parallel Systems" RM-2552 Technometrics, November 1965.

12. Owen, D. B. HANDBOOK OF STATISTICAL TABLES, Addison-Wesley Publishing Co., Inc., Reading, Mass., 1962. PP 286-303.

13. Dixon, W. J. & Massey Jr., F. J. INTRODUCTION TO STATISTICAL ANALYSIS McGraw-Hill Book Co., New York, 1969. PP 501-504.

CHECKLIST FOR TESTER IN MOE AREA

1. Define the item under test and items with which it interoperates.

   a. test set(s)

   b. AGE

   c. power

   d. other subsystems/systems

   e. facilities

2. Determine Level of the Test

   a. Multisystem

   b. System

   c. Subsystem

   d. Set

   Note: The tester should go no lower than a set due to interoperability effects.

3. Determine Grouping to Which Test Item Belongs

   a. Aircraft Systems

   b. Missile Systems

   c. Communications/Electronics Systems

   d. EW Systems

   Note: If test item is a Missile, CE, or EW subsystem of an aircraft, the tester must use both groupings. (See item 16)

4. Define the scenario

   a. mission

   b. personnel types and experience

   c. tactics for test item

d. environment (weather, electromagnetic, time of day, terrain, etc.)

e. second side system

f. second side tactics

g. second side personnel

h. mission rate

i. time sequence of events (mission profile)

5. Select Data Elements

   a. Pick all data elements that apply to test item type.

   b. Pick all data elements of items with which the test item interoperates if test item is a subsystem or set.

   c. Determine the values of all "specified times, levels, positions, etc."

6. Narrow Data Element List

   a. Delete data elements that are not applicable because of

      i. test objectives

      ii. scenario

      iii. already known

      iv. method of operation of test item

      v. tester assuming a value

   b. Justify each deletion

   c. Give source for values already known

   d. Give rationale for each value assumed

7. Set Pass/Fail Values for A, D, and C and Confidence Level

   a. Give source for each value

   b. Give rationale for each value

8. Set Confidence Level for Measured Data Elements

9. Determine A, D, and C for Alternate Item or Against Which Test Item Competes

   a. Show missions were the same

   b. Show that appropriate parts of the scenarios were the same

      i. second side the same

      ii. second side tactics the same

      iii. environment the same

      iv. operational parameters the same

      v. test systems correspond where appropriate

   c. Give values and confidence limits on alternate item's A, D, and C.

10. Determine data required to be collected to determine each data element

11. See that data requirements are incorporated into test design.

12. Calculate data elements, including confidence limits, from data collected during test for each scenario.

13. Calculate A, D, and C including confidence limits, from data elements for each scenario.

14. Calculate MOE, including confidence limits, for each scenario.

15. Incorporate data elements, A, D, C, and MOE and their confidence limits for each scenario into test report.

16. Not all types of items are covered in the list of systems. For those not covered, the tester must obtain approved data elements which can be combined in such a fashion as to yield A, D, C, and MOE for the test item.

Chapter 7

TEST DESIGN

1. INTRODUCTION

An operational test can be designed with varying degrees of sophistication. In a casual interpretation of the activity, there cannot be a test which is not designed in some way. The real question concerns how well a test is designed. When a Test Officer makes plans for collecting specific data (as contrasted with the prospect of operating a system for a specified period and then looking to see what data might be available), he has designed a test. He has not necessarily given any evidence of the usefulness of that design to the process of learning about the system under test, and he may not even know himself how he intends to gather the information (not the same as the raw data) he needs. Initially the tester may not even know what information he needs to answer the questions being posed. He is also usually faced with resource limitations and a degree of uncertainty about what significance the data he does collect will actually have. This chapter is provided to help the Test Officer understand the process of identifying specific needs for factual data and translating those specific needs into an efficient plan for conducting controlled trials during which valid data for drawing conclusions can be collected. Test design is the combined responsibility of the Test Officer and his supporting professional statistician. The statistician, of course, is the one who knows the science of test design and understands the application of the proper techniques to a given problem. The Test Officer can also play an important role in design of the optimal test to deal with a stated objective by gaining some familiarity with the types of tools statistical test designers use and with the elements of the test problem that are important to him. Then he, the Test Officer, can effectively bring his knowledge of the operational Air Force into the activity to insure that the test is realistic, relevant, and reasonable.

This is not the only way to present an overview of test design. The statistical aspects especially are organized by different persons according to different views of the science, with distinctions and generalizations made on the basis of different features. The approach of this document was chosen for accuracy, clarity, and meaning to the Air Force Operational Test and Evaluation Officer.

2. EVOLUTION OF A TEST DESIGN

Test design is the process of developing a scheme through which an experiment can be conducted and data obtained to facilitate examination and evaluation. This scheme (design) characterizes the test, which may include simulation, but does not provide all the information necessary to conduct the test. The measurements to be taken, the order in which they are to be taken, and the method by which they are to be taken are elements of test design; the means of funding the test, the offices to be coordinated with on the Test Plan, and the distribution of the Test Plan are not. An element of test design is one which could have bearing on the outcome of the test. It is usually something which, if changed itself, would change the very nature and usefulness of the test. One example is the test item; certainly if the test item is changed, the nature of the test will be different and the results would have a different meaning.

A test design is, then, a plan for selecting a sample such that the maximum amount of useful information can be derived with the minimum expenditure of resources. It is an aid to drawing valid inferences from data.

The development of a test design is as much an art as it is a science. It is a complex process in which certain steps can be accomplished systematically, but there is no single-thread, step-by-step time sequence, and some combinations of steps have to be repeated many times before the final design is considered to be satisfactory by Test Officer and statistician. In all probability it will be a long time before a truly mechanical system of test design can be developed because of the vast amounts of quantitative and qualitative information that must be considered simultaneously, along with management, operational and mathematical experience.

A discussion of the inputs to the test design process, the trade-offs that take place, and the constraints that may surface is presented to guide the Test Officer and point out the contributions he will have to make to the development of a useful test design. Reference to Figure 7-1 may be helpful.

The first requirement is that some organization be given to the thinking about variables (or sources of change) in the test. This means grouping them in categories related to

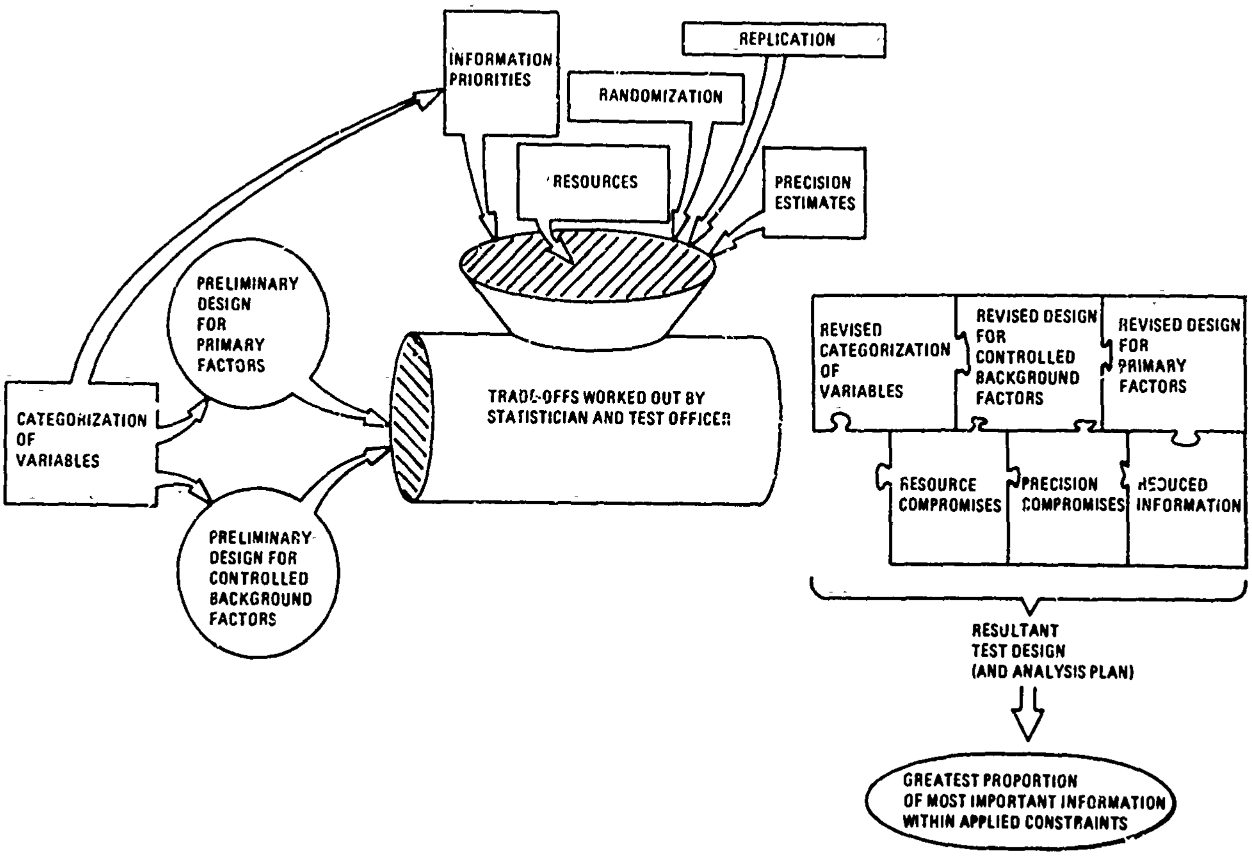


Figure 7-1. Test design evolution

different functions they will serve in the test. The initial categorization is not regarded as rigid because changing ideas and the need for compatibility with statistical methods and available resources may move the variables from one category to another.

From the grouping of variables for a test come lists of dependent variables (measures of test outcome), and primary factors, and controlled background factors (each having a potential effect on the test outcome).

Designs are generated -- at least in preliminary form -- for handling of the two most important groups of independent variables: (1) primary factors and (2) controlled background factors. These two designs are basically plans for control of the conditions under which observations will be made. Levels or settings of factors are chosen, plans are made for bringing together individual levels of the different factors in the right combinations, and means of controlling variability in the test results are explored.

The two preliminary designs (for primary factors and for controlled background factors) must be combined so that primary factor effects can be investigated under the proper background conditions and test results obtained which will be directly related to the test objectives. They are merged, with proper consideration to requirements for precision, in comparisons and estimates. This means repeating certain parts of the test (replication) to check trial-trial variability and randomizing the association of different factors at specific levels in order to avoid losing track of effects from two or more variables. The "sausage grinder" concept shown in Figure 7-1 is not too far removed from reality, for the method of arriving at a resolution for all the inevitable early conflicts will normally be relatively unstructured. The Test Officer and the statistician will be exercising ideas to see where changes and compromises can be accepted; if they cannot communicate and appreciate each other's requirements, either there will be complete dominance by one source of knowledge and ideas or else the resulting test design will be the least complicated one available.

Under the best of conditions the Test Director will probably come out with a test that does not provide all the information he had originally hoped for -- as symbolized by the piece missing from the puzzle in Figure 7-1. Variables may be moved from one category to another in an effort to accommodate conventional statistical test designs and limited resources; the pieces of the design and how they handle individual variables may be revised; the Test Officer may agree to accept somewhat less precision than he had hoped for; and if he is lucky he can talk someone into providing more money to support the test.

In Figure 7-1 the resultant test design is shown with an accompanying analysis plan. This does not imply that the analysis plan is derived easily, but rather that in the course of designing a test continual thought must be given to what will be done with the test data. This shows up in the designs for primary factors and controlled background factors, and the requirements for precision, and will drive the statistician's selection of techniques for employment in the design.

The next two sections (Sampling Experiments and Statistical Inference; Hypothesis Testing) discuss some of the statistical concepts and terminology the reader must be familiar with before he can appreciate the motivation for taking certain steps in test design.

3. SAMPLING EXPERIMENTS AND STATISTICAL INFERENCE

The task of operational testing is often formidable because it seems to require that every variation in the operational employment of a system (for an aircraft this includes airspeed profiles accompanied by three dimensional position profiles, target type and behavior, light and weather conditions, hardware performance variability, pilot performance variability, enemy defenses), be anticipated and studied so that sweeping generalizations about such descriptions of a system as operational effectiveness and operational suitability can be made. While some test objectives will be quite specific in focusing attention on a precise measurement of a particular characteristic, the evaluation of a system usually entails narrowing the number and range of variables under which the system will actually be tested and later drawing inferences about the performance to be expected under the full range of all variables. This means that the test designer must select realistic examples of missions or mission segments from the range of operations in which the system might be employed. In the case of effectiveness evaluations this may be facilitated by the use of the OT&E MOE data element diagrams. He must design the collection of data on the basis of the selected examples, and permit extrapolation to cover the missions and mission segments which could not be tested. Even with unlimited financial resources it would be physically impossible to test a system in all conceivable sets of circumstances to the extent that the exact system behavior could be said to be known with respect to all changes in all variables.

The problem is compounded by the fact that it is more difficult to characterize something with inherent variability, such as the performance of a system with human elements, than it is to measure a characteristic which does not change, such as the weight of a portable test item -- even though the weight measurement process may introduce slight variations in data from repeated weighings of the same article.

What the operational test designer must do, then, is judiciously choose the sample from which data will be collected so that he can maximize the useful information derived from any given expenditure of resources. This is analogous to the process the quality control engineer goes through in sampling a batch of items from the production line to determine whether the entire batch is to be accepted or rejected. In the quality control case, the sample is taken to be representative of the entire batch and the statistician refers to the batch as the population from which the sample has been (or will be) drawn. The operational tester's batch (called a population) is all (or a specified part of) the operational uses that will be made of the system. The characteristics of this population and the certainty of ever encountering any particular specifically-defined mission may seem rather vague, but missions are defined in official Air Force documents, operational envelopes are prescribed, enemy capabilities are estimated or known, and the Test Officer can contribute his experience and knowledge of the probability of experiencing various operational scenarios. Now the tester may only wish to draw conclusions and make predictions about selected missions or even system behavior in a limited portion of a selected mission type, but the population of inference (the population about which generalizations are to be made) must be sampled in order for supportable conclusions to be drawn about it. The operational tester's sample is composed of the test trials he has controlled and from which he has collected data.

Two types of inferences (conclusions) can be made. The first is a statistical inference, which is supported by the tools of the science of statistics and guided by rather rigid rules of the same science. The second type is non-statistical, being based on related experience and sound judgment. Where the statistical inference is drawn about all F-111 aircraft on the basis of tests conducted on randomly-selected F-111's, the non-statistical inference is drawn about all F-111 aircraft on the basis of experience with several F-111A's. A statistical sample must be drawn from the entire population of inference.

The basis for this statement is that statistics is the science of chance variations and in order to make fullest use of the tools of statistical inference there should be information available on the chance variations of the characteristics under investigation. For instance, if a test on a portable radio is conducted using one prototype radio, predictions about the characteristics of other radios are poorly supported. The tester has no information on the variability of communications system performance as different radios of that model are used. If two radios are used, there will be one source of data on radio-radio variability. As more radios are tested, the variability of the population of all radios of that type is more accurately characterized. Similarly, if only one talker is used in the test, inferences drawn to other talkers will be poorly supported (unless, of course, talker-talker variability with that particular radio system were already well-defined). The list of sources of variability goes on and on, even though they may not all be dealt with in any given test. If a single-sentence message is used repeatedly to the exclusion of other messages, the desired statistical population of inference is probably not well sampled. Another interesting source of variability is that arising from repetitions of a test on supposedly identical circumstances. This limits the precision with which any characteristic can be measured.

Holding operational variables constant for a test may indeed increase control over the test, but testing over a set of circumstances chosen to allow variability broadens the base to which test results are applicable -- in other words, broadens the population of (statistical) inference.

The use of a science of statistics to aid in analyzing test results is sometimes misunderstood. The tools of statistical inference should not be used in a purely mechanical way to make conclusive statements about a population from observations on a sample of that population. Consider a population frequency (of occurrence) distribution curve such as in figure 7-2.

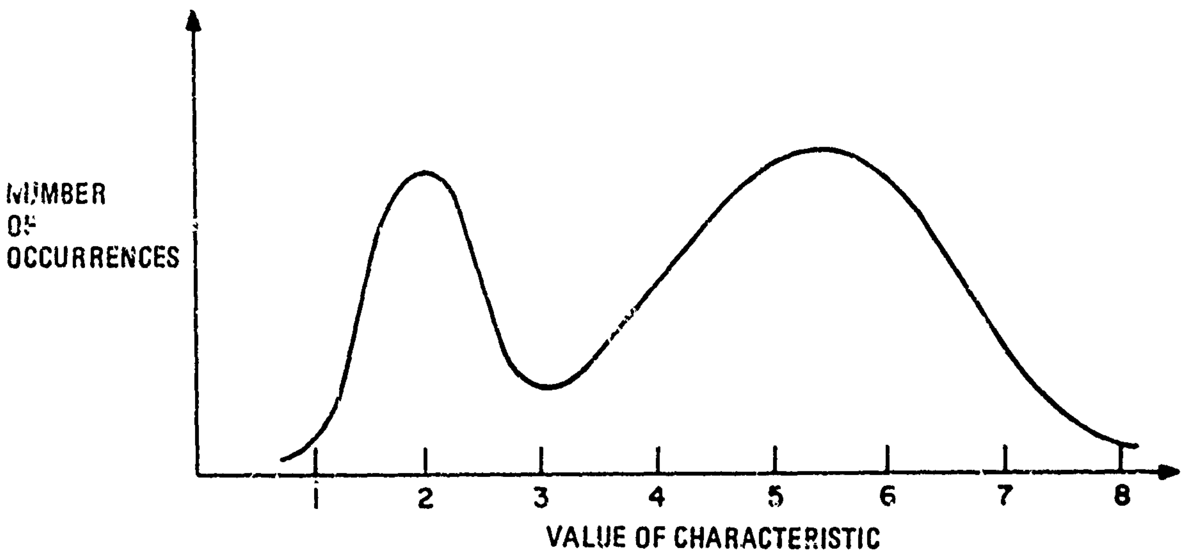


Figure 7-2. A frequency distribution plot

If one was to randomly sample from all the items in that population, he would expect to get more items with the value of the characteristic around 5 or 6 than with value around 3 or 4 just because the items with value 5 or 6 make up a greater proportion of those available to be drawn in the sample. If a bowl is filled with 300 white marbles and 100 red marbles, a person drawing marbles randomly from the bowl could be expected to get a white one three times as often as a red one. Similar statements can be made about items which are being characterized by the value of a continuous variable (such as length or velocity) rather than the value of a dichotomous (dividing into two parts, e.g., red or white) one.

The person drawing marbles might get exactly 30 white marbles in a sample of 40, but it would be no real surprise if he did not. If he drew several samples of size 40 and each sample contained exactly 30 white marbles, the observer would suspect a hoax, but the occurrence of white totals 25, 28, 30, 32, 31, 32, 29, 38, 30, etc. would not arouse any suspicion. Even the naive observer has some ideas about what to expect for a distribution of the sample statistics (measurable sample characteristics), or the sampling distribution for proportion white (red). A clustering about the known population proportion (3/4 white) with frequent occurrence of small deviations from that proportion and less frequent occurrence of the higher deviations is expected. In the continuous case, such as exhibited in Figure 7-2, the random sampling process would be expected to produce a greater number of items in the near-2 and 5-6 ranges than in the below-1 range in most samples — although perhaps not in every single one. The selection of one sample out of many with (say) five items with the value of the characteristic below 1 and two with value above 1 would not be totally unbelievable, although it would be unexpected.

On the basis of these common sense properties of sampling distributions and some theoretical calculations about sampling from known and unknown population frequency distributions, the tools of statistical inference have been created to predict the characteristics of a population based on the characteristics of items in a sample drawn from that population. As a simple example, if a sample drawn from the bowl of marbles had 3/4 white marbles, an observer who did not know the total percentage white might be willing to guess that it was also 3/4 white. <u>Inferential statistics</u> would tell the observer what chance he has of being wrong in mechanically predicting the population white/red ratio from the single sample drawn.

As suggested above, no single sample provides complete information about the population, but in a pretty good percentage of the cases it would be possible to draw sound conclusions about the population from that one sample. Statistical statements of probability will say how likely it is that the sample drawn is misleading in its makeup. The marble observer would know just how often to expect that a sample of 30 white and 10 red indicates a population ratio around 3:1 and how often it can occur with an "unlucky" sample from a population with ratio nowhere near 3:1.

It is more common to seek a statement of the probability of the population characteristic being within a certain range than fixed at a single value. A familiar table regarding the sampling distribution of proportions, for example, shows that in 19 cases out of 20 a sample of 30 white marbles and 10 red comes from a population with between 60% and 87% white.

The sample size is an essential part of the picture in drawing statistical inferences. One expects that the marble observer would be more inclined to bet a large sum of money on his prediction of the population white/red ratio if the sample showed 60 white and 20 red. The sampling distribution table gives a nineteen-in-twenty chance of being right in predicting a true white population between 20% and 99% if a sample of 3 white and 1 red is used for the prediction. There is the same chance of being right if a sample of 60 white and 20 red is used for prediction of a true white population between 65% and 84%.

Similar statements have been formulated for continuous variables. From the calculation of mean miss distance of a sample of air-to-ground missile firings, a statement of the expected mean miss distance for all missiles of that type or a statement of expected miss distance for the next missile firing could easily be derived.

Statistics can also be used to estimate the existence or non-existence of differences between samples (e.g., two types of missile characterized by some variable). It will not say whether or not there is an operationally significant difference, but if the tester can specify what the magnitude of an operationally significant difference is, statistics will help the tester decide if the samples from each missile type (say) show a real difference of the apparent magnitude or only the vagaries of the sampling of items which have variation in that characteristic. This type of statistical analysis is known as hypothesis testing. It has a vocabulary all its own and is encountered frequently enough that the tester should be familiar with this vocabulary.

4. HYPOTHESIS TESTING

Many studies of statistical data are placed in the form of a test which goes something like:

> I will assume that hypothesis A is true. If the sample data has the characteristics of data drawn randomly from a population in which A is true, the hypothesis will probably be true; otherwise A will probably be false.

The hypothesis cannot, of course, be proven absolutely true or absolutely false by sampling. Statistical methods based on the recognized properties of sampling distributions can be used to tell whether an apparent difference between characteristics of the hypothetical population and characteristics of the sample (which would be used to estimate its parent population characteristics) shows a statistically significant (i.e., unlikely to occur) deviation from the most likely sampling results.

Further discussion is conducted most easily in terms of problems in which the population frequency distribution is assumed to be of some particular shape. This field is known as parametric statistics. The extension to distribution-unknown (nonparametric) statistics is not difficult.

The ultimate uncertainty of hypothesis testing involved two types of error, referred to as α or Type I error and β or Type II error. Refer to Figure 7-3, which shows a sampling distribution for an arbitrary statistic b. If a test statistic falls in the shaded area, known as the critical region, the statistician may reject the original hypothesis that the population parameter is b_o on the belief that the test result is suitably improbable for parameter b. The size of the critical region can be adjusted at will: at a significance level of .05 the critical region includes 5% of the total area under the sampling frequency distribution curve; at a significance level of .01 the critical region covers 1%, etc. Rejection of the null (= no difference) hypothesis because a statistic falls in the critical region means that in addition to screening out statistics that apparently came from other sampling distributions (centered elsewhere), the test can reject the null

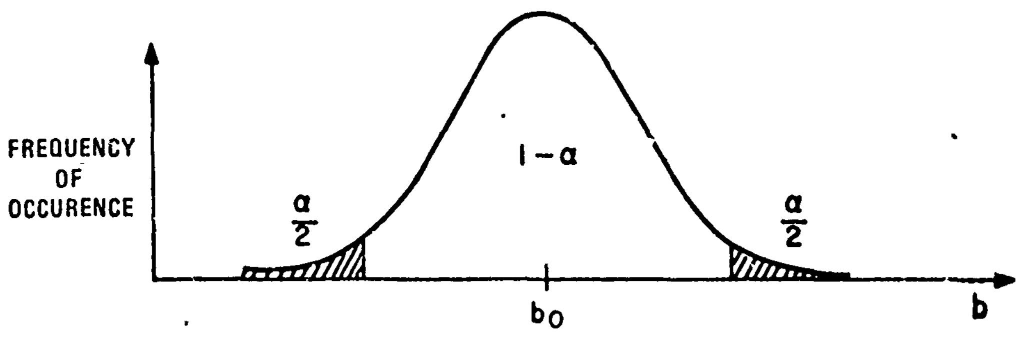


Figure 7-3. Sampling distribution for statistic b

hypothesis when it is true. If the significance level is α, this will happen $(100\alpha)\%$ of the time that the null hypothesis is true. It cannot be avoided. While the rejection of true hypotheses can be reduced by decreasing the significance level (reducing the critical area), the usefulness of the test in discriminating against all but the 100 $(1-\alpha)\%$ of the most probable sampling results is reduced. The significance level could even be reduced to zero, but then there would be no test to "pass" only the most probably true sampling results.

When the null hypothesis is false, β error may occur due to overlap of two candidate sampling distributions, the one for the null hypothesis and the true one (Figure 7-4). When the null hypothesis $(b = b_o)$ and the true $(b = b_1)$ sampling distributions overlap as shown, a statistical test with significance level α (shaded) will accept the false null hypothesis because the two values for b give the same sampling results to an extent measured by β, the dotted area. If a sample from the population with parameter b_1 gives a statistic in the dotted area [$(100\beta)\%$ of the time], it will also show in the acceptance $(1-\alpha)$ area of the sampling distribution for the hypothesized (but incorrect) population with parameter b_o. In a real experiment, only one population is sampled but the tester is forced to decide which population that is.

For the sampling distributions shown in Figure 7-4, each time α is decreased, β is automatically increased. The only way to decrease both α and β is to increase the sample size; this changes the shape of the sampling distributions as shown in Figure 7-5.

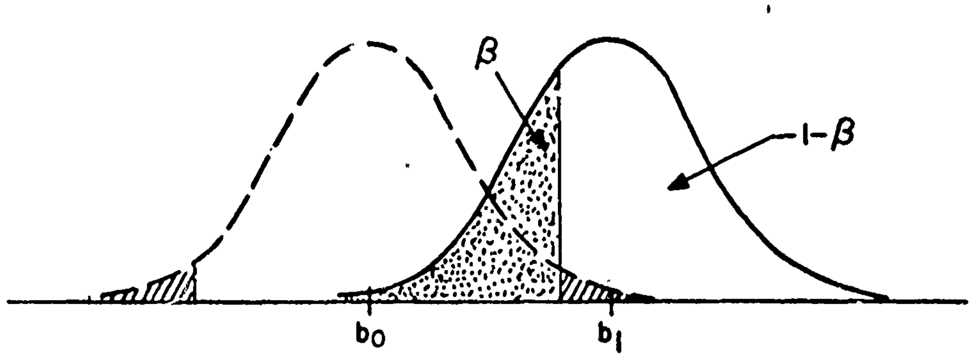


Figure 7-4. Sampling distribution for true and false b_1 and b_0

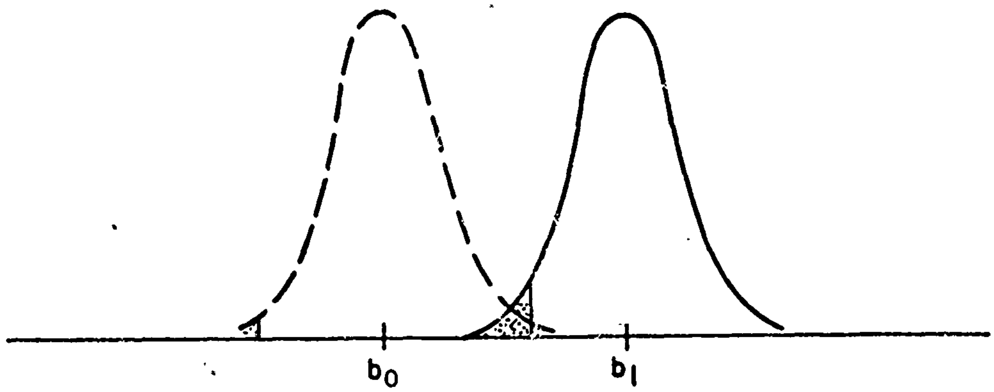


Figure 7-5. Sampling distributions for true and false hypotheses b_1 and b_0. sample size larger than in Figure 7-4.

The probabilities discussed here may be summarized in a matrix:

		HYPOTHESIS	
		ACTUALLY TRUE	ACTUALLY FALSE
DECISION	ACCEPT	$1-\alpha$	β
	REJECT	α	$1-\beta$

It can be seen that $1-\alpha$ is the sensitivity of a test -- in other words the capacity of the test to recognize a true hypothesis as such. Correspondingly, $1-\beta$ is the selectivity (known by statisticians as power) of a test -- the capacity of the test to recognize a false hypothesis as false.

5. CLARIFICATION OF TEST OBJECTIVES

The first step required in development of a test design is a detailed analysis of the test objectives and identification of needs for specific factual information. In some cases the raw data which must be collected will be reasonably obvious; in other cases the objectives may be very broad and not initially stated in terms of measurable characteristics of the system.

Certain information needed for designing a test may not be immediately obvious from inspection of the objectives written into the Test Directive. This is the stage when planning for the accomplishment of a useful test requires that questions and information requirements be translated into a scheme for the collection of specific data under carefully controlled conditions. It is likely to be a time when the spectrum of missions in which a system may be used operationally and the list of all variables in the employment of a piece of hardware in any one of those missions must be reduced to the number that can be meaningfully investigated in the limited number of trials that a test organization has the time and other resources to conduct and analyze. Definition of a few goals for this process of sorting out the problem may make the task easier and more fruitful:

<u>Define the Item to be Tested.</u>

If for no other reason than to keep a test from growing unmanageably large, there must be bounds on the item or system which is being evaluated. This is not a reference to the operational context or scenario in which some item or system is being evaluated, but to the object of the evaluation itself. The tester may ask whether it is one piece of hardware, two alternative pieces of hardware, tactics, a procedure, an organization, or something else. Is the concern with a navigation system, an aircraft, a missile, or an aircraft/navigation system/missile combination?

<u>Define/Refine the Objectives.</u>

The word "objective" can really be redefined at each step in a test depending on whether the goals are long-range or short-range, detailed or general. Therefore, the directed test objectives may not be suitably specific, data-oriented, and, as a group, thorough, to permit easy transition to identifiable data requirements. The test designer must analyze each of the directed test objectives in full consideration of available documentation on system mission and employment procedures (such as the ROC, DCP, and PMD and Concept of Operations documents for new hardware) to translate the existing objectives into statements that are detailed and clear enough to define data requirements. This process should also be used to limit the amount of data required. Not only must the tester ask what is actually needed to answer any question posed, but he must also decide how he will know when he has an answer to that question. Plans for conduct of the test proper may best be made for a number of objectives or subobjectives individually, but before the design can be finalized there should be an attempt to merge the requirements for information related to all objectives. This will minimize duplication of testing and allow the collection of more data for a given expenditure of resources.

Define the Population of Inference.

The operational scenarios or context referred to above is central to the deliberations of this step. The basic problem is to determine what population of operational employments of a system it is necessary to draw conclusions about. Since the test may be viewed as a sampling from all operational employments of the system (perhaps all of a certain type), the population of (statistical) inference will be that population which is sampled. Is it necessary to draw conclusions about the aircraft when any (properly qualified) pilot flies it? If one pilot does all the flight testing, the tester has no measure of the difference another pilot might make. If two pilots are used, at least one bit of information on the pilot-to-pilot variability is generated. Of course more are better. Does the mission of the aircraft include day/night and/or all-weather operation? Will different ground crews be used? Will different flight profiles be flown? All of these matters and more go into defining or selecting the population of inference. This is another step in bounding the test. If the scope looks excessive when compared with the initial concepts of time, money, and other resource allotments, it may be necessary to limit that portion of the test in which statistical inferences will be required and settle for nonstatistical judgments for the process of extrapolation to some untested situations. It may even be helpful to define a population of statistical inference and a population of non-statistical inference. (The latter will be larger.) The statistician may refer to these as the "sampled population" and the "target population," respectively. In this document the use of "population of inference" will mean the population of statistical inference or sampled population.

In defining the population of inference the tester has necessarily addressed the subject of variability in operational employment of the test item or system and the consequent variability in any single measure of the system's effectiveness and suitability. Now a more thorough investigation of the sources of variability must be made so the tester can decide how to handle each of these operational variables in a test.

6. SELECTION AND CATEGORIZATION OF TEST VARIABLES

Reduced to its simplest terms, a test consists of making observations on certain variables while others are in some way controlled. In operational testing these variables are all identified within the framework of an operational employment of the test item but the list of operational variables seems endless. In a missile delivery accuracy test, the list of variables includes the launch platform type, tail number and maintenance record, the crew, the target size, shape, and contrast, the time of day, relative position of the target at launch, velocity of launch platform, etc., etc. In a comparison of two missiles the missile type or mod is itself a variable. And of course the indicators by which the test outcome is measured and judged such as miss distance or percent hits are variables. Variables in the anticipated operational employment of a system are not necessarily equivalent to the variables in a test. In operational employment, knowledge of the system and concepts for its employment can be continually upgraded; an operational test takes place in a comparatively short time. Operational employment will often occur in several locations throughout the world; operational testing is usually conducted at a single base. The more of these sources of variability a test can be designed to account for, the better and more generally useful the test will be. Sometimes these variables simply cannot all be investigated in a single program and the population of inference is necessarily limited.

It will aid the test officer in organizing his thoughts and help the statistician in his construction of an applicable test design if the variables associated with a test can be categorized with regard to the role they play in operational employment of a system and with regard to the role they will play in a given test. These categories are not absolute. Operational employment variables exist and cannot be willed away or created, although they may be controlled. Operational test variables can in a sense be chosen. Some can be "excluded" by limiting the scope of a test (e.g., to a single air base). Others can be given different roles in different tests (in missile firing, time-to-target-lock-on might be either a measure of the test outcome or simply another variable which affects the test outcome.) The thinking process by which the test officer recognizes a variable as part of the operational employment scenario may give a clue as to which functional category it belongs in. The variables of a test will be categorized here as either independent or dependent and then the independent variables will be broken out into subcategories depending on the way they will be handled in the test. Figure 7-6 shows a hierarchy of the terms that will be used. The names for the types of variables have been chosen for their relevance to a Test Director's view of the test, but a one-to-one correspondence with the statistician's groupings has been preserved.

Define the Dependent Variables.

These are the measures of the test outcome. They are the means by which the results of the test will be judged, the basis on which competing systems are judged, or the variables the Air Force wants to be able to predict in operational employment. They are the variables one expects to respond differently because of variations introduced elsewhere. Some of these dependent variables will be dictated by standard MOE data elements and others will be suggested by system requirement and development

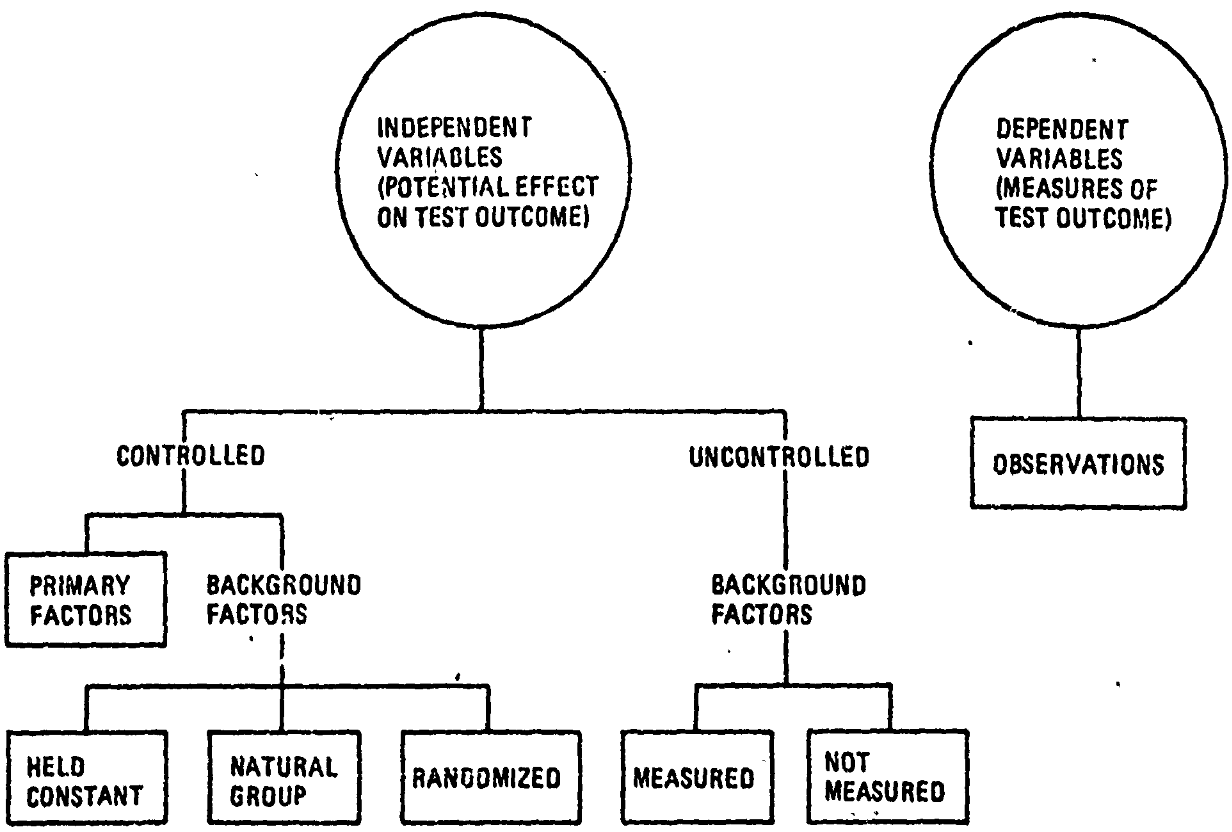


Figure 7-6. Categorization of test variables

documentation (ROC, DCP, PMD, etc.); others will not be defined, and a choice among several candidates may be necessary to select those that will be observed in a given test. In a missile test for example, does the tester want to observe the time between launch and impact, time between launch and detonation, miss distance, missile flight path? Or all? Or something else? All dependent variables will be defined in terms of response in a well-described scenario.

Define the Independent Variables.

These variables are the ones that seem important in determining the test outcome. If their importance to the outcome in the scenario(s) described is unknown after preliminary research, a couple of avenues are open. One is to test over multiple levels (settings, values, types, etc.) in the test. Another is to perform a sensitivity analysis, which is a mathematical study of the effects of changing levels of a variable (or variables) on the test outcome, as defined by the dependent variables. Some variables in an operational scenario can be changed quite freely without having any operationally significant impact on the test results; of these some are obvious while others are not. In a test of missile delivery accuracy, it is quite clear that aircraft tire pressure does not have an operationally significant impact on the test outcome. It is not so clear (except from operational experience) whether dive angle at the time of launch has a significant effect on the results. In order for mathematical sensitivity analyses to be conducted, the behavior of the system must be expressible in analytical (mathematical) form.

Independent variables may be either controlled or uncontrolled.

Controlled Independent Variables. For at least two reasons, the test designer will want to control some of the independent variables during a test. The obvious reason is that some variables identify alternatives that are being tested or investigated and compared; it would be necessary to control these independent variables in order to give any meaning to the test outcome. Other independent variables will be controlled to extend this interpretation of the test outcome by being able to give definition to the population of inference.

The controlled independent variables will now be discussed in two categories corresponding to the above reasons. They will be known as the primary factors and the controlled background factors, respectively.

a. Primary Factors. These are the factors to be compared in terms of their effect on the test outcome (as measured by the dependent variables). They are any of the sets (and there may be many overlapping ones) of alternatives which are under investigation because it is important to know the differences these alternatives make. Assessment of the effect of different lighting conditions would be important to the tester of an optically-guided missile so he would know if the missile could be employed more effectively under one condition than another. In a comparison of two radios — either two candidates for purchase or a candidate for purchase and a radio in the inventory — the radio itself would be a primary factor within the scenario used. Identification of a variable as a primary factor for the test implies either (1) that the variable can be controlled in an operational employment (i.e., the population of inference) and that control is desirable, or (2) that system deployment can and will be controlled. The range at which jamming is initiated can certainly be controlled at will. Aircraft tail number might be controlled by choosing only to fly certain aircraft, but this would be neither desirable nor probable. The aircraft type could be controlled by a decision not to employ the particular missile on a type of aircraft which tends to degrade the missile capability. Degree of pilot proficiency is not likely to be controlled (within certain limits) in day-to-day operational employment. The converse of this requirement is not true: the fact that a variable is controllable in the operational scenario does not mean that it will be among the primary factors in any given test, but only that it is a candidate for inclusion as a primary factor. Resource limitations will most likely be the constraining influence, and the tester will not be afforded the luxury of learning the detailed importance of all the candidate primary factors.

In a non-comparative test, where the objective is only to estimate population parameters, the distinction between primary factors and controlled background factors (discussion following) is unimportant.

b. Controlled Background Factors. There may be other independent variables which have not been listed as primary factors because (a) they are not of immediate interest for making any comparisons, or (b) they will not be controlled in the population of inference, or (c) they will be held at only one level in setting in the population of inference, or (d) they cannot be thoroughly investigated due to resource limitations. The tester's most immediate need is to control the effect which he suspects or knows is operationally significant. Three unique types of control are possible:

(1) Holding Constant. Use of different bases may have an effect on the take-off properties of an aircraft (different altitudes requiring different runway lengths, or rough runways encouraging a quicker lift-off) but because of the advantage of doing all of a test at one location, those variables deriving from the selection of different air bases might be held constant by definition when the test planners decide to conduct the entire OT&E from one base. This could be rationalized on the basis of past experience with several different air bases and a firm belief that the results from a test in one location could be extrapolated to predict performance at other places (nonstatistical inference).

(2) Defining Natural Groups of Trials. Rather than hold a variable fixed at one level (setting, value, choice) it may seem more appropriate to let it show up in the test results at more than one level but still with strict control. It would seem that a better test would be conducted if several radio operators were to use each of the candidates for purchase being compared. The population of inference could be expanded by incorporating information on operator-operator variability. The tool for bringing radio operators into the test picture is called the block

(and multiple sets of blocks called squares and cubes). The trials conducted with operator A participation would be carefully planned and would be compared with each other in a natural group. All of operator B's trials would be compared within another group and so forth. Then results of each of these natural group comparisons would be analyzed together. (In an actual statistical analysis these steps which are discussed as distinct and sequential are not so obvious.) By putting each level of the primary factor (or combination of levels of several primary factors) in a block an equal number of times, the effect of the level of the controlled background factor that defines the block (in this case operator A or B) weights the test outcome equally on all sides of the comparison. A disadvantage to the grouping of trials in this way is that analyses usually consider the outcome of each within-group comparison to be equally important, where the operational employment of a system may not justify it. Consider, for example, the effect of cloud cover on the delivery accuracy of an optically-guided (target contrast) missile. Suppose one natural group was defined by tests fired under a clear sky, a second was defined by 50% cloud cover, and a third was defined by completely overcast conditions. Is each of these conditions expected with equal frequency in operational employment? Overcast skies would be expected frequently in Southeast Asia and only occasionally in the Middle East, so the equal weight of each condition might be justified in a peacetime IOT&E and unjustified in an IOT&E being conducted when a conflict had started and a very specific use for the missile was envisioned.

(3) Incorporated at Random Levels. A means of controlling a background variable at several levels which is less precise (especially in small tests) but expected to be more representative of the frequency of occurrence of each level in operational employment (especially in large tests) is randomization. This topic is discussed later in more

detail. For the purpose at hand, the technique may be viewed as randomly associating the levels or settings of different factors.

Uncontrolled Independent Variables. These variables are background factors that change as a test is being conducted, although not in accordance with any test plan. Examples are ambient temperature, fatigue, and wear on mechanical parts of hardware items. Some of these may be measured and records can be compiled so that if the test results are not fully understood, these data can be studied in hopes of finding the complete answer. Other uncontrolled background factors may not be measured because the test designer chooses not to or, more significantly, because he does not know they are changing (e.g., an intermittent electrical connection) or because he does not appreciate the fact that the changes have any effect on the test outcome. (Note that the "change" is being used to indicate a condition that is not constant from trial to trial as well as one which varies with a single trial.) Uncontrolled but measurable independent variables may be brought under control, at least to some degree, by controlling the conduct of the test itself. Temperature, for example, could be controlled in effect by deciding at what time of day different trials will be run. The definition of this category, however, is intended to include only those variables that the tester does not intend to control, yet wishes to retain a precautionary record of. These are known to the statistician as concomitant observations.

It should be obvious that the "rules" discussed for deciding the category of each variable do not reflect inherent properties or qualities of the variables. Instead, each variable is pigeonholed according to the way it will be used in the test under consideration. As shown with the time-to-target-lock-on example, even the dependent and independent variables cannot always be sorted out, except in the context of a particular test objective. It becomes very important that the selection and categorization of test variables be given more than casual attention because the tester should be fully aware of the way in which all of the operational variables are handled in his test. The sources of information that can be exploited as the test designer tries to decide how a variable should be handled are summarized in Figure 7-7.

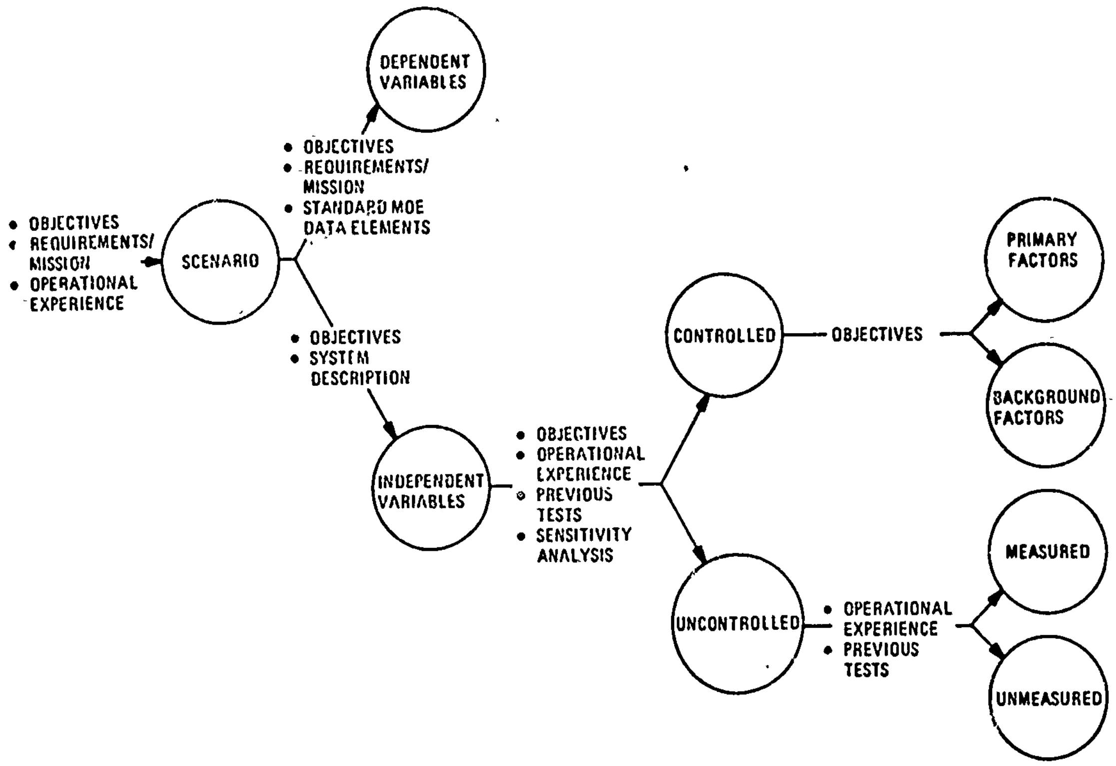


Figure 7-7. Sources of information used in identifying and categorizing variables

Although the names that have been given to the types of variables above were chosen to have a more or less apparent association with the role they will play in a test, there is a one-to-one correspondence with the variable categories that are used by the statistical test designer. The work of the statistical test designer is concerned primarily with the controlled independent variables. The primary factors correspond to his "treatment factors" for which the treatment design is developed, and the controlled background factors are handled in the units design. (A unit corresponds to the operational tester's trial.)

After making some decisions about the way in which the operational variables will become a part of the test, the tester must be more specific and decide how many different levels of any given variable will actually be tested and what those levels will be.

7. CHOOSING LEVELS OF CONTROLLED INDEPENDENT VARIABLES

The alternatives possible because of the non-constancy of a scenario with regard to some identified independent variable are defined by the "levels" of that variable. The levels may in fact be ordered on a scale, as would be airspeeds of 300, 450, and 600 knots. It may not be possible to scale levels of other variables. Except for the fact that some analysis techniques cannot be applied, the handling of these latter variables is no different. In an ECM test, jamming on or jamming off might be the levels of the variable "jamming mode." Variations in terrain (defined by irregularity, integrity, or foliage cover) might be levels of a variable. Different persons can be levels of the variable "pilot." Two radio, each produced by different manufacturers could be levels of the variable radio. Different times of the day or days of the week could be levels of a variable.

It will be noticed that all of these statements sound uncertain because they are expressed in terms of "might be", "can be", "could be," etc. In a sense all of these are operational variables simply because they can be identified. Whether they have any operational significance may be another matter (and of course operational significance depends on the measured characteristic under consideration). The tester must decide (a) that the variation in these characteristics may have operational significance and (b) that the objectives of the test warrant its consideration as a test variable. The tested levels of the variable will frequently not begin to exhaust the operationally-experienced levels of the same variable. The tester must decide, then, what range of levels to include in the test and which specific levels to choose within that range. Neither of these decision processes is likely to be completely arbitrary; resource limitations may control the number of levels on which data may be taken and test objectives may specify the range of levels to be investigated.

Range.

One philosophy is that a system should be tested to extremes (whether temperature, altitude, turn radius, recycling rate, etc.) in order to find the maximum operating envelope and identify the strong features as well as the weak links in the system. This presumes a fair degree of confidence in suitable system performance under "mild" conditions, and so is more likely to be the chosen approach in FOT&E than in

IOT&E. A pre-DSARC operational test will normally be designed to investigate the most probable sets of operational conditions, but any chance to extend the tested range of a variable will enlarge the population of statistical inference. If, for example, the populations of typical operational personnel and typical operational environments are not sampled, they should not be included in the population of statistical inference. The risk of losing or at least severely damaging an expensive piece of prototype hardware when less than the most qualified operator/maintenance personnel are in control has been questioned. The alternative is usually posed as the loss of an opportunity to get information on combat degradation of system performance, but with a little ingenuity other possibilities may be found. If scope displays or instrument faces can be recorded on film or magnetic tape, the displays can be reproduced for simulated real time tests involving any number of typical operational personnel. In some situations the better-qualified personnel can maintain loose control over hardware by over-the-shoulder participation in maintenance or operation. When a situation appears to be handled incorrectly (in maintenance) or to get out of hand (in operation) the better-qualified person can intervene to put the test back within nominal bounds. Of course, proper pre-test training of any operator and maintenance personnel reduces the risk of wasting test effort (at the very least) when the system is not used as it would be in operational employment.

Specific Levels.

There can be several criteria for choosing which levels to test at in some (now) limited spectrum of possibilities. Frequent occurrence in operational employment is desirable; there is no point in testing situations which can be expected infrequently before those which are expected frequently. The statistician may want evenly-spaced intervals between levels of a quantitative variable to simplify the analysis. It may be expensive or difficult to test at some levels (e.g., if certain hardware items are more readily available). Some levels may be crucial to the outcome of the test (e.g., the full range of interest should be covered). There may be greater uncertainty in some areas of knowledge. Perhaps system modeling can be or has been conducted to identify critical areas of knowledge, from which assumptions about the areas (levels) not tested can be made with minimal risk.

The levels first selected should be regarded as tentative unless it is directed that they be a part of the test. As an effort is made to arrive at a design that is consistent with the requirements of the statistical analyst as well as with the requirements of the information-seeker and the ability to conduct particular trials, some compromises involving the total number of levels for a given variable and the specific levels that are chosen may be made. The first possibility of conflict shows up in the next steps — selection of preliminary designs for the primary factors and the controlled background factors. Before these steps are explained, however, the nature of the test designer's thinking will be shown in two examples.

8. EXAMPLES OF TEST DESIGN PROBLEMS

Identifying Test Variables

This example will illustrate the type of reasoning a test designer has to pursue in defining the nature of a test. Because it is an illustration of that process only, it can (and will be) terminated before the test or the data analysis is completely characterized. The problem is to compare two aircraft in a fly-off. The basis for comparison will be some measure of dispersion in bomb delivery.

Do we drop the bombs in the manual or the automatic release mode? Since each could be quite aircraft-dependent, it would probably be best to use both. Now the trials have to be run four ways (2 aircraft types x 2 delivery modes).

How many aircraft of each type should be used? One thought is that by using only one of each type, the variability in test results can be held to a minimum and the comparison of aircraft types will be more straightforward. Restricting variability isn't always a good idea, however. If the operational employment of different aircraft (different tail numbers) is going to give some variability in bomb delivery dispersion, the test results should reflect this. Aircraft type A number 001 might deliver bombs with less dispersion than aircraft type B number 011, but aircraft A number 002 might deliver bombs with greater dispersion than aircraft B number 011. If all the operational testing were done with A(001) and B(011) the tester might come to the conclusion that A is a better aircraft than B. This is risky unless the tester knows that the variability in results from different aircraft (different tail numbers) of a single type is less than the "averaged" difference between the two aircraft types. The tester should also decide whether he needs an estimate of the tail number-tail number variability, even though he may think it is small. We decide that we are not primarily interested in a measurement of tail number-tail number variability, but that we will need information on this kind of variability in order to make a more conclusive statement about differences between aircraft types. Three prototype models of each aircraft should be available, so we plan to use all three.

Now we should decide whether we are comparing A(001) with B(011), A(002) with B(012), and A(003) with B(013) or something else. That's not really the object of this test. The comparison we are interested in is between type A and type B, not between individual aircraft. Thus, although we recognize that the only way to observe differences between aircraft types is through testing of individual aircraft, it will not be necessary to pit a specific aircraft of type A against a specific aircraft of type B.

Can we do the automatic bombing from tail numbers 001, 002, 011, and 013, and the manual bombing from tail numbers 003 and 012? Possibly, but it would detract from the test. In this approach the tester would not have the opportunity to obtain much data on tail number-tail number variability for a single bomb release mode. If the object is to compare some sort of "average" (over modes) bombing capability, this approach might be useful, but it is probably desirable to compare manual release bombing with manual release bombing and automatic with automatic. We need the data on each release mode from each aircraft of both types. That makes 24 (2 aircraft types x 3 tail numbers x 2 modes) unique trials. A bit of additional caution is in order. The object of the test is not to compare manual release bombing with automatic release bombing; it is still to compare aircraft type A with aircraft type B and we have chosen to do this by investigating both manual and automatic modes. The distinction will be important to the statistician.

Is the test to be run with a single pilot? That would seem to make the comparison truly between aircraft and not between pilots. It would not give any information on pilot-pilot variability though, and we recognize that it is conceivable for one pilot to get better scores with aircraft A while another gets better scores in aircraft B, especially in the manual release mode. So there will be more than one pilot. Two? Three? Six? Eighteen? The knowledge of pilot-pilot variability increases with each additional participant, and the inference drawn to these aircraft types, without regard to who the pilot is, will be better based if a larger number of pilots can be tested. Six sounds like a reasonable compromise. That would allow the tester to get two different pilots in each individual aircraft, thereby running a check on pilot-pilot differences for the six individual aircraft. Each pilot should fly each type of aircraft, however, to insure that the test is not run with three good pilots in one aircraft type and three poor pilots in the other. You don't solve this problem by running the test with nothing but good pilots either. There are bound to be some differences, even between two "good" pilots, and the tester does want some experience with a range of pilot abilities. If there are not six pilots available who have been checked out in both aircraft types, a compromise on the quality of the test will, of course, have to be made and the test run some other way. Here we assume that the six are available.

The test can be represented in a matrix as Figure 7-8 shows.

TAIL NO. LAST DIGIT	PILOT	A		B	
		MAN	AUTO	MAN	AUTO
1	1				
	2				
2	3				
	4				
3	5				
	6				

Figure 7-8. Trials matrix

Trials will be conducted to provide data for each of the cells in the matrix. A minimum of 24 trials are necessary now. It would improve the test if we could run more than 1 trial for each of the cells. After all, one measured miss distance (i.e., one bomb drop) does not tell much about the trial-trial variability under a fixed set of circumstances. If the tester knows what kind of variability there is among trials conducted under a single set of circumstances, he will be in a better position to tell if the variability observed between trials conducted under two different sets of circumstances is really due to the characteristics separating those sets of circumstances. Duplicating each trial once doubles the total, however.

It would be interesting and helpful to do the same testing for a variety of combinations of different airspeeds, dive angles, and release altitudes. This might reveal greater or smaller differences between the two aircraft that would not be detected in a test conducted for a single airspeed/dive angle/release altitude. Testing at two dive angles would double the number of trials; testing at two dive angles, two airspeeds, and two release altitudes would require eight times the original number of trials (original number x 2 x 2 x 2).

The designer must also make decisions about the order of trials. Should Aircraft A fly before Aircraft B? Should manual release trials be run before automatic release trials? Should one pilot fly before another? If the only comparison to be made is between aircraft types, the important order questions are those involving the sequence of Aircraft A trials vs. Aircraft B trials. Aircraft A trials should not all be conducted before Aircraft B trials are started, because of possible pilot learning effects. It is not obvious that alternating flights A-B-A-B-A-B-A etc. would be a mistake — but this might mean that all A flights were in the early morning and all B flights were in mid-afternoon. No pilot should repeatedly fly one aircraft type before the other; this could introduce the effect of individual pilot learning. If comparisons between aircraft types are to be made separately for manual release and automatic release, trials should not be conducted so that manual release is accomplished before automatic release on one type and automatic before manual on the other. It is conceivable that the later release is always more accurate and that this would generate a preference for one type in the manual mode and for the other type in the automatic mode — especially if releases in both modes are made on a single sortie.

Order among the six pilots is inconsequential. Similarly, order among the individual aircraft (tail number) is unimportant for this test.

The next example illustrates a different problem — that of establishing control over a test that has already been defined in some nominal way.

GIVEN: 3 MISSILES OF EACH MOD AVAILABLE
3 AIRCRAFT AVAILABLE
EACH AIRCRAFT FIRES ALL ITS MISSILES IN ORDER SHOWN

1 FACTOR OF DIRECT INTEREST (MISSILE)
0 CONTROLLED BACKGROUND FACTORS*

POSSIBLE GROUPING:

A_1: M_1 M_1 M_1
A_2: M_2 M_2 M_2
A_3: M_3 M_3 M_3

POSSIBLE FIRING SEQUENCE:

M_1 M_1 M_1 M_2 M_2 M_2 M_3 M_3 M_3

A_1 A_1 A_1 A_2 A_2 A_2 A_3 A_3 A_3

SEQUENCE-CORRELATED EFFECTS:
LEARNING (M_3 FAVORED)
CHANGING WEATHER (M_1 OR M_2 OR M_3)

RANDOMIZE FIRING ORDER AMONG AIRCRAFT:

M_2 M_3 M_3 M_1 M_2 M_2 M_3 M_1

A_2 A_3 A_3 A_1 A_2 A_2 A_3 A_1

COMPLETELY RANDOMIZED DESIGN

DO NOT COUNT OVERALL SEQUENCE

1 FACTOR OF DIRECT INTEREST (MISSILE)
1 CONTROLLED BACKGROUND FACTOR (AIRCRAFT)

GROUPING:

A_1: M_1 M_2 M_3
A_2: M_1 M_2 M_3
A_3: M_1 M_2 M_3

POSSIBLE FIRING SEQUENCE:

M_1 M_1 M_1 M_2 M_2 M_2 M_3 M_3 M_3

A_1 A_2 A_3 A_1 A_2 A_3 A_1 A_2 A_3

SEQUENCE-CORRELATED EFFECTS:
LEARNING (M_3 FAVORED)
CHANGING WEATHER (M_1 OR M_2 OR M_3)
FATIGUE (M_1)

RANDOMIZE FIRING ORDER WITHIN BLOCKS:

A_1: M_1 M_2 M_3
A_2: M_3 M_1 M_2
A_3: M_2 M_3 M_1

POSSIBLE FIRING SEQUENCE:

M_1 M_2 M_3 M_3 M_1 M_2 M_2 M_3 M_1

A_1 A_1 A_1 A_2 A_2 A_2 A_3 A_3 A_3

RANDOMIZED BLOCK DESIGN

1 FACTOR OF DIRECT INTEREST (MISSI
2 CONTROLLED BACKGROUND FACTO
(AIRCRAFT AND LIGHT LEVEL)

GROUPING:

	L_1	L_2	L_3
A_1	M_1	M_2	M_3
A_2	M_3	M_1	M_2
A_3	M_2	M_3	M_1

POSSIBLE FIRING SEQUENCE:

M_1 M_2 M_3 M_3 M_1 M_2 M_2 M_3 M_1

A_1 A_1 A_1 A_2 A_2 A_2 A_3 A_3 A_3

L_1 L_2 L_3 L_1 L_2 L_3 L_1 L_2 L_3

SEQUENCE-CORRELATED EFFECTS:
NONE - ALREADY RANDOMIZED WI
RESPECT TO MISSILE TYPE

LATIN SQUARE DESIGN

ES OF EACH MOD AVAILABLE
AFT AVAILABLE
RCRAFT FIRES ALL ITS MISSILES IN ORDER SHOWN

FACTOR OF DIRECT INTEREST
ISSILE)
CONTROLLED BACKGROUND
ACTOR (AIRCRAFT)

ROUPING:

A_1:	M_1 M_2 M_3
A_2:	M_1 M_2 M_3
A_3:	M_1 M_2 M_3

OSSIBLE FIRING SEQUENCE:

M_1 M_1 M_1 M_2 M_2 M_2 M_3 M_3 M_3

A_1 A_2 A_3 A_1 A_2 A_3 A_1 A_2 A_3

EQUENCE-CORRELATED EFFECTS:
LEARNING (M_3 FAVORED)
CHANGING WEATHER (M_1 OR M_2 OR M_3)
FATIGUE (M_1)

ANDOMIZE FIRING ORDER WITHIN
LOCKS:

A_1:	M_1 M_2 M_3
A_2:	M_3 M_1 M_2
A_3:	M_2 M_3 M_1

OSSIBLE FIRING SEQUENCE:

M_1 M_2 M_3 M_3 M_1 M_2 M_2 M_3 M_1

A_1 A_1 A_1 A_2 A_2 A_2 A_3 A_3 A_3

RANDOMIZED BLOCK DESIGN

1 FACTOR OF DIRECT INTEREST (MISSILE)
2 CONTROLLED BACKGROUND FACTORS
(AIRCRAFT AND LIGHT LEVEL)

GROUPING:

	L_1	L_2	L_3
A_1	M_1	M_2	M_3
A_2	M_3	M_1	M_2
A_3	M_2	M_3	M_1

POSSIBLE FIRING SEQUENCE:

M_1 M_2 M_3 M_3 M_1 M_2 M_2 M_3 M_1

A_1 A_1 A_1 A_2 A_2 A_2 A_3 A_3 A_3

L_1 L_2 L_3 L_1 L_2 L_3 L_1 L_2 L_3

SEQUENCE-CORRELATED EFFECTS:
NONE - ALREADY RANDOMIZED WITH
RESPECT TO MISSILE TYPE

LATIN SQUARE DESIGN

6 MISSILES OF EACH MOD AVAILABLE

2 PRIMARY FACTORS (MISSILE AND TARGET)
1 CONTROLLED BACKGROUND FACTOR (AIRCRAFT)

GROUPING AS FOR NONFACTORIAL RANDOMIZED BLOCKS

SEQUENCE-CORRELATED EFFECTS POSSIBLE
UNTIL BLOCKS ARE RANDOMIZED.

RANDOMIZED BLOCKS:

A_1:	M_1 T_1	M_2 T_2	M_3 T_1	M_1 T_2	M_2 T_1	M_3 T_2
A_2:	M_3 T_1	M_2 T_2	M_1 T_2	M_1 T_1	M_3 T_2	M_2 T_1
A_3:	M_3 T_2	M_1 T_2	M_2 T_1	M_3 T_1	M_2 T_2	M_1 T_1

FACTORIAL ARRANGEMENT IN RANDOMIZED
BLOCK DESIGN

Figure 7-9. Steps in bringing experiment under control comparing 3 missile mods

Controlling a Test

By means of a simple example it can be shown how the design for controlled background factors and the process of randomization in bringing primary factors and background factors together act to bring an experiment or test under control. The application of different grouping designs will be illustrated for a single primary factors design; then the impact of changing the primary factors design will be illustrated. Figure 7-9 summarizes the example that will be used. Some of the terminology used has not been explained yet, but the reader should be able to follow the discussion nevertheless.

Consider a comparative evaluation of three mods of an air-to-ground missile in which the criterion for judgment is target miss distance. At the start it is assumed that there are certain constraints on the test: a total of 9 missiles — three of each mod — are available; three launch aircraft are available; and each aircraft will fire its load of three missiles in the order indicated by a left-to-right reading of any sequence shown. The first two constraints could reasonably be expected in an operational test; the third helps keep the example simple.

This is a typical nonfactorial (Section 9 of this chapter) experiment. There is only one primary factor — missile mod — and the comparisons are being made between the levels of that one factor — designated according to mod M_1, M_2, and M_3.

The naive tester may not do a great deal of planning at all. The missiles are loaded on aircraft as they arrive — three M_1's on A_1, three M_2's on A_2, and three M_3's on A_3 — and fired in that order (Figures 7-10 and 7-11).

Whatever the outcome of this test, the results could be hopelessly confused by the existence of effects that change over the nine missile firings. Suppose, for example, that the weather was bad in the middle of the test: M_2 performance would probably be degraded. Or suppose that the pilot of aircraft A_1 finds that he has to fire the missiles later than the instrumentation suggests in order to get good scores. He wouldn't learn this without firing one or two of the M_1 missiles, but he might report this fact to the pilots A_2 and A_3 and then they could take advantage of this increased knowledge

A_1:	M_1	M_1	M_1
A_2:	M_2	M_2	M_2
A_3:	M_3	M_3	M_3

Figure 7-10. Possible combination of aircraft and missiles

M_1	M_1	M_1	M_2	M_2	M_2	M_3	M_3	M_3
A_1	A_1	A_1	A_2	A_2	A_2	A_3	A_3	A_3

Figure 7-11. Possible firing sequence

for all three of their firings. The pilot of A_3 might also learn from A_2's experience. A better design must be found.

These sequence-correlated effects can be distributed over the firings of all other mods by randomizing the order in which missiles are fired. This is shown in Figure 7-12. The tester has completely randomized the missile mods with respect to the only significant variable he has identified: sequence. This is a Completely Randomized Design (section 10 of this chapter).

M_1	M_2	M_3	M_3	M_1	M_2	M_2	M_3	M_1
A_1	A_2	A_3	A_3	A_1	A_2	A_2	A_3	A_1

Figure 7-12. Randomized firing sequence

Notice, however, that M_1 is always fired from A_1, that M_2 is always fired from A_2, and that M_3 is always fired from A_3. The tester cannot be sure whether observed differences in results for missile mods are really due to differences in missiles or whether they are due to differences between aircraft. Nor can he tell if a failure to show differences means that there are no missile differences or that real missile differences are being canceled out by equally real aircraft differences.

This can be controlled by grouping the missile firings according to aircraft number and insuring that each missile mod is fired from each of the aircraft. This grouping is shown in Figure 7-13.

There could once again be sequence effects (See Figure 7-14), so the tester randomizes the firing order within each block as shown in Figure 7-15.

This is known as a Randomized Block Design (section 10). A single background factor has been controlled by use of a grouping design.

A_1:	M_1	M_2	M_3
A_2:	M_1	M_2	M_3
A_3:	M_1	M_2	M_3

Figure 7-13. Grouping to combine each mod with each aircraft

M_1	M_1	M_1	M_2	M_2	M_2	M_3	M_3	M_3
A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_1	A_3

Figure 7-14. Possible firing sequence

A_1:	M_1	M_2	M_3
A_2:	M_3	M_1	M_2
A_3:	M_2	M_3	M_1

Figure 7-15. Randomized Block Design

The tester may additionally suspect that there is an effect on missile guidance caused by the changing of the ambient light level from bright sun to cloud-obscured sunlight. He therefore chooses to identify three light levels and set up testing blocks defined by those light levels so that each missile can be tested at each light level. This simultaneous grouping by two variables is shown in Figure 7-16. Now each missile mod has been fired in each light level an equal number of times. The design used to control the two background factors by grouping is called a Latin Square (section 10).

	L_1	L_2	L_3
A_1:	$\ddot{M}_1$	$\ddot{M}_2$	$\ddot{M}_3$
A_2:	M_3	M_1	M_2
A_3:	M_2	M_3	M_1

Figure 7-16. Latin Square Design

This increase in control has been realized without firing any additional missiles. There has been some tradeoff, however, because the tester has forfeited any knowledge of the trial-to-trial variability in test results for a given set of circumstances (combination of factor levels). When distinctions were not drawn between trials in the Completely Randomized Design there were 3 trials to be compared for a study of "within missile mod" variability. In the Randomized Block Design shown, M_1 only gets fired from A_1 once, and similarly with M_2 and M_3. M_1, M_2, and M_3 do all get fired from A_1, however. In the Latin Square Design this equivalence no longer exists. The combinations of levels of the two background factors which appear in combination with mod M_1 never appear in exactly the same way for tests of M_2 or M_3.

Now the changes brought about by the introduction of another primary factor will be shown. Perhaps the tester wants to compare the three missile mods still, but in addition he wants to find out whether they all can be delivered with equal accuracy against two types of target. The combinations of primary factor levels to be compared are indicated by checks in the Factorial Arrangement (section 9) matrix of Figure 7-17.

	M_1	M_2	M_3
T_1	x	x	x
T_2	x	x	x

Figure 7-17. Factorial Arrangement for Primary Factors

Now each of the combinations of missile mod and target type is considered equally, just as the missile mods were before. These six combinations may be put into a

Completely Randomized Design or into aircraft-defined Randomized Blocks of six trials each or into a Latin Square Design utilizing six aircraft and six light levels. See Figure 7-18. Increasing the number of combinations of primary factors rapidly increases the number of trials required.

A_1:	M_1	M_2	M_3	M_1	M_2	M_3
	T_1	T_2	T_1	T_2	T_1	T_2
A_2:	M_3	M_2	M_1	M_1	M_3	M_2
	T_1	T_2	T_2	T_1	T_2	T_1
A_3:	M_3	M_1	M_2	M_3	M_2	M_1
	T_2	T_2	T_1	T_1	T_2	T_1

Figure 7-18. 3x2 Factorial in Randomized Blocks

This has been a simple example, but it begins to show the criteria (number of primary factors, number of levels for each primary factor, number of controlled background factors, and number of trials per block, etc.) by which a test designer is moved from consideration of one design to another. Now a discussion of the two parts of the overall design (primary factors and controlled background factors) and of randomization can be given. An early discussion of the types of problems addressed in a test design should have established a frame of reference for the next three sections.

9. SELECTION OF THE DESIGN FOR PRIMARY FACTORS

Together with the response, or dependent variables, the primary factors define the heart of the statistically designed experiment. If a test objective is to compare the delivery accuracies of two missiles, a simple outline of the experiment might be: "Compare the miss distances realized when the two missiles are fired under similar circumstances". If a test objective is to find out whether a new mod makes an operationally significant improvement in the effectiveness of a deployed jamming pod, a simple experiment outline might be: "Compare the effectiveness of the ALQ-299 Mod 1 in jamming the SUR SAM-guidance radar (as measured by missile-target miss distance) with the effectiveness of the ALQ-299 Mod 0 in performing the same function. An operationally significant improvement is defined as a decrease by 10% in the total proportion of missiles guided to within a lethal radius of the target." If a test is required to find out whether Level 4 maintenance crews can really diagnose the cause of malfunctions more quickly than Level 3 maintenance crews, a simple outline might be: "Compare the average time to locate faulty components in the preamp circuit of the PXR-2". In each case none of the background factors have been identified. The primary factors and levels essential to the definition of the test are two types of missile, two ALQ mods, and two proficiency levels, respectively. To work out initial intentions for the primary factors design it is not necessary to decide under what conditions (i.e., levels of controlled background factors) the comparisons will actually be made. It will probably be true that the two designs -- primary factors and controlled background factors -- will be derived in parallel, and some changes in categorization may be made.

An outline of the basic alternatives in the design for primary factors is given in Figure 7-19. The two types of designs for comparative experiments are in the lower left and lower right corners (factorial and nonfactorial), while the other possibilities show that the object of the test will be to estimate population characteristics rather than to make comparisons.

The design of a non-comparative experiment consists of holding certain factors constant while others vary at random; although the design work may be difficult, the

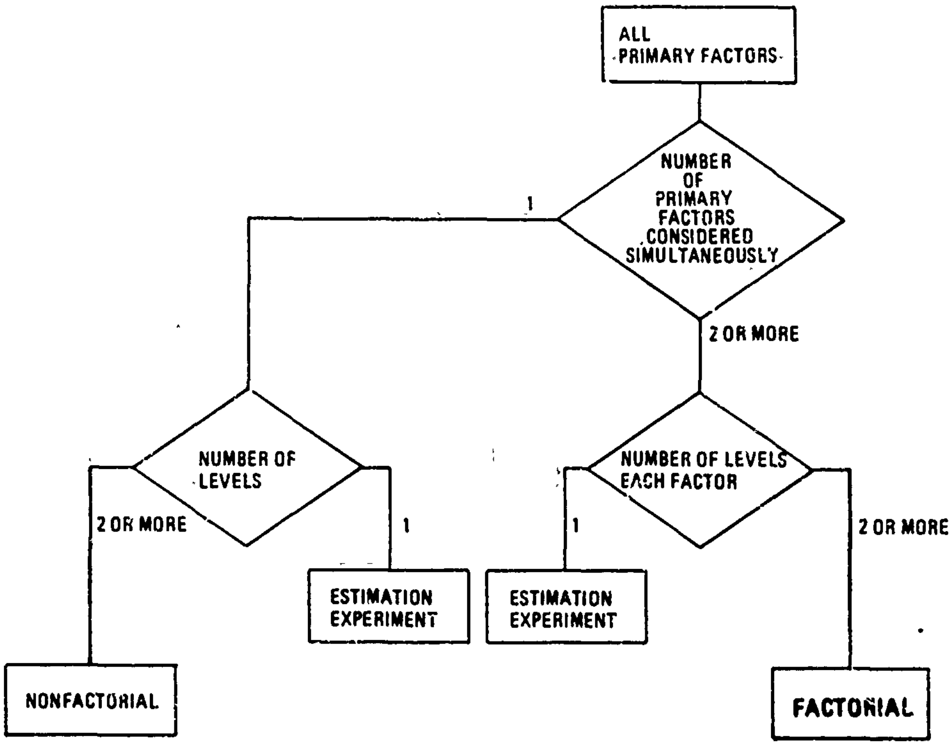


Figure 7-19. Design for primary factors

problems encountered are not hard to visualize. The discussion following will be aimed at the comparative experiment.

The primary factors design for an experiment with a single primary factor is straightforward. The only requirement is to specify the levels of that factor at which observations will be made. The arrangement is called nonfactorial, for reasons that will be apparent shortly, and with it the tester makes a series of observations throughout which the level of that factor is changed repeatedly to provide for observations at each level. A tester could, for example, be comparing two types of missile, two ALQ mods, or two proficiency levels. There may be repetitions of the entire set of trials under similar conditions (replication) to get data on trial-trial variability. The ordering of the trials is to be decided when the primary factors design is merged with the controlled background factors design, but it is apparent from the start that some sort of randomization will be required.

The alternative to the nonfactorial arrangement is called a factorial arrangement; it is required whenever multiple levels of multiple primary factors are being investigated in the same experiment. The objective of an experiment might be to study the effectiveness of two different air-ground missiles against four target types (e. g., brick building, SAM launch installation, and moving and stationary tanks) on two types of terrain (e. g., target fully exposed and target surrounded by dense foliage) and provide data to be used in a decision on which missile to use under each of the eight sets of circumstances. Figure 7-20 below shows the combinations of factor levels at which observations will be made.

TERRAIN	TARGET	MISSILE A	MISSILE B
EXPOSED TARGET	BRICK BUILDING	x	x
	SAM INSTALLATION	x	x
	MOVING TANK	x	x
	STATIONARY TANK	x	x
PROTECTED TARGET	BRICK BUILDING	x	x
	SAM INSTALLATION	x	x
	MOVING TANK	x	x
	STATIONARY TANK	x	x

Figure 7-20. Complete factorial arrangement

Because of the number of levels for each factor, this is referred to as a 4x2x2 or $4x2^2$ factorial experiment. Each of the 16 combinations of a type of missile, a type of terrain, and a type of target will be tried under similar circumstances to make comparisons. The experiment in which each level of each primary factor is tested at each level of every other primary factor is said to have a complete factorial or full factorial arrangement. Sometimes this is shortened to factorial arrangement but there are also fractional factorial experiments.

A fractional factorial arrangement is laid out on paper much as the complete factorial is, but observations are made at only some (carefully chosen) factor level combinations (figure 7-21).

TERRAIN	TARGET	MISSILE A	MISSILE B
EXPOSED TARGET	BRICK BUILDING	x	
	SAM INSTALLATION		x
	MOVING TANK	x	
	STATIONARY TANK		x
PROTECTED TARGET	BRICK BUILDING		x
	SAM INSTALLATION	x	
	MOVING TANK		x
	STATIONARY TANK	x	

Figure 7-21. Fractional factorial arrangement

A statistician refers to this as a fractional replication (in this example, 1/2) of the complete factorial arrangement. The fractional factorial arrangement has significant limitations in that it does not provide data on all interactions. In spite of the fact that each missile is fired at each type of target, that each missile is fired in each type of terrain, and that each target appears in each type of terrain, missile A never gets fired at an exposed SAM installation and so the firing of missile A at a protected SAM installation never gets compared with the firing of the same missile at the same target in an exposed environment. For the same reason, missile A never gets compared with missile B when both are fired at the same target on the same terrain. Overall, the effects can be averaged, but if missile A is much better against all fixed targets and against a moving target in an open field but loses its track on a moving target in a cluttered terrain the tester will never know. This would be important if missile B performs admirably against moving targets in heavy ground clutter.

The foregoing is an example of a three-factor interaction. Main effects which are the changes in the mean observation (e.g., reduction in miss distance) from level to level of one primary factor (e.g., missile type), averaged over the levels of the other primary factors (e.g., terrain type, target type), can be observed in fractional factorial experiments. Interaction effects are the unequal changes in the mean observation from level to level of one primary factor when at different levels of other primary factors, are in general not separable in a fractional factorial experiment. Some fractional factorial arrangements permit separation of selected interaction effects. "Paper plans" for many fractional factorial arrangements are given in various handbooks along with a list of the effects which are separable with each design. The principal motivation for doing fractional factorial experiments is the reduction in the number of trials in which resources (time, money, missiles, etc.) need be expended.

Selected interactions can be separated in a fractional factorial arrangement by choosing observations that delineate a few interactions at the expense of confounding (essentially, hiding or mixing together) others. While confounding is usually regarded as undesirable, there are instances in which it is done intentionally.

The object is to confound high-order interactions (three or more factors involved) while increasing the trials that can be run for data on low-order (two-factor) interactions and main effects. Confounding may also be used in certain complete factorial experiments for the same purpose.

Factorial arrangements may be used for economy even where interactions are known to be insignificant as long as there are multiple levels of several factors to be investigated. Consider the nonfactorial experiment in which missile delivery accuracy is being compared for 2 targets, then for 2 terrain types, then for 2 countermeasure environments. One approach is to vary one factor at a time while holding everything else constant (Figure 7-22).

	TARGET 1	TARGET 2
1. TERRAIN 1 - CM 1	X	X
	TERRAIN 1	TERRAIN 2
2. TARGET 1 - CM 1	X	X
	CM 1	CM 2
3. TERRAIN 1 - TARGET 1	X	X

Figure 7-22. Part I of nonfactorial arrangement

This does not cover all possible combinations of factor levels, so next the tester initially sets everything at level 2 and again varies one factor at a time.

	TARGET 2	TARGET 1
1. TERRAIN 2 - CM 2	X	X
	TERRAIN 2	TERRAIN 1
2. TARGET 2 - CM 2	X	X
	CM 2	CM 1
3. TERRAIN 2 - TARGET 2	X	X

Figure 7-23. Part II of nonfactorial arrangement

Each factor has been varied, but only one has been selected for variation at a time and a total of 12 trials has been run (Figures 7-22 and 7-23) compared with the 8 required for the factorial arrangement of Figure 7-24.

		TARGET	
TERRAIN	CM	1	2
1	1	x	x
	2	x	x
2	1	x	x
	2	x	x

Figure 7-24. Factorial arrangement for economy

Each factorial level combination has been observed, but in fewer trials. Additionally, this has the advantage that the order of all combinations can be randomized at once rather than doing it piecemeal as might happen in the two-part nonfactorial experiment.

Statistical texts use different terminology to refer to what are called here factor levels and combinations of factor levels. Some books call each combination a treatment, while others call each factor level a treatment and each combination a treatment combination. The design for primary factors of this volume is often called the treatment design.

10. SELECTION OF THE DESIGN FOR (CONTROLLED) BACKGROUND FACTORS

Once the comparisons to be made in a test have been decided upon, additional steps can be taken to reduce the uncertainty in these comparisons. Some books refer to these as designs for the reduction of experimental error, but when the analyst makes no unrealistic assumptions about uniformity of the test data the problem is more correctly identified as one of excessive variability in test conditions and the consequent inability to connect the variability in test data with specific origins. Two basic approaches to bringing the test conditions under control will be described --holding (operational) variables constant in the test and randomizing the application of the different levels of a factor to the combinations of other test variables. A third approach, which consists of grouping trials to maximize homogeneity in the group with respect to a background variable, will also be described. It has aspects of both randomization and holding variables constant.

The basic unit of the actual data collection (in fact often called a unit) is the trial, which is defined as the smallest subdivision of a test in which a single measurement or observation is made. There should be a clear similarity or a purposeful and identifiable dissimilarity between the conditions which define different trials. A single sortie may correspond to a single trial or it may, in another test, support a number of trials such as multiple passes through a chaff cloud or multiple attempts to locate targets. This definition of trial insures that the results of individual trials are compared with the results of other individual trials, the results of groups of trials with the results of other groups of trials.

In holding a variable constant at one level throughout, the tester has indicated that he wishes to compare the results of each trial without any consideration of that variable. From one point of view that is fine because it focuses attention on the effect of changing the primary factor(s) alone. This is common in experiments designed to measure an effect when all other (potential) variables are normally constant, but the operational employment of Air Force systems typically involves simultaneous variations of many types (ranges of several threat variables, several physical environment variables, and several friendly hardware/personnel/procedure variables). It is often preferable to take observations of an effect over a range of levels for other variables to get an "averaged" result.

Simply "letting things happen" without exercising any control over a particular variable is probably the best handling of a variable that will typically be changing from operational employment to operational employment...............in a very large test. This means that if radio operators with a particular heavy accent make up 1% of the total Air Force population of radio operators, they will appear in about 1% of the trials, quite at random, over the long run. The operational tester selects a limited number of participants for a typical radio evaluation, however, and he could be caught in a biased (influenced in some direction) test if the radio operators are not representative of the entire population. Barring the possibility that the tester may have the time and money to study the entire population of Air Force radio operators and scientifically choose a sample representative of the entire population with regard to the proper variables, he can at least identify as many of the important variables as possible and choose a random sample in a way that includes the opportunity for variability with regard to these important factors. By random selection of individuals within the group to be sampled he can hope to get the right proportion of persons with each of the relevant characteristics. The smaller the test will be, the harder it is to both insure that a range of levels for different variables is tested and still select participants at random. What can be done in every case is to randomly associate the selected variable levels (in this example radio operators) with the selected combinations of (levels of) other factors in the test. At times the distinction between these two acts of randomization becomes blurred, as when a tester randomly chooses and introduces the ambient temperature to the test by mixing up the order in which trials will be run. Methods of randomization are described in section 11.

Another means of coping with variation in certain background factors is to hold a variable constant at one of a limited number of levels (quantitative or qualitative) and run several trials at each of those levels. The groups of trials will be defined by the one background factor level and will be constrained to interpretation as comparisons within one group (at least with regard to levels of that background factor). This amounts to holding a variable constant over part of the test, holding it constant at another level over another part of the test, and so forth. The combinations of primary factors being compared in the experiment (the statistician's "treatments") should each appear an equal number of times in each natural group. This causes the effect of that level of the defining background factor to appear with equal weight on all sides of whatever comparisons are being made. This is not always possible because the number of trials in any natural group is not always equal to a multiple of the number of combinations of factor levels being compared. There are designs to deal with this by running group-to-group controls, but they are not so efficient as when all comparisons can be made at least once within each group.

The statistical test designer's name for one of these natural groups is a block. A block design contains a number of blocks. If there is only one "block", the tester has decided to hold that block-defining variable constant throughout the experiment, and this is not the intent of block designs. Blocks are used when the tester either wants to or has to incorporate different levels of a variable. An example of the latter can be seen in the fact that a limited number of trials can be run on any one day; it might be useful to make different days into blocks of the experimental design and (try to) insure that all primary factor combinations are run each day.

It is common to randomize the order of appearance of the primary factor combinations within each block (in time sequence or position sequence or with respect to the appearance of some other factor) to thwart the occurrence of unexpected confounding within blocks. This randomized block design is a common design for the control of background factors. If no blocking is used, the design is probably a completely randomized design; this term was not brought out earlier because it is not so typically contrasted with an "incompletely randomized design" as with the introduction of blocks.

An example of the use of the completely randomized and the randomized block designs can be given by reference to the example in Figure 7-25. Here missile delivery accuracy was investigated in eight situations involving different targets, different terrain, and different countermeasure environments. A different way of drawing the factorial matrix, this time with each combination of factor levels identified by a letter is:

TERRAIN	CM	TARGET	COMBINATION OF PRIMARY FACTORS
1	1	1	A
		2	B
	2	1	C
		2	D
2	1	1	E
		2	F
	2	1	G
		2	H

Figure 7-25. Factorial arrangement

Suppose now that the tester has 24 missiles to be fired in three days. Then he might randomize the order of the 24 firings (assume 3 for each primary factor combination) and get them over with as quickly as possible (Figure 7-25).

ACHBEDG	AGFHDCEFFGH	EDCBAB
Day 1	Day 2	Day 3

Figure 7-26. Completely randomized design

If lucky he might finish by noon on the third day. But suppose that terrain 2 is the dense foliage giving some shelter to the targets and that the second day was completely overcast. Because 8 of the 12 terrain 2 shots take place on the second day, test results might show the foliage-surrounded targets much harder to hit than those on the (open) terrain 1. An observant analyst might notice the correlation, but that is about all he could do. He could not work backwards through the results to say what the true terrain 1 - terrain 2 difference is because of the unbalanced testing on the terrain 2.

Combination F isn't ever tried on days 1 or 3, and B is never tried on day 2. Maybe B would be highly successful on a cloudy day and F a dismal failure on a sunny day but the tester will never know.

Perhaps instead the tester fires the missiles as shown in Figure 7-27.

Day 1	AHCDEFBG
Day 2	HEDAGCBF
Day 3	BEDACFGH

Figure 7-27. Randomized Block Design

As a result of this design, the analyst and the tester know how much of the total variability can be accounted for by day-to-day variability. Results that previously looked like so much noise on top of the day-to-day differences now seem more significant. Without any particular order, the radial miss distances might be as shown in Figure 7-28.

31,	26,	39,	32,	27,	28,	32,	23,	20,	22,	41,	22
27,	29,	40,	31,	26,	29,	32,	29,	28,	32,	41,	28

Figure 7-28. Randomized data

When exhibited (and analyzed) in blocks they become as shown in Figure 7-29.

Day 1:	28, 31, 22, 28, 30, 31, 26, 27
Day 2:	40, 39, 32, 32, 32, 41, 41, 32
Day 3:	26, 28, 22, 23, 27, 29, 29, 32

Figure 7-29. Data of Figure 7-17 in Blocks

Further analysis (not obvious to the reader) would show the terrain 2 data to be consistently higher than the corresponding terrain 1 data.

The tester has not had to run any additional trials to separate out the day-to-day source of variability, but he has had to structure the test more precisely. In this hypothetical example, he was prohibited from completing the test in the shortest time possible but the scheduling impact may be insignificant. In other examples the blocks might not even involve time. Aircraft tail numbers, pilots, weather conditions, approach patterns, and any number of other test variables or combinations of variables might define blocks.

Simultaneous sets of blocks defined by two different variables can be incorporated in what is known as a Latin Square design. To simplify presentation of an example, an investigation of only four of the combinations of primary factors (designated A, B, C, D below) will be considered (no terrain effect investigation). Now, instead of the testing on three different days, a tester feels that the important effects will come from the use of four different pilots and four possible target approach patterns. He wants to group the trials simultaneously according to test pilot and approach pattern. This is shown in Figure 7-30.

PILOT	PATTERN 1	2	3	4
I	A	B	C	D
II	D	A	B	C
III	C	D	A	B
IV	B	C	D	A

Figure 7-30. Latin-Square design

Note that each combination of primary factors (A, B, C, D) is tested at each approach pattern and each combination of primary factors is tested with each pilot. Also of note is the fact that each pilot flies each approach pattern. A Latin Square may be viewed as two overlaid randomized block designs (although there are restrictions on the "randomization") with the blocks restricted to be of a certain size (Figure 7-31).

PILOT	COMBINATION OF PRIMARY FACTORS			
I:	A	B	C	D
II:	D	A	B	C
III:	C	D	A	B
IV:	B	C	D	A

PATTERN	1	2	3	4
COMBINATION OF PRIMARY FACTORS	A	B	C	D
	D	A	B	C
	C	D	A	B
	B	C	D	A

Figure 7-31. Latin Square resolved into two sets of blocks

The Latin Square design presupposes that there are no interaction effects between the primary factors and pilots or approach patterns. It is also restricted to situations where each of the three factors involved has an equal number of levels. If these conditions are met, however, the tester can use this design to establish precise control over two sources of uncertainty in the comparison of combinations of primary factors without being required to increase the total number of trials.

The example shows only one trial at each overall set of conditions, so in this sense the tester has had to give up something from the randomized block design of 4 trials per block x 4 blocks. That information lost is on trial-to-trial variability for the same set of conditions. If he requires the information for his analysis he can run a repetition of the Latin Square, but he might be able to use the variability among the 4 D trials (for example) to check for a statistically significant D effect.

There are extensions of the Latin Square idea to incorporate more than two sets of blocks simultaneously.

If the block size (i.e., number of trials per block) in either block or multiple sets of blocks (square) designs is not and cannot be made equal to the total number of combinations of primary factor levels being compared, the test designer will use other patterns. There are a number of different incomplete block designs for different purposes. The most commonly encountered are probably the balanced incomplete block designs and the Youden Square designs.

Balanced incomplete block designs have fewer trials per block than there are combinations of primary factors to be compared with the following three conditions satisfied.

1. Each block contains the same number of trials.
2. Each combination of primary factors occurs the same number of times altogether.
3. All specific pairs of combinations of primary factors occur together in all blocks the same number of times.

Figure 7-32 shows a balanced incomplete block design that might be used if only four of the trials of the example for Figure 7-25 could be run in any given day.

Day 1:	A	B	C	D
Day 2:	E	F	G	H
Day 3:	A	B	G	H
Day 4:	C	D	E	F
Day 5:	A	C	F	H
Day 6:	B	D	E	G
Day 7:	A	D	F	G
Day 8:	B	C	E	H
Day 9:	A	B	E	F
Day 10:	C	D	G	H
Day 11:	A	C	E	G
Day 12:	B	D	F	H
Day 13:	A	D	E	H
Day 14:	B	C	F	G

Figure 7-32. Balanced incomplete block design

In order to show the balanced nature of the design, no randomization has been given to the positions within blocks (whether they represent time, positions, aircraft, or something else) or to the order of the groups of four combinations as they are assigned to blocks. Both of these steps would be taken in applying this "paper plan" to the real situation. Incomplete block designs are less efficient than complete block designs, which is to say that it takes more trials to achieve the same precision in estimating primary factor effects. Furthermore, it takes more trials to get a balanced design (very helpful in analysis) with incomplete blocks than with complete blocks.

Youden squares can be viewed as a combination of a set of complete blocks and a set of incomplete blocks, put together the same way a Latin Square is, or as a Latin Square with one or more blocks (i.e., one or more levels of one of the background factors) missing. The number of combinations of primary factors and the number of blocks in one set are still equal and this number is the same as the block size (trials per block) for the other set of blocks. The conditions are still quite restrictive and Youden Squares suffer from the same drawbacks as incomplete block designs. For an example, see Figure 7-33.

APPROACH PATTERN

PILOT		1	2	3	4
	I	A	B	C	D
	II	D	A	B	C
	III	C	D	A	B

Figure 7-33. Youden Square

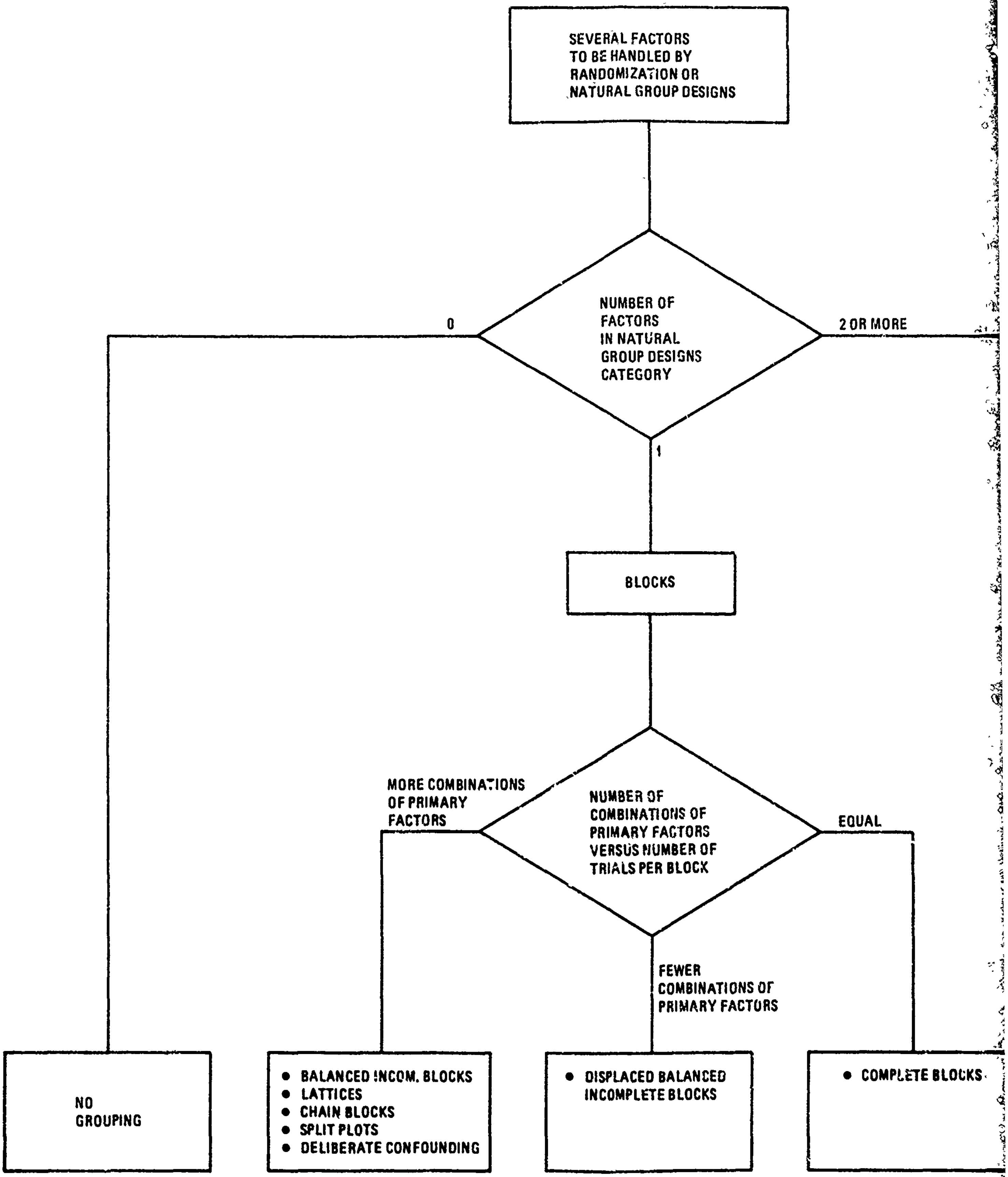
"Paper plans" for incomplete block designs and Youden Square designs are given in a number of handbooks. There are not always balanced designs available for arbitrary values of the number of combinations of primary factors and the number of levels (or block size) of each background factor.

A hierarchy of different grouping designs for the controlled background factors is diagrammed in Figure 7-34. Each of the diamonds shows the criterion for differentiating between the designs in the rectangles immediately below it. Although there is no single ordered scale on which to rank the different designs, it is true that in general the control over and knowledge of the test conditions increases as one moves from left to right and the precision with which trial-to-trial variability can be measured is increased. Note the relative locations of the completely randomized, randomized blocks, and Latin Square designs.

Also as one moves from left to right, the restrictions on the number of levels of each factor that are permissible and on the freedom in associating specific levels of different variables increase. At times this may be done without increasing the total number of trials, but when that happens the number of trials that can be compared at any given level of specification decreases. There are simply fewer trials which can be described as similar. This diagram is not intended to be exhaustive either in scope or in detail. It does cover the most frequently encountered designs and includes a few of the more specialized designs for purposes of orientation.

The Test Officer may encounter references to "one-way classifications," "two-way classifications," etc. These are statements of the number of sources of variation the experiment is designed to accommodate explicitly. Thus a two-factor factorial arrangement combined with a randomized block design for the background factors calls for a three-way classification analysis model.

The various blocking or grouping designs have an advantage over the randomization and holding constant methods of background factor control in that they allow a certain amount of variability with regard to the selected background variable(s), but the nature of the representation of the sampled population must be known before the experiment. Blocking and randomization probably in reality control several sources of variation in an operational test. Identification of a particular pilot as the defining variable level for a block also restricts level of training, age, and physical stature. Randomization of trials with respect to time may simultaneously effect randomization with respect to runway use, maintenance crew, batch of replacement parts, RF interference levels, and/or target used.



SEVERAL FACTORS TO BE HANDLED BY RANDOMIZATION OR NATURAL GROUP DESIGNS
NUMBER OF FACTORS IN NATURAL GROUP DESIGNS CATEGORY
0
2 OR MORE
1
BLOCKS
NUMBER OF COMBINATIONS OF PRIMARY FACTORS VERSUS NUMBER OF TRIALS PER BLOCK
MORE COMBINATIONS OF PRIMARY FACTORS
EQUAL
FEWER COMBINATIONS OF PRIMARY FACTORS
NO GROUPING
• BALANCED INCOM. BLOCKS
• LATTICES
• CHAIN BLOCKS
• SPLIT PLOTS
• DELIBERATE CONFOUNDING
• DISPLACED BALANCED INCOMPLETE BLOCKS
• COMPLETE BLOCKS

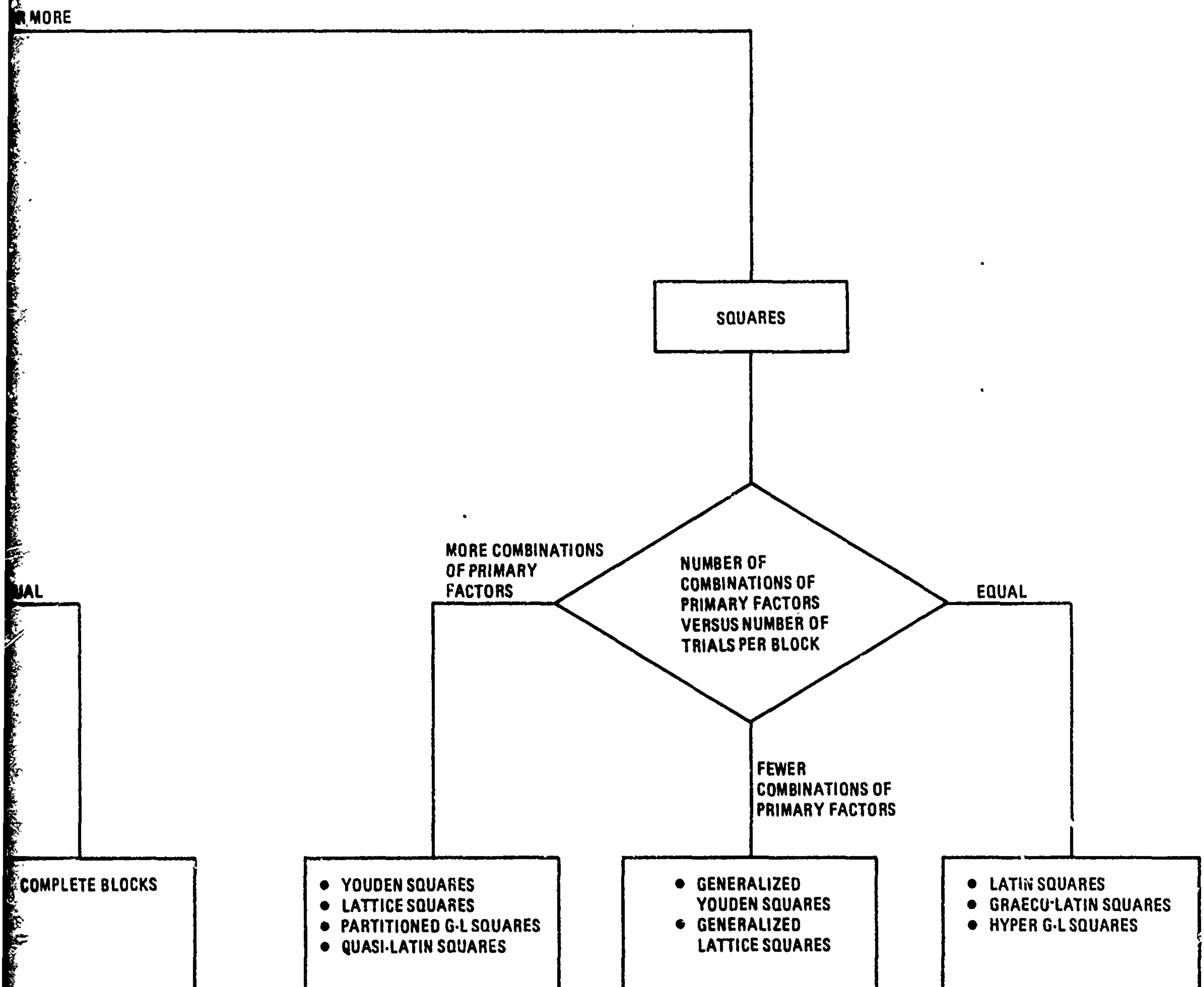


Figure 7-34. Design for background factors

11. RANDOMIZATION

Although a design for primary factors and a design for controlled background factors may have been found useful for the test at hand and compatible with each other, the method of combining the two is not fixed. Different possibilities for combining the levels of primary factors and background factors will elucidate some effects and bury (or "confound") others. The method of randomization will determine which effects show up in the test data and which are hopelessly inseparable -- whether real or not.

Randomization is the act of "mixing things up" so that the effect of certain identified influences (or potential influences) on experimental results are random and will not bring into question the extent of influence from the factors under investigation. It is accomplished by randomizing the selection of test items for observation or for influence by an experimental variable. This may mean randomization in a time sequence randomization in a position sequence, or more generally stated, randomization of the occurrence of levels of one factor with the levels of another.

If the influence of a factor is being investigated by application of that factor in the test at four different levels, say, the effect of that factor must not be confounded with the effect of some other factor by the regular occurrence of a fixed pairing of each level of one factor (say, A) with selected levels of another factor (say, B). Two examples will illustrate the problem.

TRIAL	LEVEL OF FACTOR A	LEVEL OF FACTOR B
1	1	3
2	4	2
3	3	1
4	4	2
5	3	1
6	2	4
7	1	3
8	2	4

Figure 7-35. Confounding of A and B effects

In Figure 7-35, note that each level of factor A is paired with one and only one level of factor B, even though the sequential appearance of each may appear quite thoroughly mixed. There is no way to tell whether, for example, the effect observed in trials 1 and 7 is from level 1 of factor A or from level 3 of factor B. Level 1 of factor A never appears in the test with any other level of factor B. A one-to-one correspondence is not necessary before this confounding of effects can take place. Consider the example of Figure 7-36.

TRIAL	LEVEL OF FACTOR C	LEVEL OF FACTOR D
1	1	2
2	2	1
3	3	2
4	2	1
5	1	2
6	3	2
7	3	2
8	1	2
9	2	1

Figure 7-36. Confounding of C and D effects

Here, level 2 of factor D is confounded with levels 1 and 3 of factor C while level 1 of factor D is confounded with level 2 of factor C. A better test would randomize the occurrence of the different levels of factor D with levels of factor C (Figure 7-37).

TRIAL	LEVEL OF FACTOR C	LEVEL OF FACTOR D
1	1	1
2	2	1
3	3	2
4	2	1
5	1	2
6	3	2
7	3	1
8	1	1
9	2	2

Figure 7-37. Confounding of C and D effects removed

Now each level of factor D appears to have no correlation with any one particular level of factor C.

Randomization at different levels or of different aspects of a test design can be useful. In the last example, for instance, the order of occurrence of each level of factor D for a given level of factor C might confound the test results by prohibiting separation of sequence-correlated effects. The example given is as well randomized with regard to sequence as could be expected for a small number of trials, but if, say, level 1 of factor D got tested before level 2 at each level of factor C, the results might be revealing an effect of that sequence rather than of factor D. Examples of sequence-correlated effects are learning processes, weather changes, and fatigue, as well as indirect effects like the intermittent failure of a rotary switch with two "on" positions every time it is in a particular one of those two "on" positions, or the use of different operating procedures by one crew that happens to appear regularly for every third trial.

Because so many variables get linked together in a complex operational test, intentional randomization with respect to one variable may effect randomization with respect to other variables even though the detailed origin of the effects cannot be traced. An individual pilot, for example, carries with him peculiarities involving physical height, level of training, preference for tactics, eyesight, reaction time, age, and crew rest. It will be up to the tester to anticipate whether any one of these variables could have a significant effect on the test outcome and should this be controlled individually, or whether randomization of pilot encounter with levels of some other variable will be sufficient control of all the pilot-associated variables.

Randomization is usually brought about by ordering observations, combinations of factor levels, etc. on the basis of tables of random numbers. Although most tables of random numbers are computer-generated, they come close enough to satisfying the definition of random numbers. (1. Any number in a specified range equally likely at any time; 2. All numbers in the specified range occurring with equal frequency; 3. No number having any influence on what the next number shall be) to be useful for most scientific experiments.

Occasionally, the results of a randomization process do not have the desired mixing effect, especially when a short series is sought. One nine-digit sequence of the numbers 1, 2, and 3, for instance, would be 3-3-3-2-2-2-1-1-1 and this sequence could definitely be derived from a long table of random numbers. Thus it is advisable to screen all random number sequences to identify those which do not appear useful and replace them with new sequences wh ich are more random in appearance. If the series 3-3-3-2-2-2-1-1-1 appeared somewhere in a sequence of 1000 random digits, it would not be significant, but when used by itself, it is not an effective tool to achieve the desired end.

Has a human influence been introducted by rejection of this series in favor of another? Not necessarily. There are many different random number series based on the same digits (the number possible depending on the number of choices possible for each position and the length of the series) and the rejection of one of these alternative possibilities as unsuitable will not affect the validity of the test design. Selection of the series to be used on the basis of observed characteristics of that series, however, might reflect some human bias and would be a poor technique. It should also be apparent that a mechanical screening of the series of random numbers would lead to an ever-present bias in the series that pass the screening.

12. REPLICATION

With an idea of the test conditions that will define the different trials, the tester must decide how many times he has to repeat certain test conditions before he can be sure of the true effect of those conditions. This is the problem of replication.

The only way a tester can learn how precisely he is estimating a parameter is to repeat the test. In chemical analysis this might mean repeating the measurement of pH by taking several samples from the same batch and performing the same chemical tests on each sample or even repeating the tests (if they were non-destructive) on the same sample. In operational testing the trials are often dynamic chains of events which cannot be held constant for repeated measurements and at times are impossible to record completely for reproduction and remeasurement. This is not of critical importance, however, because the trial-to-trial variation is usually much greater and of more interest than the measurement-to-measurement variation for a single trial. It is important that use of the word "replication" be reserved for attempts to repeat a trial under a single set of defining conditions, and not be applied to multiple attempts to measure a single trial.

Within a single test in which several parameters are being varied, repetition of events takes place at several levels. In a test of air-to-air missiles where each crew/aircraft fired at two different targets with four missiles each--one Monday morning, one Monday afternoon, and two Tuesday morning, they have tested with each missile-target combination four times, they have conducted 6 morning firings, 2 afternoon firings, 2 Tuesday morning single target firings, 4 Tuesday morning firings, 4 Tuesday firings, 4 Monday firings, , 1 Monday afternoon single target firing, etc.

Each of the numbers greater than one can be viewed as a repetition of some set of circumstances and is called a replication. Individual test requirements will determine whether it is important to replicate day-by-day firings, morning firings, firings at a particular target, firings on a single afternoon at a single target or only missile firings (or whatever).

Replications of a set of circumstances give information on the trial-trial variability within the bounds of that set of circumstances but do not, without more detailed replication, give information on the sources of variability within those bounds. Reflection on the nature of the population of inference should help the test designer decide at what level replication is necessary.

Statements of replication should tell what is being replicated, but at times statisticians will talk about "two replications" without saying what is being replicated; if this happens it can be assumed that the entire experiment is being duplicated as nearly as possible with each total combination of circumstances being realized exactly twice. It can be seen how this becomes ambiguous when dynamic scenarios are "replicated" because time (and time-correlated changes) do not stop to wait for the tester to replicate his experiment.

The amount of replication required can be predicted before a test under the right conditions. That is fortunate because it allows the tester to know, at least approximately, the degree to which his test data will give conclusive results. This is discussed in the next section.

13. PRECISION AND SAMPLE SIZE

In a test conducted to obtain data for a point estimate of a parameter (discussed in Chapter 13), the importance of sample size -- that is, the number of observations taken in different trials under similar conditions -- cannot be quantified. The intuitive feeling that a larger sample gives a better estimate of the true population parameter holds, but statements of a point estimate alone do not quantify the uncertainty of that estimate.

Most estimates are a combination of the point and interval type, and it is the interval estimate which tells how "good" the point estimate is. In constructing an interval estimate one takes into account (1) the point estimate, (2) the sample (including measurement) variability, (3) the sample size, and (4) knowledge of the sampling distribution of the statistic on which the point estimate is based. The point estimate gives a location for the center of the interval, while higher variability among the observations and smaller sample size each give a larger interval for the estimate than do lower variability and larger sample size. The reasons should be apparent. A point estimate is the best (by some criterion) single-valued prediction of where the population parameter really is. As the size of the sample goes up there is an increase in the probability that the random sampling has given a truly representative "picture" of the population of values being sampled. Yet, if the variability of values in the population is large, a given size random sample can pull the estimate further away from the population parameter than it can in a population with less variability. The statistic sampling distribution (section 3 of this chapter) relates the certainty of the estimate and often the sample size to a coefficient governing interval size.

For an interval estimate of the arithmetic mean the equation looks like:

$$\begin{matrix}\text{point}\\ \text{estimate}\\ \text{of}\\ \text{mean}\end{matrix} + \left(\text{coefficient}_1 \times \begin{matrix}\text{square root}\\ \text{of point}\\ \text{estimate}\\ \text{of variance}\end{matrix}\right) < \begin{matrix}\text{population}\\ \text{mean}\end{matrix} < \begin{matrix}\text{point}\\ \text{estimate}\\ \text{of}\\ \text{mean}\end{matrix}$$

$$+ \left(\text{coefficient}_2 \times \begin{matrix}\text{square root}\\ \text{of point}\\ \text{estimate}\\ \text{of variance}\end{matrix}\right)$$

Mean and variance are described in Chapter 13. Here, information on sample size and the sampling distribution of means has been combined into the "coefficient." The way in which these coefficients vary with sample size and certainty (confidence level) is shown in Figure 7-38. As a rule of thumb useful for sample sizes greater than five, to halve the desired interval length requires an increase in sample size by a factor of four. As an exact example, at the 95% confidence level increasing sample size from 9 to 36 (a factor of four) reduces the interval length by a factor of 0.473. Figure 7-38 shows this "inverse square root" effect graphically. The exact relationship between sample size and interval length is illustrated in Figure 7-39 and can be put in the following form:

$$\frac{\text{sample size}}{\text{slowly changing function of sample size}} = \left[\left(\frac{4}{\text{desired interval length}}\right)^2\right] \times \left[\text{anticipated variance}\right]$$

Before the required sample size can be known, the tester must have a grasp on the trial-trial variability of the test data. This a priori knowledge can come from reports of the variability observed in previous tests of similar systems, from DT&E or contractor data on the system under test, or perhaps from a short preliminary test conducted specifically for the purpose of shaking out an estimate of the required quantity.

The variability (in this case the anticipated variance) need not be known exactly. A ten percent uncertainty in variance would give only a ten percent uncertainty in the required sample size -- sample size 11 instead of 10 or 110 instead of 100. The desired interval length will not be a firm number anyway. If variability is estimated somewhat low, the resulting estimated interval (for the same confidence level) will be somewhat longer; who cares until it is longer by 50 or 100 percent? The impact of uncertainty in the anticipated variance can be quantified as follows:

$$\text{relative uncertainty in length of interval to be obtained} = \text{relative uncertainty in anticipated variance}$$

The relationships given above are understood most easily in terms of an experiment conducted to estimate a population mean. The general principles also hold in approximate form at least for other estimations and comparisons involving means and proportions.

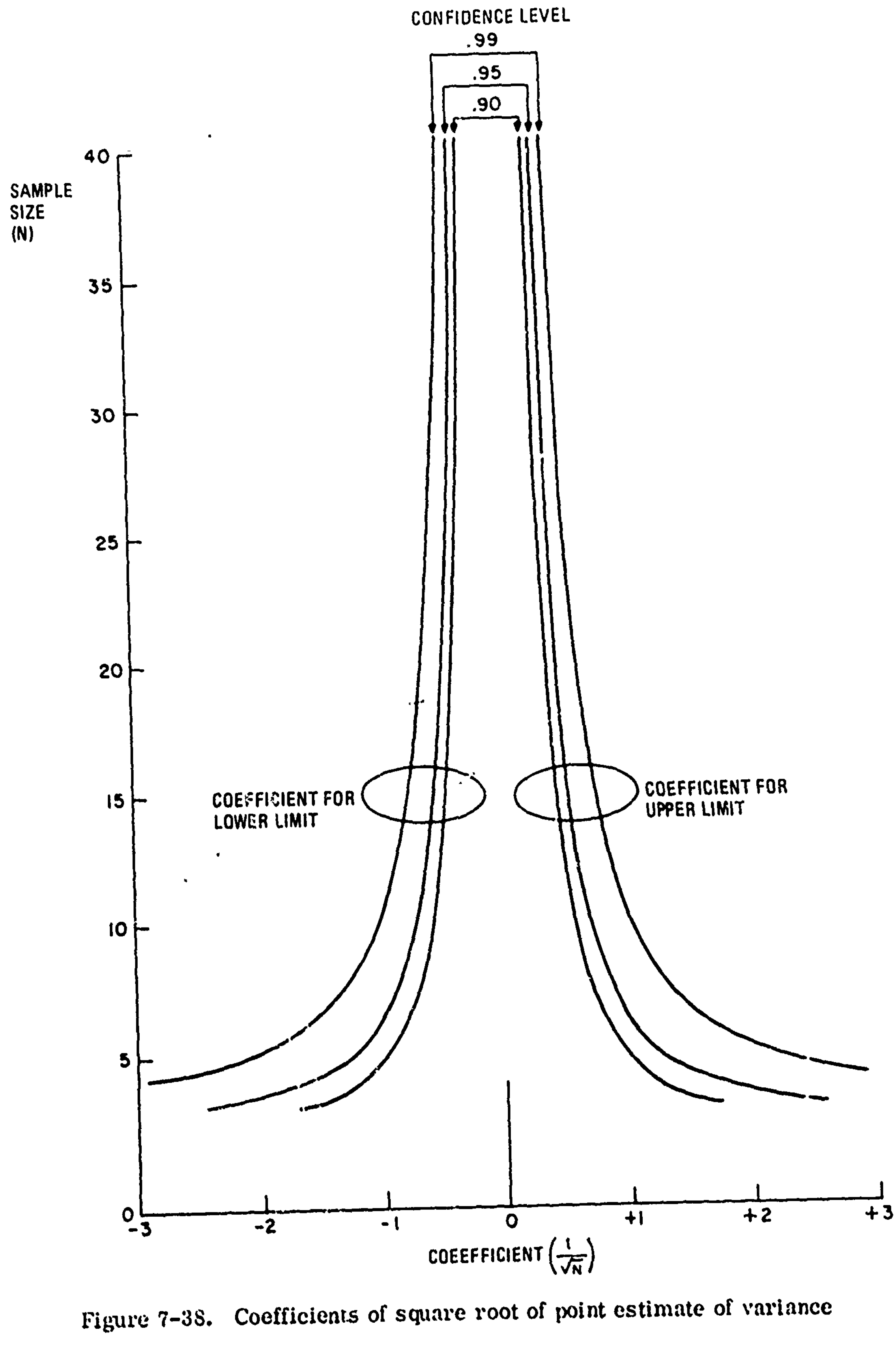


Figure 7-38. Coefficients of square root of point estimate of variance

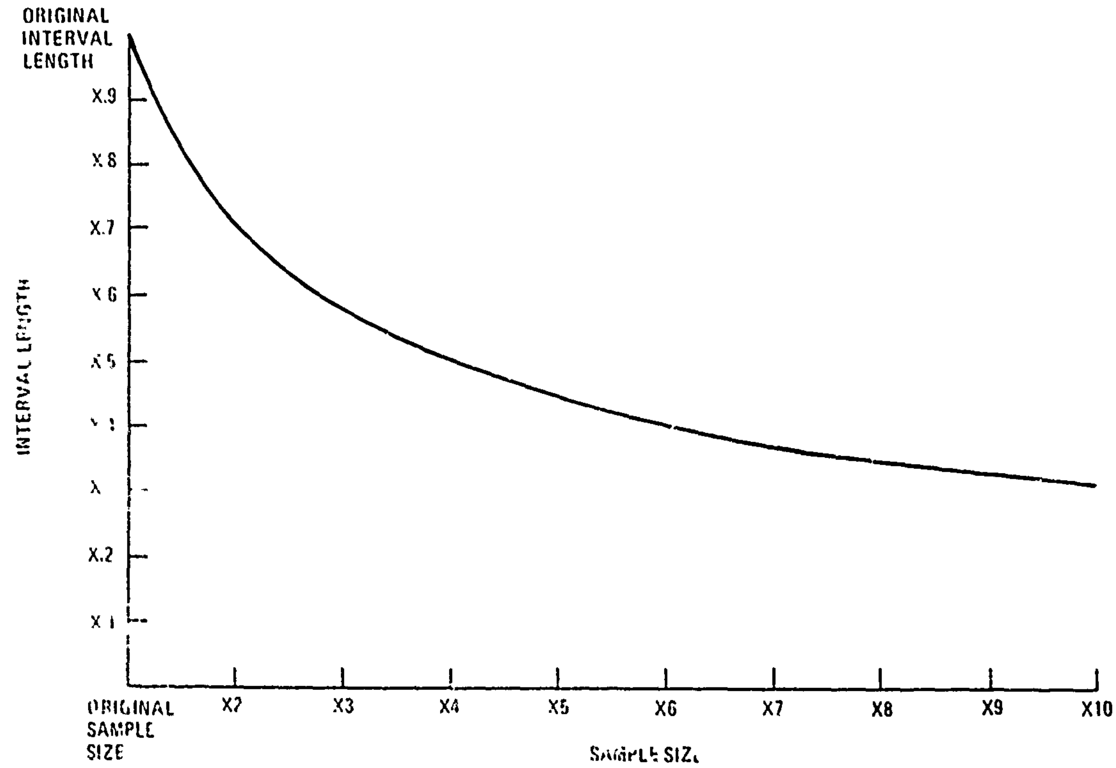


Figure 7-39. Approximate effect of increasing sample size

When applied to comparison problems, the sample size is the sample required in each one of the samples to be compared. (Samples of unequal size can be compared with proper adjustment of the formulae, however).

The interval estimate of variance equation is of the form:

$$\left[\text{coefficient}_1 \times \begin{array}{l}\text{point}\\ \text{estimate}\\ \text{of}\\ \text{variance}\end{array}\right] < \begin{array}{l}\text{population}\\ \text{variance}\end{array} < \left[\text{coefficient}_2 \times \begin{array}{l}\text{point}\\ \text{estimate}\\ \text{of}\\ \text{variance}\end{array}\right]$$

These coefficients are shown in Figure 7-40 as a function of sample size. They are not symmetrical and the inverse square root rule is not much help.

Estimation of the required sample size for all but the most simple experiments is a difficult matter. The trained statistician should be able to help.

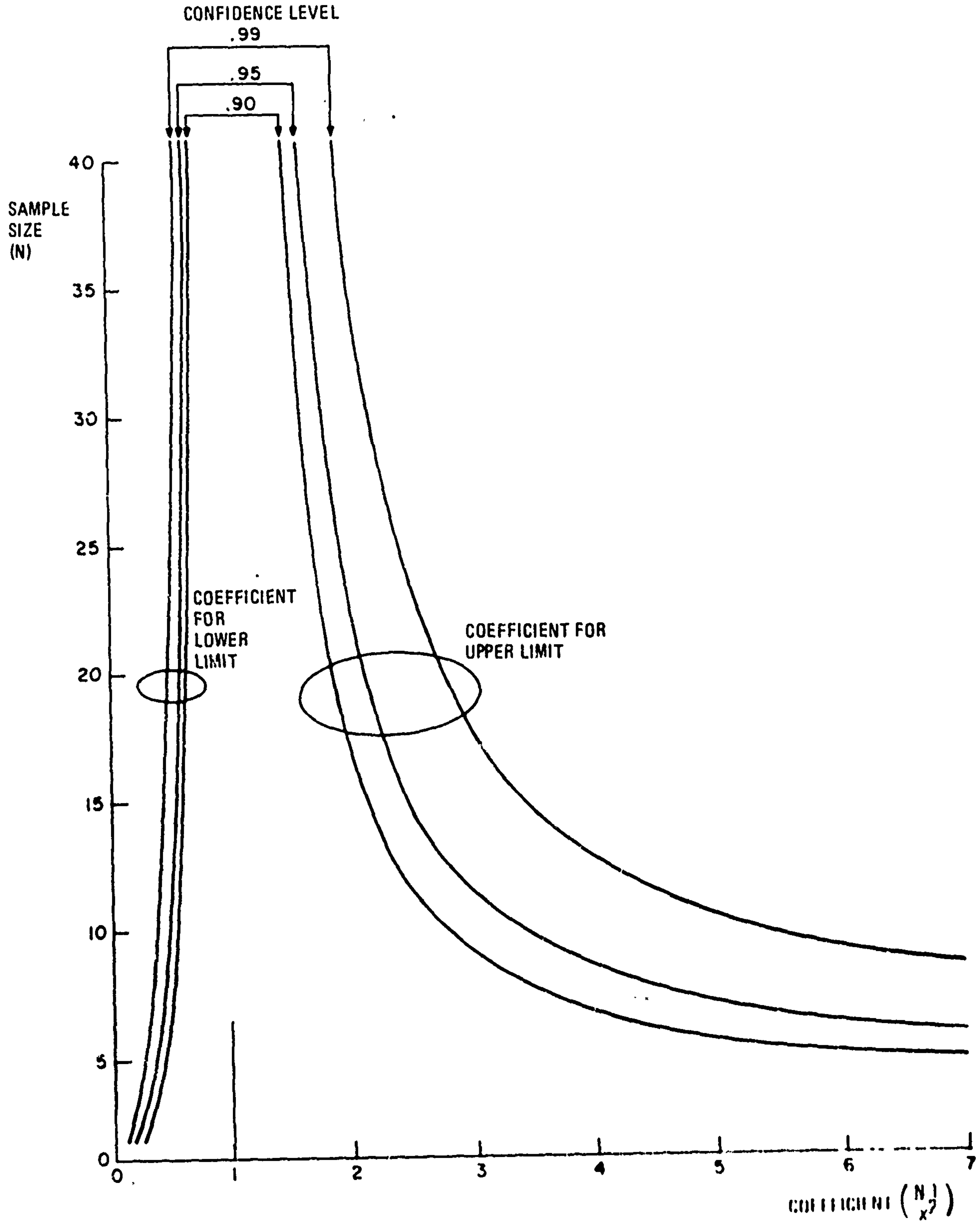


Figure 7-40. Coefficients of Point Estimate of Variance

In some instances it will be advantageous to start an experiment without knowing what the sample size will be. This does not really involve working completely in the dark, but rather uses techniques for revising the estimate of the required sample size in steps as the experiment progresses, and at times terminating the experiment early because the results have had such a strong bias in one direction that the trend cannot be significantly reversed except at the preset (low) risk level. These sequential experiments are applicable only in certain situations but if the circumstances are right they can make two important contributions to the operational test:

(1) sequential experiments provide a constructive approach to tests in which the variability is not well known; and

(2) sequential experiments generally require fewer trials to generate the same statistical information than the comparable non-sequential experiments.

Sequential tests are best explained by reference to a single estimation or comparison -- that is, a test having only one primary factor and in which the background factors are controlled either by holding constant at a single level or by randomization. Extension to cover the more complex designs will be addressed later.

The basic sequential test consists of analyzing the test data after each trial to make a decision about whether a statistically-supported conclusion can be reached (in which case the test can be terminated) or whether the test should go on to the next trial. This means that the tester must specify, in the case of an estimation, the confidence level with which he wishes to know the estimation interval and the length of the required interval. The flexibility in sample size, then, comes from the uncertainty in knowledge of variability. As can be seen from the discussion above, if the variance turned out to be lower than anticipated, a smaller sample size would be sufficient. An example is shown in Figure 7-41. The tester has specified that he wishes to know the mean one-dimensional miss distance (range only) within 5 meters and wants a 95% confidence level for the resulting interval. The variance of impacts is believed to be 25 meters^2. The statistician tells him that 18 firings will be required for the desired interval/

confidence combination. Conducted as a sequential experiment the results might appear as in Figure 7-41.

Trial	Miss Distance (Meters)	95% Confidence Interval Length (Meters)
1	-0.44	·
2	-5.08	58.93
3	-0.18	13.67
4	-1.86	7.16
5	0.72	5.64
6	-1.12	4.27
7	-9.64	6.75
8	1.56	6.14
9	5.16	6.44
10	-2.46	5.67
11	4.26	5.54
12	-5.50	5.28
13	0.14	4.83
14	1.02	4.48
15	0.76	4.17
16	3.38	4.04
17	-0.08	3.77
18	5.04	3.77

Figure 7-41. Miss distance data

It can be seen from the solid line in Figure 7-42 that the mean is bounded within a 5 meter 95% confidence interval after only 6 trials. The true variance of the population from which these numbers were taken is 16 meters squared (standard deviation 4 meters instead of the anticipated 5 meters) because the numbers were taken from a table of random, normally-distributed numbers with a standard deviation of 4. If the

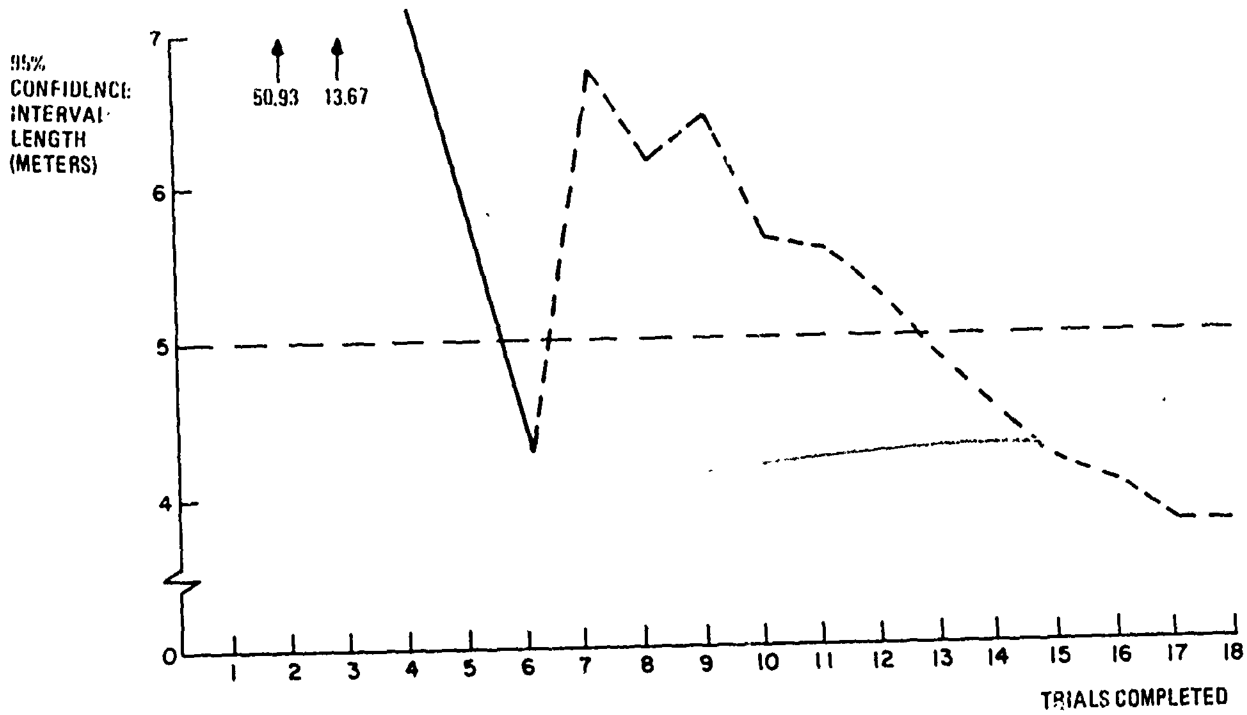


Figure 7-42. Miss distance data

variability from trial to trial consisted only of random deviations about the mean, this is a valid example of one of the infinite number of different samples that could occur.

The results may appear somewhat suspect at first. But it is true that although the tester may have a better gut feeling about running all 18 tests (possible trials 7-18 are also shown in Figure 7-41) than about basing his decision on the first 6, there is every bit as much validity in the 95% confidence interval estimate of the mean after 6 trials (-3.46 meters to 0.81 meters) as in the one computed after 18 trials (-2.12 meters to 1.65 meters). Although the 18-trial interval is shorter, both are less than 5 meters in length, which was the requirement. Notice that the seventh trial takes the 95% confidence interval back up to 6.75 meters. This is not a contradition; no statistical inference is a guarantee and it has just happened that from the random sample obtained in 6 trials, the interval is shorter than the one based on 7 trials would be. Even after 18 trials, the tester is not guaranteed that the true mean is inside the interval estimate.

For a 95% confidence interval, the tester is 95% sure that it is. It may be that the interval size does not get down to 5 meters even after 18 trials. Everything depends on the (random) sample selected. If the sample estimate of variance (not the same as the anticipated variance) is not 25 meters^2 or less, the 18-trial interval will be more than 5 meters long.

There may be valid reasons for not stopping after 6 trials. If the tester is not sure he has included a good random sampling of all sources of variability (maybe the sky was clear and sunny for all of the first 6 trials), he should continue the test. Sequential techniques can save money, however, and should be considered for simple tests.

Comparison experiments can be conducted sequentially also. Here the tester is looking for statistically significant differences between samples or between a sample and a hypothetical population and he must specify two things: (1) the risk he will accept in declaring a difference when there is in fact no difference, and (2) the risk he will accept in declaring no difference when in fact there is a difference. These are the α and β error risks discussed in section 4 of this chapter. After each trial the tester will (1) accept equality, (2) reject equality, or (3) continue testing.

The extension of these ideas to the handling of more complex tests depends on the ability to repeat "similar" trials sequentially. Intentionally-introduced variations (i.e., different levels of primary and/or controlled background factors) make succeeding trials fundamentally different from one another. It may be possible to review groups of trials on a sequential basis.

14. CONSIDERATIONS IN AGREEMENT TO A TEST DESIGN

It has been pointed out that the test design activity is not a single-thread mechanical process. Initial preferences get sorted out by several different criteria, the statistician's bag of tricks is invoked, and finally the Test Officer and statistician work together to hammer out the optimal design for the problem at hand. While the Test Officer is working primarily from his concerns for operational realism, conclusive results, resource expenditure, and relevance, the statistician makes his contributions in terms of the number of levels for each factor, the number of primary factor combinations, the number of trials in each background factor block, and trial-to-trial variability.

Some mutually-acceptable criteria for a good test as well as options for flexibility can be used to place mental priorities on various aspects of a test design and to guide progress toward the appropriate design.

Properties of a Good Experiment

A good experiment does not contain systematic error. Systematic error, or bias, is not limited to physical measurement processes, but can also be present in the form of personal habit and personal preference. The experiment should be planned in a way that precludes the influence of systematic error on data generation, data collection, data analysis, and data interpretation. A good experiment allows inferences to be drawn to a large population. The conclusions drawn, that is, have validity in a wide range of the circumstances to be encountered in operational employment. A good experiment is controlled to such a degree that the results can be reproduced. Only in this way can the tester support the results he claims. A good experiment includes an analysis of the uncertainty of the results. This means that the tester must be aware of the significance of conducting the experiment in different ways, of the control that is exercised over each source of variability (insofar as possible), and of the "degree of credibility" that can be given to the generalizations made about the test item as a result of this test. A good experiment is sensitive to real, operationally-significant effects. If there is a meaningful difference between two items or employment practices or scenarios under comparison, it should be revealed by the test. A test that is too coarse and vague is apt to be worthless -- unless worth is measured in terms of the satisfaction of having spent a certain amount of money. A good experiment is simple. Consideration of a large number of sources of variability dictates against a truly elementary test, but in the rush to do everything the tester must not entangle the essential elements of the investigation in resource-consuming and distracting details. A test that is elegant in its simplicity is probably a lot more convincing to others, too.

Selection of the Areas of Investigation.

If money is tight (Money is always tight!), the tester may have a need to assign some priorities to different areas of a potentially very far-ranging and thorough test and then investigate only the top-priority items at the expense of foregoing all information on others. Obviously, the tester should not waste any time on unrealistic situations; if it isn't going to be done in actual operational employment, there is no need to do it in the test. This requires the tester to have a thorough knowledge of the potential operational employment of a system and to be able to bound that employment both in terms of levels of specific variables and in the more general definition of missions. It is important to test the whole system, including the interfaces between hardware items as well as those between a hardware item and personnel. There are real differences in personnel involvement (quality, not degree) due to differences in motor skills, knowledge, motivation, and stress tolerances, among other areas, and because these differences do affect weapon system effectiveness they must not be overlooked. If one interface fails, the entire system operation may fail. The ideal operational test has many of the characteristics of a tactical exercise because elements of surprise and variable tactics are required for a truly operationally-realistic check of the system. Free play exercises, however, are hard to control (by definition), to measure, and to reproduce. A middle ground will be found which strikes the proper balance between the need for the results of operations and the need for operationally-meaningful results. Certainly the agreed-upon test design must address the areas of greatest risk (i.e., uncertainty about what capability the dollar spent will really purchase) in system acquisition.

There may be an initial temptation to make all the controlled independent variables primary factors. This is undesirable as it implies a requirement for measurement of the effect of different levels of each variable (and maybe interactions) to some particular level of precision when employment practices will not be based on such information anyway. The OT&E should be limited to a test which will simulate operational employment practices; if serious problems surface the system can be returned to DT&E for diagnosis of the detailed nature and causes of the problem.

Testing of the Proper Level of Detail.

Relating observations of operational effectiveness and suitability to specific causes has been historically difficult. The structure now available for outlining operational employment of a system, as presented in the Measures of Effectiveness chapter of this volume (chapter 6), will help, but it remains up to the Test Officer to decide at what level of detail cause and effect relationships will be sought for any particular test. As more money is spent in controlling variables at fixed levels and precisely exploring the effect of varying a few over a range of levels, less money is available to cover the full selection of variables and consequently greater leaps of faith must be taken in drawing inferences from the test results. It will not serve any useful purpose to draw distinctions at a level that will never be important in actual operational control of employment.

Other Economic Considerations.

The order in which an experiment is run can be important if it does not go exactly as planned. One principle says that the easiest trials should be run first because they will be the least expensive (fewer aircraft, personnel, etc., involved), and if the system shows faults under only moderately strenuous trials, the test can be terminated with the least expenditure of resources and the least risk to prototype hardware. Another principle says to schedule the complex and difficult-to-realize trials first (e.g., a trial involving extensive coordination of personnel and hardware employment with particular weather conditions) because if something falls through, the simpler trials can be rescheduled instead without necessarily delaying completion of the entire test program. These scheduling controls must be coordinated with randomization requirements to insure validity of the conclusions drawn from the sample actually obtained. A further goal of the tester should be to maximize nondestructive testing. If the right information can be obtained, it is usually cheaper to retest a single item several times than to blow it up and buy new ones for each trial.

Segmenting the Test.

The possibility of doing a preliminary test or small pre-test exercise to get a better fix on test variability was raised in the earlier discussion of precision. These preliminary tests can also be used to make sure the right subject areas are being investigated, to train testers, operator and maintenance personnel, and data collectors, and to debug the Test Plan. Sometimes the test will be too complex for the statistician to handle in one good design, and it will be more feasible to run two or more smaller tests where the first one looks only at the effect of a couple of prominent independent variables and later tests serve to amplify the work completed in the first. In other situations, a factorial arrangement may be split into fractional factorials to be run separately.

An Alternative Test Design Philosophy

When confronted with very limited resources, the tester sometimes decides to try a "shotgun" approach in which as many levels of as many variables as possible are tested just once in an effort to locate any real holes in a system's performance. The feeling is that one failure in some modes would be enough basis for rejection of the system, and that a statistically significant result is not required. A tester should keep in mind, however, the essential variability of trial results and the fact that variations of degree only or simply between passing and failing may be happening randomly and very infrequently. It is theoretically possible for all the molecules of water in a lake to move upward at the same time and lift the lake out of its bed. The tester must be able to satisfy himself and others of the fact that he has observed a real effect and not something as improbable as the jumping lake example.

15. CONCLUSION

This introduction to test design is presented as an aid to the Test Director in doing his job well and as an aid to communications. No attempt is made to create test designers; rather, the goal is to give the Test Director an idea of what to expect from his professional support staff so he will be in a better position to use that support and to make reasonable requests for assistance. In addition, the Test Director will be able to plan and conduct a more useful test if he is personally familiar with the basic concepts surrounding a designed test as well as with the importance of a well-designed test.

REFERENCES FOR TEST DESIGN

A good library will have a large number of related books for the Test Officer who wishes to learn more about any of the topics discussed in Chapter 7. The problem is to find one that suits the educational background and style of the individual user. Some of the best references for the non-statistician are:

Cox, D. R.; Planning of Experiments; John Wiley & Sons, Inc.; 1958. An introduction to the general concepts of statistical test design, based on the need to deal with different problems rather than on mathematical technicalities.

Davies, O. L., and Goldsmith, P. L. (ed); Statistical Methods in Research and Production. Hafner Publishing; 1972. Especially useful chapters on frequency distributions and statistical inference.

Dixon, W. J., and Massey, F. J., Jr.; Introduction to Statistical Analysis (3rd ed); McGraw-Hill; 1969. Especially helpful discussions of samples, sampling distributions, and estimation.

Mace, A. E.; Sample-Size Determination; Reinhold; 1964. The approach to determination of the proper sample size does not vary from problem to problem, except in specifics. This book gives those specifics.

Natrella, M. G.; Experimental Statistics (National Bureau of Standards Handbook No. 91); U. S. Government Printing Office; 1966. Also available as Engineering Design Handbook - Experimental Statistics (U.S. Army Materiel Command Pamphlets 706-110, 706-111, 706-112, 706-113, and 706-114). Contains a good problem-oriented discussion of statistical design techniques and some associated analysis techniques.

Snedecor, G. W., Statistical Methods (4th ed); Iowa State University Press; 1946. Later editions are more readily available, but this one has a reputation for clarity. Helpful sampling and sampling distribution discussions.

Wilson, E. B.; An Introduction to Scientific Research; McGraw-Hill; 1952. Although dealing primarily with research problems, this book provides a lot of good ideas on the means and principles of experimentation. Available in paperback.

Each of these books contains references to more advanced texts.

TEST DESIGN CHECKLIST

1. Are the required measurements of the independent and/or dependent variables identified?

2. Do you know what is sufficient to conclusively answer any of the questions being posed?

3. Has the population of inference been defined?

4. Have all operational employment variables been considered for a special type of control in the test?

5. Can the test objectives be addressed without knowledge of those variables in the Unmeasured category?

6. Can the test objectives be addressed without control of those variables in the Uncontrolled categories?

7. Are the primary factors controllable in operational employment?

8. Do the levels of primary factors adequately represent the questions being posed?

9. Do the levels of controlled background factors adequately sample the full population of inference?

10. Are requirements for precision being satisfied?

11. Can the uncertainty in results be quantified?

12. Are effects being confounded (unintentionally)?

13. Are requirements for information on trial-trial variability being satisfied?

14. Have sample sizes been chosen to give results with the proper level of confidence or to detect the required differences at the proper level of significance?

15. Are the sizes of the Type I and Type II errors acceptable?

16. Are the go/no go criteria rigidly set before the test?

17. Do any blocking designs used represent the population of inference in the correct proportions?

18. Do controls imposed by randomization satisfy the requirements for sampling the entire population of inference?

19. Can human bias affect the test results?

20. Can this experiment be reproduced?

21. Will this experiment detect operationally-significant effects?

22. Can the test be simplified?

23. Does the test measure more detail than is required for operationally-useful conclusions?

24. Is the data analysis plan complete?

25. Would there be an advantage to running a short preliminary test?

26. Has full advantage been taken of the ability to do nondestructive testing?

27. Has the possibility of running a sequential test been considered?

28. Does the fact that certain background variables are held constant at a single level prevent you from drawing inferences in certain areas?

29. Is the proposed test compatible with the documented (ROC, DCP, etc.) system requirements?

30. Has the statistical analyst agreed that he can derive the information you need from the test data?

31. Are resources available to support this test design?

32. Can the trials be ordered to minimize the cost of a premature termination of the test?

33. Can the trials be ordered to minimize the probability of a delay in test completion if bad weather or scheduling problems appear?

Chapter 8

SIMULATIONS

1. INTRODUCTION

A wide use of simulations is made in OT&E. In fact, all of OT&E is in itself a simulation except for the infrequent condition of a live war with test items being investigated in combat. In all other cases the degree to which the simulation represents the real war must be considered. A simulation may be defined as an artificial representation of a process or a situation whose complexities are too great to explore by other means. Its value lies in being able to control an input and easily observe an output and thereby be able to explore a wide range of values of the inputs and outputs.

2. TYPES OF SIMULATIONS

Simulations can range over a great breadth [illegible] and complexity:

The simplest type of simulation may be done with pencil and paper and the flow diagram of a process. For given inputs, calculations may be made of conditions at various points throughout the process.

A more complex simulation might be done with a computer and the program for the flight of a missile. With this simulation characteristics such as times required to accomplish various functions (such as intercept) can be determined and used in test planning.

An even more complex example might involve the representation of the enemy target, including its evasive maneuvers, run against the friendly weapon with its constraints. At times it may be useful to couple a man into the loop to obtain operator reaction times or human decisions.

The various types of simulations are summarized in Figure 8-1.

- Paper and pencil
- Mock-ups
- Mathematical formulae and desk calculators
- Computer and program
- War games
- Exerciser

Figure 8-1. Types of simulations.

3. PHASES OF APPLICABILITY

Simulations can be useful in many phases of the OT&E process from early preliminary planning to the analysis of data:

In the information gathering phase a simulation can be extremely useful as an instrument to learn more about the system to be tested. Various aspects of the system's response to different stimuli can be explored as well as the determination of the degree of interaction that may occur when several parameters are varied simultaneously.

In the categorization of variables during the early phase of test design a helpful estimate of the importance of various independent variables can be obtained by using a simulation to examine the sensitivity of the dependent variables to changes in the independent variables. In this way unimportant independent variables can be dropped from consideration and the limited resources available to the Test Director can be used on more interesting and productive items. In this same phase once the important independent variables have been defined, the simulation can be used to establish the range of interest for the variables. In this way the levels or settings for the primary factors can be ascertained for the test design.

Once the first cut preliminary test design has been obtained, the test conditions can be "dry run" by use of the simulation to explore for any unexpected results that might limit the scope, range, or usefulness of the test.

In fact, it is possible in limited cases (the more simple cases) that the simulation might be the main instrument of OT&E with the physical tests providing inputs to the simulation and being used as validation of the model at several specific points. The validated simulation can then be run to provide more continuous data over a large number of test conditions. This approach should not be attempted without full consideration of the precautions and disadvantages discussed later.

During the conduct of the test a simulation can be a useful near real time diagnostic tool if unexplained results occur that raise doubts about the advisability of continuing the test. The troublesome test conditions can be inserted into the model for a quick model result comparison with the actual test results. The simulation can quickly explore the whole region in the vicinity of the difficulty and, if the results are valid, determine the range of test conditions over which the surprising test results might be expected.

4. ADVANTAGES

There are several marked advantages to the use of simulations:

Simulations can be accomplished in a relatively short time given the existence of the computer program, personnel who know it and/or documentation to fully explain it.

The cost of accomplishing the simulation with the above stated conditions can be substantially less than the cost of performing comparable physical tests.

The number of test conditions that can be explored are very great providing virtually continuous data over ranges of interest.

Future conditions can be explored which today cannot be reproduced physically. For example, a missile's performance may be explored against a future enemy target whose altitude capability or speed capability cannot be physically reproduced within the present state of the art.

The above advantages all contribute to a better understanding of the test item's capabilities and limitations.

5. DISADVANTAGES

There are also a variety of disadvantages or precautions that must be observed in the use of simulations.

A current scientific author states, "A successful scientist knows that all models are somewhat defective and that certain aspects of his visualization do not apply to the problem in hand." To be successful with simulations one must pay close attention to idealizations, simplifications, and assumptions which may frequently rule out the realistic use of the simulation for the purposes desired. Frequently the assumptions made are more for mathematical convenience than for realism.

Simulations are useful only if they have been adequately validated. The validations that do exist have probably come from previous OT&E which immediately raises a warning. It is, therefore, to be expected that the range of validation is quite limited, and the information in previous test plans and test reports on conditions surrounding the data acquisition is so limited or indefinite as to be of little use.

Large simulations are to be regarded with particular caution. The dangers in this area are well summarized in a recent report (ARPA Report R-1060-ARPA/RC, dated May 1972; subject: "Models, Simulations, and Games--A Survey",) and is quoted below:

"The evident preference for large, all-machine models and simulations is questionable on several grounds. Large-scale, finely detailed Models/Simulations/Games that try to deal with problems having significant uncertainties may only serve to generate errors, not clarifying anything. Given what appear to be weak-to-poor data, extremely fine temporal and spatial levels of model resolution, and low levels of demonstrated concern for supporting research, the Models/Simulations/Games produced may have doubtful reliability.

Large models are usually complicated, expensive to build and use, take extended periods to operate and interpret, and are the least scientifically defensible. They quickly begin to suffer from the disorganization created by changes in purpose and personnel, bad documentation, gaps in logic, and problems of data-base preparation, maintenance, and validation.

If large models must be produced, the key to control seems to be in continuity of personnel. Changes of personnel have significant effects. Usage decreases because no one knows what a model is supposed to do, how it does it, or why. Where sunk costs are great, there is a tendency to use a large and expensive model anyway, even though none of its caretakers can determine its validity for new applications. Documentation should ameliorate this problem, but it seldom does."

Finally, appropriate validated simulations are difficult to find for most OT&E purposes. However, some do exist and some effort is warranted in developing from experience a catalog of existing simulations, and also researching the literature to locate models. Some existing catalogs and models are listed below:

"Weapon System Effectiveness, Analysis, Optimization & Simulation"
Tech Report AFALT-TR-71-20
AF Armament Lab

"A Piloted Power Approach Simulation"
Tech Report AFFDL-TR-73-27
AF Flight Dynamics Lab

"Evaluation of Air Defense Analysis, Digital Simulation for Aircraft Vulnerability"
Tech Report TN4565-3-73
Army Material Sys Analysis Agency

"Current Mathematical Models (Simulations) for Digital Computers"
Tech Report NWL TR-2390
Warfare Analysis Department, Naval Weapons Laboratory

"The Literature of Gaming, Simulation, & Model Building Index & Critical Abstracts"
ARPA Report R-620-ARPA
Rand Corporation

6. SUMMARY

In summary, simulations can be useful in many phases of OT&E; however, caution is advised in checking on inherent assumptions, simplifications, and idealizations, as well as on the range of validation the simulation has attained. Large simulations are particularly suspect lest the Test Director becomes placed "at the mercy of the model."

Chapter 9

DATA

1. INTRODUCTION

The full value of a well designed and conducted test can not be realized unless adequate data are collected to support the test objectives. A common error in designing test is to select a set of trials as the driving factor of a test design and then to try to force fit a data collection scheme that may or may not support an adequate analysis and evaluation. The selection of test profiles and the derivation of data requirements are inseparable and of equal importance. Well executed tests with inadequate data collection are as useless as irrelevant tests with excellent data collection.

The following provides an approach to development of a data support scheme and outlines essential considerations and procedures associated with the role of data in test and evaluation.

2. DATA COLLECTION

Data collection is the act of capturing and recording raw data, to include the gathering of the recorded data to a prescribed location. The three basic methods of performing this function, each with attendant advantages and disadvantages, are manual, semi-automatic and automatic. Occasionally the required data will dictate a particular method, but generally a choice exists.

Manual data collection involves a human observer with no machine assistance. It is generally divided into objective and subjective categories. Collection of objective data is supported by forms that record quantitative data that does not require the judgment of the recorder. An example of this approach is the USAF maintenance data collection system where maintenance actions, running time hours, and other prescribed functions are recorded on specified forms. There are numerous existing objective data collection systems in the USAF and where possible the test designer should incorporate their usage in the data collection scheme. Examples of standard T& E data collection forms in common usage are AFSC Form 258 and AFFTC Form 0-294 used in conjunction with the AFSC SEDS aircraft evaluation system and AFTO 349 and 350.

Subjective data collection employs questionnaires. Data collection forms are used to record objective and verifiable data; questionnaires are used for collecting opinions and not to record actual measurements or events. They are used to record likes and dislikes but not the capability or performance of an object under test. They are particularly valuable in assessing the suitability and compatibility of the man-machine interface and should be used for all tests characterized by such interfaces. The construction of questionnaires is discussed in greater detail at the end of this chapter.

Manual collection of data is attractive from a cost standpoint and is an effective approach within certain limits. It is suited to the recording of events that do not occur in rapid fashion and hold their value constant for a reasonable period. Examples are maintenance actions, pre-flight checks, counting the holes in a target, and measuring the impact point of a bomb. The accuracy of the observation deteriorates rapidly as the frequency of observation is increased or when external influences are inserted. The recording of performance, position, or engagement data by a pilot will result in progressively poorer quality data as the intensity of the test increases. Manual recording of altitude values from an altimeter by a dedicated observer tends to become less accurate and reliable as the rate of change or rate of recording increases.

Semi-automated data collection involves the use of both man and machine. It should not be confused with the manual collection of objective data which may include the recording of meter readings, etc. It is a system that is dependent upon both an operator and some device such as a manually operated tracking camera. The actions of the two are interdependent since either camera malfunction or poor tracking by the operator can contribute to loss of data. The distinction between manual and semi-automated data collection is made to point out that manual methods are seldom employed to measure performance while semi-automated systems, with less predictable limits, are often improperly stretched beyond repeatable capabilities to provide data that should rightfully be collected using automated techniques.

Automated data collection refers to the measurement and collection of data with little or no data collector involvement. The distinction between semi-automated and automated systems is governed by whether an operator must actively participate in gathering the data as it occurs. An auto-tracking radar with appropriate data recording devices is typical of this category, although it may require operator assistance in initially acquiring track of the target. An on-board or integrated monitoring recording system on the test object is also representative even though it may require someone to start it, stop it and change its tape. The measurement accuracy, precision, and performance of the automated system is the most predictable of the three approaches to data collection. Also, the use of automated data collection is often dictated by the rate of change of the measured values and the duration of the coverage requirement. The principal disadvantages of automated systems are that they may produce nearly unmanageable quantities of data for manual reduction techniques, and that their expression of measurement may be misunderstood or abused during the processes of reduction and analysis. Unmanageable quantities of data may be resolved through proper planning and application of automated data reduction techniques while it is advisable to use personnel with a thorough understanding of test instrumentation to avoid the second disadvantage.

3. DATA VERIFICATION

Data verification is the process of assessing whether the data correctly represents the variable it characterizes and insuring that sufficient data is collected to support the test design. The process is performed prior to conduct of the tests (validating collection method); during conduct of the test (insuring that critical data is being recorded), and after completion of the tests (when quick-look analysis indicates that the recorded data consists of questionable values). The results of verification may lead to rerunning tests where critical data was lost, revision of test profiles where the degree of realism prevents dependable collection of data, or revision of the test design where valid data cannot be collected without unduly compromising the realism of the test.

Manual, semi-automated, or automated methods may be employed to verify data. Usually, the approach used to verify data is the same regardless of the method used. The data is first inspected for general content, format, and continuity:

- Does the data appear as specified in the data plan?
- Is alphabetic data appearing on forms where only numeric data should be entered?
- Is expected analog data erroneously represented by pulse or stepping type data?
- Does the data recording comply with the coverage requirements or does it have gaps where coverage was specified?
- Is the data item annotated with prescribed events, phases and recorder start/stop times?
- Are the proper timing or reference signals present?

The second level of verification is to check the recorded data against what was expected:

- Do the data values represent the anticipated range of values for the test or do they grossly exceed what was predicted or known?
- Is the frequency or pattern of the data representative of what could be expected?
- Do variables held constant during conduct of the test fluctuate in the data representation?
- Does data which is known to vary appear constant?

Manual verification is straightforward and requires few supplemental tools. Where quantity of data is not a factor, manual inspection is most effective. The human observer is capable of detecting a larger variety of data errors than most automated methods. Assessments of data item quality are easily performed by humans.

Criteria for selecting a manual or automated approach - or some combination of both include data quantity, the form of the data, and the degree of repetition. Large quantities of data with considerable repetition are most suited to automated and semi-automated verification techniques. For example, verification of the data from a 16-channel recorder that will be used on a series of 12 identical tests would best be handled by a properly programmed automated system that would verify each test's data in a matter of minutes.

The tools for manual verification of data are relatively simple and inexpensive. The verifier must have a description of the data presentation format and the predicted data characteristics for the individual trial or test. This information should be readily available from test profile charts, data item description forms and operating test logs (described in subsequent sections). Scales for translation of chart data to meaningful representations are basic and may be prepared by hand or mechanically for each data channel. An inexpensive and effective scale is easily constructed by duplicating the chart recorder calibrations on a piece of chart paper and then using the individual channel calibrations as scales for each channel. Accurate verification of chart data may be accomplished with magnifying glasses, professional scales (Gerber, Matson, etc.) and data scanners. Film readers and appropriate playback and display units are required for review of film and video data. Their use should be programmed for the verification process as well as analysis applications. Other devices, such as audio reproduction equipment or photo-detection devices may be required depending upon the type and form of data collected. A reasonable rule of thumb in selecting devices and aids to assist manual inspection is that the same devices used for translating and displaying data for analysis are often required for verification.

Verification of data entries on forms and questionnaires is particularly suited to manual processing. Aside from verifying that each block on a form has the required entry, the format of the data is easily checked, and the content of the entries may often be verified for logic and magnitude. The review of manually prepared data (forms and questionnaires) is always important as omissions and errors in preparation are common.

Semi-automated verification is appropriate for a wide variation of data situations. An example is data that is manually inspected for format and completeness and then entered into a computer for verification of values. Other examples are the use of calculators and tabulators for running total checks, and semi-automated inspection of photogrammetric data values through use of an X-Y scanner. Roll chart scanners with chart illumination and transluscent scales are effective for review of roll chart data.

Automated data verification is usually applied through the use of general purpose or special purpose computers. The principal value of computers is the ability to rapidly check anything that may be reasonably predicted. Because of their speed they are appropriate for checking large volumes of data. A further advantage is the ability to rapidly correlate or manipulate data to provide a higher degree of confidence in the verified values. Airspeed, altitude, target position, and range to target may be cross correlated rapidly by the computer to determine if the results are reasonable. Their biggest drawback is that computers must be told exactly, step by step, what they are required to do.

A computer may be able to verify 100 trillion data entries in the amount of time it takes an individual to prepare the computer instructions required to perform the check. However the quantity of data may be such that the data can be verified manually in considerably less time than it takes to develop the computer instructions. The approach must further be weighted in terms of how many manhours are available during testing.

Many standardized computer routines are available for checking various characteristics of data. Data formats and continuity are routinely verified by computer and existing routines can accommodate a wide range of variables like alphanumeric characteristics, length of the data entry, whether an entry is required, etc. Use of this capability requires defining the structure of the data to be checked by the computer. Verifying the value of the data is more complex and is related to how well the range of values can be predicted, how often the values change, and how many data elements have common ranges. The implementation of computer verification is best determined by a coalition comprising the test designer, support officer for data, and data processing and operations analysis personnel.

In summary, data verification is necessary because it may force revision to either the test profiles or test design. It may also impact on the value of the testing. The verification method selected relates directly to the time available for review of the data and the relative costs and quality of the method selected.

4. DATA REDUCTION

Data reduction is the process of transforming raw data into useful, ordered, or simplified form. Data reduction may be accomplished either manually or in an automated fashion. A typical example might be the transformation of tracking radar range, azimuth, and elevation raw data into aircraft position according to some coordinate system. The same radar raw data may also be "reduced" to present aircraft velocity, bearing, altitude, range to target, dive angle, etc. Another example may be the translation of telemetry analog values into units that represent the actual factor measured (i.e., 3.5 volts on a 0 to 5 volt analog scale may equate to 70% fuel level).

The definition of raw data is often ambiguous but should be thought of as the form of the data at the time it was recorded. The confusion occurs when there are various possible levels of reduction in presenting the data. To illustrate, the analog values of a measurement, if recorded on magnetic tape in digital format, would have to be transformed back into analog values on a chart or other analog type display before they are understandable. One analyst may view the reduction to an analog representation as being adequate because he translates the data directly from the chart using a scale representing the chart values in units related to the factor being measured. Another may feel the data is usable only after he receives the measurement values represented in list form versus time.

There are many reasons for reducing data. Along with presenting the data in intelligible form the analyst may only want to sample segments or regular intervals of a large quantity of data, reserving the option to reduce the unsampled data at a later time. Reduction of event data may be limited to presenting the times of each event rather than displaying an entire recording and then searching for the events. The same data may have to undergo several reduction processes, resulting in different presentations of the same values. Different combinations of data may be summed, subtracted, or otherwise manipulated during reduction for analysis purposes. Because of this wide variation in meaning of data reduction it is essential that data reduction requirements be clearly stated in a step-by-step fashion. Block diagrams and flow charts are valuable tools in defining reduction requirements to avoid misunderstandings.

Manual reduction methods consist of translating recorded measurements on charts or displays into units (degrees, feet, etc.) related to the factor being measured (dive angle, miss distance, etc.). It may also involve calculating sums and ratios, determining elapsed time between events, and otherwise putting the data in a form suitable for analysis. A manual approach to the reduction of large quantities of data is neither smart nor recommended. Aside from the prohibitive time requirements, accuracy and precision of data are degraded by operator fatigue in a relatively short period of time, no matter how simple the reduction process. Complex reduction processes are best assigned to computing machines with less susceptibility to error. Unless the data reduction process is relatively straightforward and addressing small quantities of data it should be accomplished by data processing equipment. Typical tools supporting manual reduction of data include pencil and paper, slide rules, hand calculators, data scales, reference tables, data chart scanners and film projectors.

The principal considerations in selecting automated data reduction methods are the availability of Automatic Data Processing (ADP) equipment, applicable reduction methods and procedures, quantity of data to be reduced, complexity of reduction processes and time available. If appropriate reduction methods and procedures are not available these must then be considered from the standpoint of planning the development of new ones (i.e., computer program development). Availability of reduction equipment is an easier problem to cope with as the data may be reduced on a variety of computers at locations removed from the test site, again, assuming routines and methods are available or can be developed for use on the candidate computers.

When operators record readings from data display devices this is termed semi-automated data reduction. Data display devices include semi-automatic data scanners and film readers, TV displays, position plotting boards, data comparators, etc. The process is often limited to sampling of the data in a prescribed scheme (time interval, events, phases, etc.) Because of the human element it is susceptible to error and misinterpretation. Accuracy and precision of the data from semi-automatic reduction is about on a par with that produced by manual techniques. The advantage of semi-automated reduction is that it is usually faster than strictly manual reduction and thus reduces operator error caused by fatigue.

5. DATA SUPPORT SCHEME

Preparation of the data support scheme for a particular project is interrelated with the test design and associated test profiles. The objective of the data support scheme is to provide the data specified by the test design as necessary for analysis and evaluation of the test. This imposes the requirement for a thorough understanding of the test design and test profiles as well as data collection methods and systems.

Prior chapters of this document described the approach to test design and the identification of discrete elements for analysis called variables. The data support effort must provide data corresponding to these variables. The variables may equate directly to data elements on a one for one basis or several data elements may be required to produce one variable (Figure 9-1). Measurement of the linear distance between target center and weapon impact would result in a data element which would relate directly to the variable "Radial Miss Distance." The variable "aircraft altitude" might correspond to a single data element produced from altimeter measurements or it might be produced from a combination of range, azimuth and elevation data elements from a tracking radar. A data element may also be used with other data elements to produce more than one variable. Range, azimuth and elevation data elements from a tracking radar may be used to provide the variables "aircraft altitude" and "aircraft velocity" while contributing to a "range to target" variable.

The following is an organized approach to development of the data support scheme. It includes definition of variable "characteristics" within the test profiles, selection of methods (data collection and reduction) to provide the required variables, and guidelines for maintaining control of the developing data support scheme.

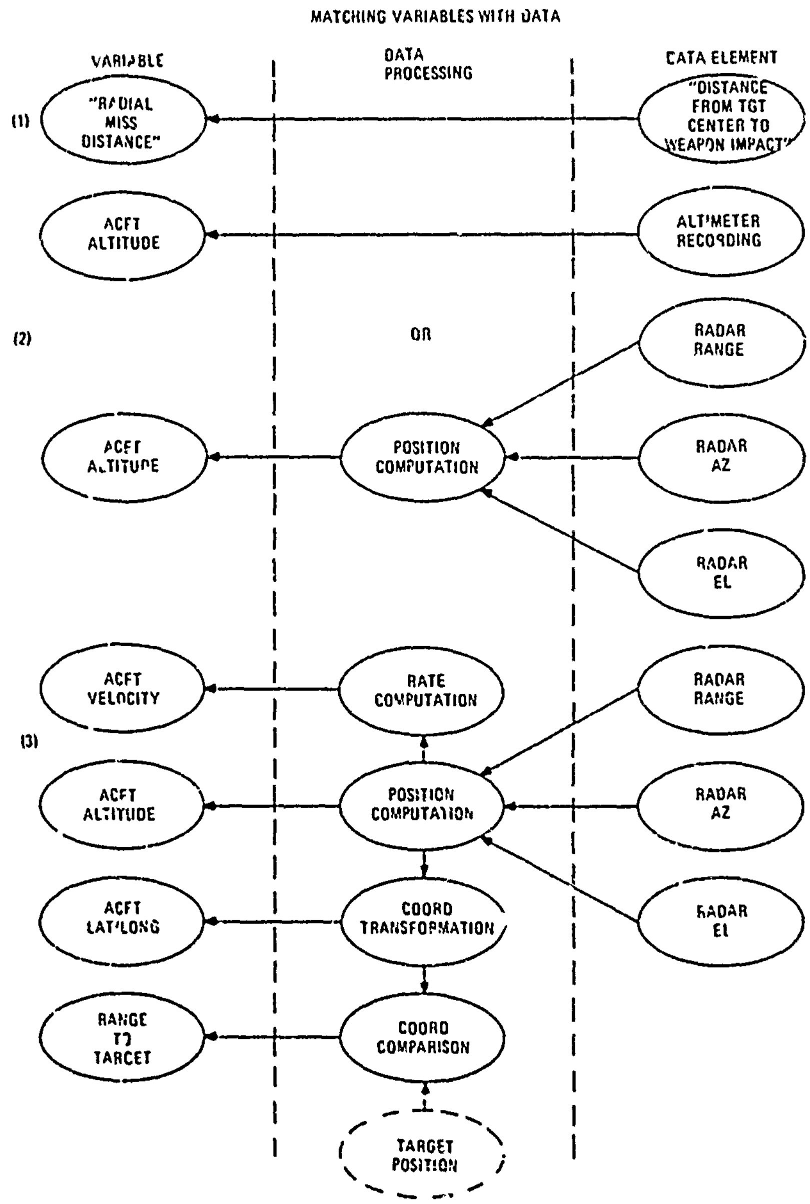


Figure 9-1. Matching variables with data.

<u>Defining test design variables</u> is often not sufficient in its self to permit selection of the data collection and reduction methods. The test profile may exert a greater influence on the selection. The range of values or required accuracy for a variable often changes during the course of an individual test. As this may dictate the selection of different data sources for the same variable during different phases of a test, the Support Officer for Data (whom we shall call the SOD) requires an organized approach to separating and assessing the impact of these influences. One approach is the use of time-line

techniques, referred to as Data Profile Time-Line charts (DTL). The purpose of the DTL is to display the variable characteristics versus test profile segments. A representative DTL is illustrated in Figure 9-2*. Along one axis of the chart the SOD lays out those significant test segments during which data must be collected for analysis or test control purposes. The characteristics of the variable that are likely to be affected by variations in the profile are listed on the other axis. Below the variable characteristics space is reserved for noting when the data is required and recording what methods will be used to provide the data. Once the chart is set up the SOD should enter the appropriate characteristic requirements under each test segment. This process should be repeated for each variable until all variables are accounted for. Characteristics to be considered are the range of values, accuracy, precision, frequency of change, coverage requirement, unexpectedness and priority of the variable.

Range of values refers to all the values that a particular variable may have. For the variable "Altitude of Object A" typical values might range from 1500 feet to 50,000 feet for one segment of an air combat maneuvering test profile and remain relatively stable at 20,000 feet during a "station keeping" segment.

The utility of defining the range of values for each variable is that when applied to the measurement capabilities of instrumentation or observers it places boundaries on the test area, the degree of realism and freeplay in the test, and on the knowledge that can be obtained from the test.

*This chapter illustrates various representative "forms" or "formats" that are shown simply to demonstrate how such tools are used. They are not, in any respect, advanced as standard forms. When the program is assigned to a given test range (or test area) the forms/formats in use there should be sought out and applied. In some cases these might be from the RCC UDS. If such is the convention--this should be used.

DATA PROFILE TIME LINE CHART

Test Project #		NAME		SOD	Subtest #
Data Parameter/ Attributes	Subtest Profile				
	Taxi	Take Off	Climb	Cruise	Penetration Etc.
Altitude, Object A					
• Range	-	0-200 ft	200-5K ft	5K-30K ft	20K-30K ft
• Accuracy	-	± 10 ft	± 100 ft	± 200 ft	± 100 ft
• Precision	-	10 ft	10 ft	10 ft	10 ft
• FOC	-	Continuous	Continuous	Controlled	Moderate
• Coverage	None	Continuous	Continuous	Check Point	Continuous
• UF		Low	Low	Low	Moderate
• Priority		Desirable	Desirable	Required	Mandatory
(when required)					
• Real Time				Range Safety	Range Safety
• Quick Look					+ 2 Hours
• Final		+ 24 hrs.	+ 24 hrs.		

Figure 9-2. Data profile time-line chart.

Accuracy refers to the degree to which the measurement differs from the true value. Using the example, "Altitude of Object A", accuracy refers to how close the altitude measurement represents the actual object position, e. g., to ±10 feet. Care should be taken to assure that accuracy of measurement is not over-stated as the cost of measurement rises rapidly with increasing accuracy. Frequently the accuracy required for a particular variable changes within a test profile. For example "Aircraft pitch attitude", may require varying degrees of accuracy in aircraft position for the delivery of different types of weapons. With this condition the required measurement accuracies for each weapons delivery mode should be stated separately rather than by stating the most stringent accuracy as the requirement for all modes. This approach allows a logical determination of what specific requirements may not be met without inferring that the overall accuracy cannot be met simply because it relates to the most stringent requirement.

Precision relates to the degree of exactness of a measurement. It is sometimes confused with accuracy which specifies the degree to which a measured value deviates from the true value. Data precision is a term that corresponds to the number of significant digits in a measured value. For example if the measured height of an object that is exactly 72 inches high results in a value of 72 inches the measurement has 100% accuracy and adequate precision. Extending the precision to tenths or hundredths of an inch will not change the accuracy, and in this case not improve our knowledge of the object. If, however the object were 72.53 inches tall, extending the precision could improve our knowledge of the object's height yet not improve the accuracy of the measurement. There are practical reasons for not extending the precision beyond the ability of a system to accurately measure a value. For example, a precision requirement to display altitude values to a tenth of a foot when the accuracy of the measurement is ±100 feet is poor application of precision criteria and could result in wasted analysis of meaningless inaccurate variations.

Frequency of change (FOC) is a reasonable estimate of how often the variable value changes and must be stated by the test designer. This characteristic is determined during test design and relates to how often the variable must be measured or observed to satisfy the analysis requirements and provide a desired data confidence level. The thoroughness in addressing this characteristic is reflected in the quantity of data that will be collected and reduced, and could have an impact on test analysis if adequate reduction capability cannot be provided to cope with unexpected large quantities of data.

Coverage requirement, the period during which the measurements are required, must be well defined. In conjunction with the frequency of change characteristic, it provides an indication of data volume. The combination of the two characteristics more completely defines measurement requirements and often identifies the need for redundant recording. Redundancy refers to the physical recording of the same variable on separate recording media. Variations in frequency of change characteristics during a mission could dictate periodic bursts of high rate data collection that could not be sustained by a single recorder for the entire mission. Another example is where coverage duration simply exceeds maximum recorder capacity. There are efficiencies to be obtained through careful analysis and definition of the coverage requirements. For example the recording of measurements during periods when the data is not required for analysis or conduct of the test unnecessarily adds "noise" to the overall data support effort.

The Unexpectedness Factor (UF) relates to how predictable the assessments of the variable characteristics are likely to be. In information applications the richness of the factor is measured by the unpredictability of the information and relates to recording only that information that you cannot reliably predict. The factor expresses how much confidence is placed on the values predicted for each variable characteristic and should affect the selection of those values. At a minimum, some latitude should be considered when selecting data sources for those variables with a high degree of unexpectedness.

Priority indicates the degree of importance attached to the variable in the test design. Typical categories of priority include "mandatory," meaning that the test would be a failure if the data were not gathered; "required," indicating that loss of the data would have a serious impact upon the value of the test; and "desirable," which implies that the data would contribute to knowledge gained from the test but that its loss would not have a serious impact upon test results.

Once the characteristics have been estimated for each test segment the SOD should then define when the variable is required. Test control, analysis and management are considerations at this point. Test control relates to data required for conducting each test in accordance with the test objectives and procedures. Test management refers to determining whether the test program (as opposed to a specific test) is progressing satisfactorily and what alternatives in the test program may be exercised. The management requirements are usually fulfilled by abbreviated analysis of data related to critical objectives or otherwise supporting the decision process.

Three generally accepted terms for describing when data is required are "real-time," "quick-look," and "normal." "Real-time" refers to near instantaneous collection, reduction, and display of the data and is usually reserved for test control purposes. "Quick-look" is intermediate or simplified data produced in a relatively short time frame (within a few hours) after completion of the test and is used extensively to support test management objectives. "Normal" refers to routine data reduction requirements normally accomplished within a period of days or even weeks. The length of time required to accomplish routine data reduction requirements varies significantly dependent upon the quantities of data, type of reduction and test priority. In addition to defining the reduction category, the data delivery time from test completion in hours, days, or weeks, should be specified for each data item.

Selection of Data Collection and Reduction Methods naturally follows completion of the data profile time-line charts. This process requires a high degree of familiarity with data collection methods, instrumentation, appropriate test ranges, and data reduction methods. Depending upon the size of the test project it may be a formidable task; it is therefore expected that the SOD will seek technical assistance from in-house or range agencies established for this purpose. When using this technical assistance the SOD function is that of evaluating the general applicability of proposed data collection methods and systems, and selecting the most valid and dependable approaches.

Selecting the method of data collection is the first step in fulfilling the data requirements. Method relates to the manner in which the data is collected. Will the data be provided through measurement or observation? Will it be recorded by instrumentation, or people, or both?

The preferred data collection method is usually measurement by instrumentation to prevent the influence of subjectivity. However, operator opinion data is often desired as the man-machine interface is one of the principal considerations of Operational Test and Evaluation. In fact, all tests characterized by man-machine interfaces should have provisions for human judgment data that expresses the operator's opinion of the utility of the object.

Using the Data Profile Time-Line chart, the SOD may proceed to match data sources to the time sequenced data requirements. Initial inspection can show data characteristics vary considerably along the time-line with respect to the stringency of the measurement and the SOD should be prepared to consider alternate methods of collecting the data rather than attempting to apply the highest level of stringency to the data collection method. An example is where test article position accuracy requirements vary considerably between range safety (surveillance) and a weapons delivery phase (precision tracking). Obviously, applying the stringency of precision tracking to the range safety requirement is not cost effective. The test designer should bracket those areas on the DTL where the characteristics vary significantly. Segregation of the significant areas should be applied to each variable until all requirements have been scrutinized.

Often the initial inspection of the DTL, or command policy, will dictate a particular test range (or ranges) that may be used to support the tests. In this case the SOD submits his completed DTLs, along with supporting descriptions of the test design profiles, objectives, and proposed schedule to the Range Requirements Office for initial review and translation into specific methods of fulfilling the data requirements. The Range Commanders Council Universal Documentation System formats can be used for this purpose if in use at the selected range. The requirements translation phase, whether conducted in-house or by a candidate test range, will require considerable dialogue between the data support officer and the agency selected to translate variable requirements into potential data sources. The translation will often dictate a reduction or processing requirement to produce a specified variable from diverse data elements, further complicating documentation of the translation process. To maintain control of the process SOD should enter the proposed data sources and any required reduction processes below the variable characteristics on the DTL, or alternatively as attachments to the charts. The translation should include what instrumentation will provide the data elements, data element format, recorded measurement rate, recording medium, recording identification and data channel assignments, as well as the recorded data range, accuracy, and precision of the raw and reduced data. These specifications should also be documented with each DTL to facilitate comparison and insure continuity.

As translation of requirements into data sources often results in variations between the DTL requirements and range capability, the SOD must record these deviations on the DTL and determine whether their effect upon the test design warrants revision of the test profiles, possible purchase of instrumentation, or revision of the test design. Larger projects often require all three approaches in design of the data support plan.

During identification of data sources the SOD must assess the dependability of the data source. Decisions for data redundancy at this point may prevent considerable difficulty later in the project when a marginal data source fails to provide essential data. No hard and fast rules apply to redundancy although cost of measurement is always a factor. Consideration should be given to essential data requirements and data that is difficult or impossible to duplicate or approximate. Other factors influencing redundancy are the ability to monitor the recorded data in near real-time, the amount of control possible over each phase of the test, the unexpectedness of the data content, and the validity of redundant sources of data. Where data collection is provided by an established test range the issue of redundancy is satisfied by the test range in accordance with the priorities assigned to the variables.

Defining the data reduction and display requirements is the responsibility of the SOD in conjunction with determining the data sources. Display requirements refer to the form of the data presentation, and when and where it will be presented. The data reduction and display requirements are derived principally from the test design, i. e., what data will be displayed in what units of measure, according to what order or functions, and referenced to what coordinates. When and where the reduced data will be displayed, and how many copies of the data will be produced, are also questions germane to data reduction. Of basic importance in describing the reduction and display requirements is clearly defining the units of measure, the coordinate system to be used, the scales of the presentation, and the precision and accuracy of the reduced data. Where possible, the reference for a particular process or definition should be cited (i. e., CEP) to avoid interpretation or misunderstanding by the element reducing data.

Where there are alternate sources of data (redundancy) there should be flexibility incorporated in the scheme to allow for reduction of the back-up data when the primary source fails or is questionable. The question of when the data is reduced relates to the urgency associated with the results anticipated from the data. Data required for test control or safety must be presented in real-time to facilitate making decisions concerning conduct of the test. The same raw data providing this real-time display may be subsequently reduced hours or days after completion of the test for analysis purposes.

Data reduced for analysis is also ordered by its importance to the test project. Here we refer to data that indicates whether the test was a success or failure versus data that merely provides background information on performance of the article under test. Data display requirements should be stated in terms of (near) real-time, quick-look, preliminary, and final with delivery times stated in terms of hours or days from mission termination. In specifying the reduction and display requirements the data should always be identified by source name, data item (recording) number, and recording channel assignment.

Properly selected variable descriptions often imply the reduction requirements Identifying variables as aircraft altitude, aircraft range to target, etc. when these are the desired variables indicates the manipulation of metric tracking data that must be transformed (reduced) into the desired variables. Processing requirements on the desired variables for analysis purposes (frequency analysis, regression analysis, etc.) should be prepared by the analyst supporting the test design.

It is essential that data reduction requirements be communicated clearly between the SOD and the agency supplying the service. Regardless of who determines how the requirement should be met, the procedures of the reduction process should be documented in step-by-step detail, accompanied by block and flow diagrams.

As the data sources and processes required to supply the variables are identified on the DTL, the SOD should prepare data origin charts providing an overview of the data flow. A representative data origin chart is illustrated in Figure 9-3.

In block 1 the name and number of the analysis variable are entered as these are principal objectives of the data collection scheme. Block 2 contains the name and number of the data item (recording) containing the variable and the channel, block, track, or axis within the data item where the variable appears. The name or description of any reduction or processing requirements applied to data elements for producing the displayed variable are entered in block number 3. In block 4 are entered the data element(s), one per line, required to derive the variable. Block 5 identifies the source of the data elements and block 6 is used to identify where the data elements are recorded prior to

DATA ORIGIN CHART

PROJECT NUMBER:				NAME:					SOD:						
1 VARIABLE		2 DATA ITEM		3 RORD REDUCTION/ PROCESSING	4 DATA ELEMENTS	5 DATA SOURCE	6 DATA ITEM		7 TEST						
NBR	NAME	NBR/NAME	CHANNEL/ BLOCK				NBR	CHNL/ BLOCK	1	2	3	4	5	N	

Figure 9-3. Data origin chart.

being processed in producing the analysis variable. Block 7 is used to identify what tests require collection of the data element(s) to produce the variable. Where the variable is produced by more than one method (multiple reduction techniques or different sets of data elements and sources), a separate entry for each process is required by repeating the variable entry and each new procedure. The completed form should allow the SOD to trace the source(s) of each variable and assist him in maintaining control of the developing data support scheme.

In conjunction with accounting for the individual variable requirements the SOD should identify the composition of the individual data items as in Figure 9-4. In addition we must also account for the associated calibrations and recording environment. Under "environment" the items of interest are recording rate (frames/ sec., pulses/sec., etc), recorder speed, timing reference and coverage requirements (recorder start/stop keyed to test segments). For continuity, the source of the data elements, associated variables, and originator of the data item are included. The completed form provides the test designer with an overview of the contents of data items he must control.

DATA ITEM IDENTIFICATION

PROJECT NBR　　　TITLE　　　SOD

DATA ITEM		ORIGINATOR	SOURCE	CHNL TAB/ BLOCK	DATA ITEMS	VARIABLES SUPPORTED	CALIBRATIONS			ENVIRONMENT			REMARKS
NBR							RANGE	WHEN	HOW	DATA RATE	RECUR RATE	COVERAGE REQMT	

Figure 9-4. Data item identification.

The data control form (Figure 9-5) represents a means of checking that all required data items for each test are accounted for by recording who should receive each data item when each item is to be delivered, and the organization that will serve as a repository for each data item. It should include data release stipulations, and any other instructions that facilitate retrieval of the data during its life cycle. Classification of the data item is also indicated on the form with any special handling requirements covered in the remarks section. Classification down-grading instructions would also be included for handling by the data repository.

With the required variables accounted for and the subsequent flow and characteristics of the data documented, the SOD may proceed to verification of the data support scheme. Data verification ranges from insuring that all required data elements are accounted for to checking the recorded data for validity. The most successful technique, but often the most difficult to accomplish, is simulation. Dependent upon test complexity, verification through simulation may involve nearly as much planning as does the scheme undergoing validation. Exceptions to this are largely repetitive tests with a small number of data elements.

PROJECT NBR						TITLE				SOD:		
DATA ITEM			RECORD DISPOSITION		DISTRIBUTION		TESTS					REMARKS
NBR		TYPE	OPR	REMARKS	PRI	ORGANIZATION	1	2	3	4	...... n	

REMARKS - ( ) DATA ITEM NBR IS OF THE FORM A B C/D A - ORIGINATING ORGANIZATION. B = DATA SOURCE.

(1) C = UNIQUE DATA SET NBR ASSIGNED BY TEST DESIGNER. D = SECURITY CLASSIFICATION OF DATA ITEM.

EG WSMR S360-13/S MIGHT EQUATE TO A DIGITAL RECORDING TAPE, CLASSIFIED SECRET, CONTAINING XYZ DATA FROM FPS-16 RADAR NBR S360 AT WHITE SANDS MISSILE RANGE

(2) DATA ITEM TYPE REFERS TO RECORDING OR DISPLAY MEDIUM. i.e. T=TAPE, C=CHART, ETC.

(3) DISTRIBUTION PRIORITY IS QUOTED IN (NEAR) REAL-TIME OR HOURS AND DAYS FROM END OF TEST

(4) THE NUMBER OF COPIES OF EACH DATA REQUIRED BY EACH RECIPIENT IS LISTED UNDER EACH TEST COLUMN

Figure 9-5. Data control form.

Simulated data may be produced by a trial test that produces data similar to that expected during a real test, or the data may be produced manually or by hardware without a trial, according to some pre-conceived plan. Although the more valid of the two, the trial test is usually the most difficult approach due to normal time constraints on test project milestones and the unavailability of hardware, instrumentation, or funds at this point in test planning. The second approach commonly used in checking reduction and display processes, involves the creation of artificial data, distribution of the data, and subsequent reduction, display, and analysis of the data. With the exception of creating artificial data, all other processes are performed in accordance with the procedures developed in the data scheme. In the case of forms or questionnaires it often proves whether the data elements were properly selected or described. To a limited extent, control and distribution of data may also be verified.

Creation of artificial data, in the case of forms and questionnaires, involves establishing the criteria of the situation to be described and selection of a representative group of personnel to complete the forms or questionnaires. To be a success the group must know the criteria and understand the purpose of the simulation. The completed data items would then proceed through reduction and analysis to determine if the displayed data represents what was anticipated and whether it is of actual or potential value to the analysis. At this point the test designer will often discover that some items on the questionnaires produce completely unexpected results. This is usually caused by ambiguity of the questions that leads to misinterpretation of what is requested.

The creation of artificial instrumentation data requires establishing the analysis indications and then deriving the raw data characteristics and quantities to support the analysis. The method of creating these characteristics and quantities via instrumentation must be carefully planned and controlled to insure that the artificial data resembles the expected test data. The difficulty of creating realistic data cannot be understated and may often appear not worth the derived benefits.

Regardless of the difficulty encountered in creating artificial but realistic instrumentation data the price should often be paid where the reduction process is unique and where different organizations are used to support the collection and reduction processes.

Problems detected during the verification process may reflect as far back as the test design so it is imperative that the verification process be well documented. The selected criteria, method of accomplishment, participating elements and a detailed description of the results are minimum aspects to be documented. Properly treated, the documented verification may assist in solving data problems that occur during conduct of the test.

The verified data support scheme should now be integrated with the Test Plan. Data control procedures and check lists may also be incorporated in the test plan at the option of the Test Director although they are primarily oriented toward insuring that the data support scheme is implemented as designed.

6. DATA CONTROL

With the basic data support scheme reasonably developed the SOD should develop procedures to insure compliance with that scheme. Procedures and checklists should be prepared to insure (1) data validity, (2) proper distribution, handling and control of the data items, (3) proper conduct of the data collection effort during each test, and (4) that data support related events occur as required within the overall test project schedule.

Checklists and schedules should be prepared for the project as a whole and for each unique test type when the data support effort varies. Project related events should include approval of the data scheme by data collection and reduction elements, data collection training requirements, arrival and checkout of instrumentation, publication or acquisition of required data forms, acquisition of data recording and display supplies, and retirement of data items to eventual repository. Events related to individual tests or missions are more detailed and serve to insure that the data collection/reduction/display requirements are carried out as specified. Typical of test related events are checkpoints during preparation and conduct of the test to insure instrumentation and displays are ready for support, review of "red-line" readings, start and stop times of data calibrations and data recordings, scheduled changes to data or recording environment (e. g., revisions in sample rate or recorder speed), collection/ recording contingency plans, delivery times of data items, and acknowledgement of data item validity. The test related checklists should be used as the foundation for pre- and post-test briefings for all personnel participating in the data support effort.

During conduct of the active test phase procedures are necessary to verify authenticity and quality of the data and to insure that all data is distributed properly and is accounted for.

Distribution and accounting for the data is easily handled by distributing data item lists (similar to Figure 9-5) to all data recipients prior to each test and instructing them to report any deviations from the delivery schedule or in the data items received.

The SOD should establish procedures for inspecting all data items to determine quality. The inspection is best performed by the individual responsible for analysis of the data as he is more apt to identify gross deviations in the anticipated data characteristics.

The inspection should include, as a minimum, that the data item identification (sheet) is present and filled out, that prescribed calibrations are present, that the data representations appear normal and indicative of an operable condition, that the data item is properly annotated with respect to test profile segments and events, and that the specified reference or timing is present on the recording.

Adequate training and indoctrination for the individuals involved in collecting and handling the data is essential to insuring that the data support effort is a success. It also provides a degree of confidence that the procedures and objectives of the support effort are understood and complied with. Development of training or indoctrination plans will often lead to the development of additional procedures or checks as their need becomes more apparent when considering, in detail, what each individual's role and limitations are with respect to the overall data support effort.

Procedures establishing an unrestricted communications channel between data collectors, processors, analysts and controllers ensures that the specialized talents of all individuals have the opportunity of contributing to the success of the data support effort. Free flow of information and ideas should be encouraged with the understanding that absolutely no deviations from the support scheme or procedures are to be made by any individual without the concurrence of the SOD.

7. CONCLUSION

In summary, the objective has been to point out that no matter how simple the test project may be, it cannot be assumed that useful data will be collected unless an organized approach is used in preparing for the data collection effort.

Definition of data related terms and a discussion of the functions associated with collecting, reducing, and verifying data have been presented to acquaint the Test Officer with the role of data in support of testing and evaluation. The following chapter "Range Instrumentation" treats many of the systems supporting the different data functions and should be used for familiarization with ranges and test instrumentation. However it is no substitute for the use of qualified specialists in instrumentation and test facilities during the development of the data scheme.

An organized approach to preparing the data support scheme for test projects has been presented to assist the Test Officer in determining what data must be collected, how it is to be collected and presented, and what procedures are nec ssary to insure the proper data is collected. A simplified overview of this process is illustrated in Figure 9-6. To assist the Test Officer in maintaining order during development of the data support scheme and conduct of the test a generalized checklist is presented at the end of this chapter.

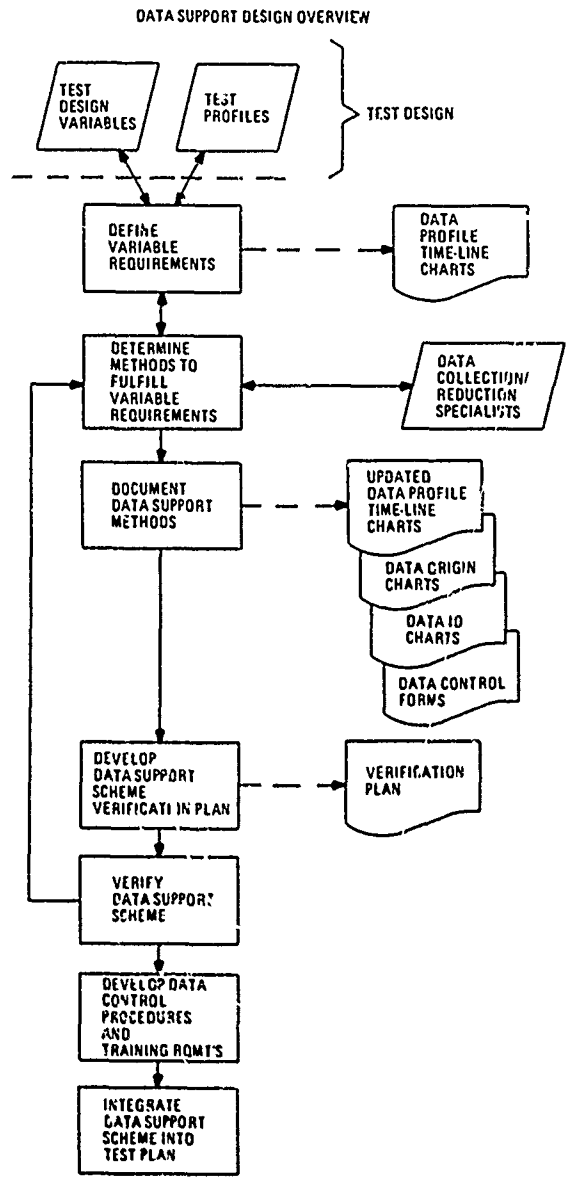


Figure 9-6. Data Support Design Overview

8. SUBJECTIVE DATA COLLECTION AND ANALYSIS

a. Introduction

Operator and maintenance personnel must be considered as important contributing factors to the overall suitability and effectiveness of every system being considered for operational deployment. Because procurement specifications cannot anticipate all user-desirable features, an objective of OT&E should be to collect subjective opinions of the system's suitability. Numerous data collection systems and instrumentation techniques are available to acquire objective data and measurements relative to a system's performance. Often, however, the Test Officer must custom design questionnaires for the collection of subjective data relative to the system he is testing. This paper treats with considerations for the design of questionnaires and the analysis of data collected with them.

The following conventions are used: (1) OBJECTIVE DATA refers to measurements and factual observations of events, made by a human data collector and recorded by him on data collection forms or voice tracks on magnetic tape recorders. (2) SUBJECTIVE DATA refers to the data collector's opinion of a situation presented to him in questionnaire form; the key word here is "Opinion." Measurements and factual observations ideally leave no room for opinion, and are not discussed here. (3) QUALITATIVE DATA refers to "ballpark" measurements and observations not necessarily recorded by a human data collector. This data can be objective, but never subjective. (4) QUANTITATIVE DATA refers to measurements and observations, not necessarily recorded by a human data collector. When these measurements and observations are acquired and recorded by means of instrumentation, that instrumentation is usually calibrated, and the validity of the data can be described in terms of confidence intervals. This data can be objective, but never subjective. Objective data is recorded on data collection sheets; subjective data is recorded on questionnaires.

One major application for questionnaires is to determine human factors effects in system design and operation. Data of this type is very difficult, and often impossible, to acquire and record in any other manner. MIL STD 1472 defines human factors design criteria, and can be used effectively as a basis for formulating a questionnaire.

Another major application of questionnaires is for reporting upon the quality of services provided by supporting agencies. In this case, the questionnaires are usually composed by the supporting agencies and distributed to the agencies receiving the service. The data is used as a "self-policing" measure to insure that the quality of service is maintained at a high standard.

b. Questionnaire Construction

The Test Officer should consider the following procedure for construction of questionnaires:

(1) Select topics in MIL STD 1472 which are applicable to the system being tested.

(2) Determine which topics should be addressed to operator personnel, and which should be addressed to maintenance personnel.

(3) Decide upon a method of administration for the questionnaire. Among the methods available are:

(a) Post-test Debriefing - advantages include:

1 Responses are still "fresh" in the mind.

2 Misinterpretations of questions can be avoided.

3 Spontaneous responses can be secured.

4 100% sample return is assured.

(b) Post-test Debriefing - disadvantages include:

1 Free-form answers may not lend themselves to statistical analysis.

2 The debriefer may unintentionally bias his questions or his recording of responses.

3 Interviewers may require specialized training.

4 Debriefings require more of a respondent's time than any other method of questionnaire administration.

(c) Take Home or By Mail - advantages include:

1 This method requires less time and expense to administer.

2 The need for trained interviewers is eliminated.

3 Unintentional biasing of questions is reduced.

4 The answers provided may be more complete and well thought out than they would be in a debriefing.

5 This method is usually most convenient to the respondent.

(d) Take Home or By Mail - disadvantages include:

1 Many questionnaires will not be returned.

2 Some questionnaires will be unusable because of illegibility, incompleteness, or unintelligibility.

3 Memories of the test may not be fresh in the respondent's mind.

4 Completion of data collection may require more time.

5 Written answers may be misinterpreted.

(e) "Over the Shoulder" Interviews (while testing is in progress) - advantages include:

1 Continuous, real-time observations may be recorded. This method can be effective in assessments of supportability, maintenance, and training requirements.

(f) "Over the Shoulder" Interviews - disadvantages include:

1 Potential interference with the conduct of the test.

2 The requirement for a trained data collection team.

3 "Bits and pieces" answers from numerous respondents may require careful analysis to derive meaningful data.

(4) Decide upon a form for each question which is appropriate both for the method of application of the questionnaire, and for the information which must be extracted from the response to the question. For example, the possibilities for free form answers to debriefing questions will be wasted if the respondent is allowed only a "yes" or "no" answer. Likewise, a multiple choice answer may not include the best possible response choice, and the respondent is forced to compromise his answer to fit the choices available. An example is used to illustrate the types of information which may be derived from a questionnaire by proper wording;

(a) Should the AN/PRC-XXX Two-Way Radio be adopted for use in USAF motor vehicles?

yes ☐ no ☐

Comment: This approach will yield the easiest data to process and analyze, as only two answers are possible. This is done, however, at the expense of visibility into deficiencies which produced the "no" answers, and deficiencies which were accepted within the "yes" answers.

(b) What is your opinion of the overall suitability of the AN/PRC-XXX Two-Way Radio for use in USAF motor vehicles?

Outstanding	(no deficiencies observed, no reservations or doubts about its merit)
Good	(few doubts about its merit, and none serious)

Adequate	(can probably serve its purpose, despite some shortcomings)
Inadequate	(unacceptable in its present state, can probably be made adequate with minor fixes)
Poor	(unacceptable in its present state, probably requires major re-design and major fixes)

Comment: This approach yields a more precise insight into the test participant's opinion of the system he is testing. Data analysis is still relatively straightforward. No insight into the nature of the system's shortcomings is provided, however.

(c) Please comment upon the following aspects of the AN/PRC-XXX and suggest improvements when possible:

1 The Channel Selector

2 The Whip Antenna

3 The Hand Microphone

4 The Volume Control

5 The Squelch Control

Comment: This approach calls for free-form open-ended answers in sub-categories of the overall system. Insights into deficiencies are provided but data analysis becomes difficult. Often, however, trends may be spotted even in a relatively small number of samples.

(d) Which of the following statements most closely describes your feelings about the AN/PRC-XXX Two-Way Radio:

☐ 1 If it came with the motor vehicle, I'd replace it with another radio.

☐ 2 If it came with the motor vehicle, I'd use it even though I didn't particularly like it.

☐ 3 I see very little difference between the AN/PRC-XXX and a conventional Two-Way Radio. I'd use either one just as readily.

☐ 4 Given a choice between a vehicle with a AN/PRC-XXX, and a vehicle with a conventional radio, I'd probably choose the vehicle with the AN/PRC-XXX.

☐ 5 I'd strongly recommend that every conventional radio be replaced with a AN/PRC-XXX.

Comment: This approach simplifies data analysis and decision making. If a limited number of options is available to the Tester, he simply presents them to the respondent and asks him to choose the option which most closely matches his wishes. In this example, the options were to: (1) Remove all conventional radios, (2) Return all conventional radios and (3) Remove some conventional radios.

(e) Describe your experiences with the AN/PRC-XXX Two-Way Radio.

Comment: This "open-ended" approach is probably the least desirable. It allows the respondent to give as minimal an answer as he wants, just to satisfy the requirement. It does not inform the respondent about what, in the realm of his experience, is relevant to the test at hand.

(f) Which features do you consider most desirable about the AN/PRC-XXX? Assign each a number (1, 2, 3, etc.) in order of descending desirability:

☐ The Channel Selector

☐ The Whip Antenna

☐ The Hand-Held Mike

☐ The Volume Control

☐ The Squelch Control

Comment: As the number of items to be ranked increase, the difficulty in organizing and assigning rank increases for the respondent. Excessive numbers of items should be avoided. This approach focuses attention on desirable aspects of the system. A ranking in order of ascending desirability would focus attention on system weaknesses.

(g) What is the maximum distance from base you have successfully operated your AN/PRC-XXX?

( ) 10 to 20 miles

( ) 30 to 40 miles

( ) 50 to 60 miles

( ) more than 60 miles

Comment: This type of approach can be used effectively to provide the respondent a means of answering a question when he does not know or cannot remember the precise value being asked for.

(5) Determine what information is necessary to identify the respondent. Typically, this may include:

Name

Rank/Grade

Unit/Mail Code

Description of training received for participation in this test.

Degree of familiarity/experience with the test item

Formal training received in test item operation/maintenance

This information may be requested in questionnaire form, if desired, by adapting each question to one of the forms described above.

(6) Assemble the questionnaire.

c. Reviewing Completed Questionnaire

Apply the following checklist to each question. Unless every answer is affirmative, consideration should be given to rewriting that question.

(1) Is the topic of the question relevant to the test?

(2) Does the question require consideration of only one topic or idea which must be responded to?

(3) Will the answer to the question provide the information it was intended to? i.e., is the question ambiguous or likely to be misunderstood?

(4) For multiple choice items, are the choices mutually exclusive, clearcut, distinct, and separate?

(5) Has a third response (e.g., "No Opinion," "Neither," "Don't Know," etc.) been considered?

(6) Has provision been made to keep the third response from becoming a "cop-out" that enables the respondent to avoid providing useful data?

(7) Is the wording of the question clear and precise?

(8) Have "leading questions" that may bias the respondent been avoided?

(9) Have lengthy lists of items which must be ranked according to preference, been avoided?

(10) Have "antagonistic" questions which force the respondent to reveal self-derogatory information been avoided?

(11) Is the scope of the question sufficiently limited to enable the respondent to decide what experience to base his answer on?

(12) Apply the following checklist to the questionnaire as a whole. Unless all answers are affirmative, consideration should be given to restructuring or limiting its size.

(a) Does each question "stand alone?" Unless they do, a misunderstanding, provoked antagonism, or inaccurate answer may propagate itself and invalidate a "chain" of questions.

(b) Does the physical appearance of the questionnaire convey the impression that there is importance attached to it? Sloppy, handwritten or corrected questionnaire forms do not encourage sincere, meaningful answers. Professional-appearing questionnaire forms suggest that considerable time and effort has gone into their preparation, and that the answers will be carefully looked at.

(c) Is the length of the questionnaire reasonable? Having an understanding of what duties the respondent will have performed before participating in a post-test debriefing (e.g., several hours in a cockpit, a full day of flight-line maintenance, etc.), the Test Officer should ask himself whether he could (or would) maintain concentration throughout the entire questionnaire.

(d) Are preliminary instructions kept as brief as possible, and as uncomplicated as possible?

(e) Is the amount of writing required of the respondent reasonable and kept to a minimum?

(f) Has consideration been given to the possibility that the questionnaire may not be returned, may be illegible, or unintelligible. The successful accomplishment of data collection should not hinge upon obtaining usable data from every respondent.

d. Statistical Considerations

In the event that the sample size of subjective data for an OT&E test program is small, many commonly used statistical methods may have limited application. The sample size will be equivalent to the number of operators or maintenance personnel who participated in the test. Conceivably, this sample size could be one. Therefore, considerable importance must be attached to thorough identification of the participant's relevant backgrounds and previous experience with the test item or similar items. Dissimilarities between the respondents and actual USAF personnel who will operate and maintain the system will bias or negate the value of the data.

The Test Officer should seek qualified data analysis assistance in dealing with subjective data. He should coordinate his questionnaire with data analysts prior to the test (1) to insure that meaningful data can be derived from its application, and (2) to formulate a plan for analysis of the data.

CHECKLISTS FOR DATA

A. DATA SUPPORT SCHEME PLANNING CHECK LIST

1. Set up a project data support file. Establish data support log for recording appropriate data events.
2. Review project objectives
3. Review test design and test profiles
4. List 'variables' required by test design
5. Add all other potential data requirements to list.
6. Identify 'variables' that are supplied by established data collection systems (AF Maint Data Collection, WSEP, etc.)
7. Plot 'variable' characteristics versus test profile segments on data profile time-line charts (DTL)
8. Prioritize the data requirements on the DTL in accordance with test design, test control, and test management needs.
9. Enter when the data is required to the DTL.
10. Verify that 'variables' and other data candidates are accounted for on DTLs.
11. Segregate significant variations in data requirements on DTLs.
12. Identify candidate data collection and reduction methods for each variable.
13. File copy of each DTL in project data file.
14. Submit DTLs, test profiles, project schedule and other supporting information to candidate test range(s) or data collection agencies for review and translation into collection methods.
15. Document identified collection/reduction methods as they occur on the appropriate DTLs.

16. Advise test designer of potential problems in supplying data in accordance with the test design. Enter correspondence with appropriate DTL in project data file.

17. Acknowledge receipt of range statement of support capability (SSC).

18. File copy of range SSC in project data file.

19. Prepare preliminary cost estimate of data support scheme.

20. Review Range SSC with test designer to identify data support problems.

21. Coordinate acquisition of required instrumentation and establish schedule for same.

22. Prepare revised DTLs, as necessary, based on range SSC and revisions to test design or test profiles.

23. File copy of revised DTLs in project file.

24. Prepare data origin descriptions to correlate collection/reduction processes to each 'variable'

25. Coordinate data origin descriptions with test designer, test analysts, and collection/reduction agencies.

26. File copy of data origin descriptions in project data file.

27. Prepare descriptions of each data item to be collected and file copy of each in the project data file.

28. Prepare "data item" distribution control forms and check forms for continuity and completeness against data origin descriptions.

29. Coordinate distribution control forms with data recipients and Test Director.

30. File a copy of distribution control forms in project data file.

31. Order necessary supplies to support data support effort.

32. Submit manpower requirements for support of data collection reduction effort.

33. Prepare any required data collection forms or questionnaires.

34. Prepare project schedule of data support events to include the following:

   a. Procurement of required instrumentation

   b. Data support personnel arrivals/departures

   c. Data support supplies acquisition

   d. Receipt of required instrumentation

   e. Instrumentation integration with test article

   f. Checkout of required instrumentation

   g. Data training schedule

   h. Data support procedures distribution

   i. Any decision forks in the program affecting data support

   j. Completion and verification of associated software development (computer programs)

   k. Distribution of draft and final test plan

   l. Active test operations

   m. Project phase-down operations

B. ACTIVE TESTING CHECKLIST

1. Prepare pre-test briefing for each test and include the following:

   a. Data collection requirements

   b. Data reduction requirements

   c. Data disposition procedures

d. Test countdown data events (calibrations, status, etc.)

e. Test profile, as it affects data collection

f. Data contingency plans

g. "Operator Log" recording requirements

h. Deficiency reporting procedures

2. Define red-line data criteria and develop associated procedures

3. Develop test countdown of the following data events:

   a. Calibrations start/stop

   b. Status checks (equipment manned, ready, etc.)

   c. Contingency points

   d. Coverage requirements for each data collector

4. Check data collection for completeness after each test.

5. Check data distribution after each test against data distribution requirements.

6. Follow-up on data disposition as project progresses.

7. Document all completed data events and deviations in data support log.

8. Provide required information to Test Director for Test Report.

9. As project is completed, document the provided data support and file in project support file.

10. Recommend what data should be retained for future reference.

11. Coordinate data retention requirements with Test Director.

12. Document final data disposition and retrieval methods and file copy in project data support file.

Chapter 10

RANGE INSTRUMENTATION

1. INTRODUCTION

The satisfactory completion of any test program requires the availability of adequate testing resources (facilities and services). The cost of these resources is frequently quite high. If provided solely by the test agency for its exclusive use, the cost would be many times that of the basic program. It is because of this fact that general purpose test facilities (laboratories, test tracks, test ranges, etc.) are vital contributors to test missions and their maximum use represents the most economical overall advantage to the Air Force.

This section provides a general introduction to the services offered by DOD test ranges.

A test range is generally a designated geographical area remote from any areas of population or significant civil activity. Most ranges are under cognizance and control of one of the major Military Departments.

Testing activities, depending upon the range, can include singularly or in combination, missiles, ordnance, aircraft, remotely piloted vehicles, avionics systems, electronic systems and various ground vehicles.

Each range provides instrumentation, facilities and services to support the various tests. Test data is gathered from the range instrumentation and is normally processed by a central data reduction facility. This data, when reduced and in final form to facilitate test analysis and evaluation is submitted to the range user.

Twenty-six high use test and evaluation activities have been designated as elements of the DOD T&E Facilities Base (TEFB). The TEFB over which The Office of the Secretary of Defense (DDR&E) exercises management monitoring of the planning, programming and budgeting consists of extensive land and sea testing areas and major testing facilities which are set aside for specific purposes. TEFB facilities, instrumentation, technical test resources and data processing capabilities are of planning interest to all test and evaluation programs. The DOD Component exercising management responsibility and location of each of the 26 elements of the TEFB are listed in Table 10-1.

Use of most of these ranges, particularly those designated as National Ranges, involve common OSD prescribed funding policies and use of standard test planning and operation formats jointly developed through the Range Commanders Council, Universal Documentation System (UDS). Guidance concerning the policies, procedures and use of the UDS may be obtained from the Secretary, Range Commanders Council, White Sands Missile Range, New Mexico. Other test ranges are generally classified as "Service" ranges and are controlled by the responsible DOD component following established policies and procedures.

TABLE 10-1. MAJOR ELEMENTS OF THE DOD TEST & EVALUATION FACILITY BASE

ARMY

Arctic Test Center, Fort Greely, Alaska
Tropic Test Center, Fort Clayton, Canal Zone
Yuma Proving Ground, Yuma, Arizona
Jefferson Proving Ground, Madison, Indiana
*White Sands Missile Range, White Sands, New Mexico
*Kwajalein Missile Range, Huntsville, Alabama
Electronics Proving Ground, Fort Huachuca, Arizona
Dugway Proving Ground, Salt Lake City, Utah
Aberdeen Proving Ground, Aberdeen, Maryland

NAVY

*Pacific Missile Range, Point Mugu, California
Atlantic Undersea T&E Center, Andros Island, Bahamas
Naval Air Test Center, Patuxent River, Maryland
Naval Air Propulsion Test Center, Trenton, New Jersey
Naval Air Test Facility, Lakehurst, New Jersey
Naval Aerospace Recovery Facility, El Centro, California
Naval Weapons Center, China Lake, California
Atlantic Fleet Weapons Range, Roosevelt Roads, Puerto Rico

AIR FORCE

*Space & Missile Test Center, Vandenberg AFB, California
*Eastern Test Range, Patrick AFB, Florida
Satellite Control Facility, Sunnyvale, California
Arnold Engineering Development Center, Tullahoma, Tennessee
Flight Test Center, Edwards AFB, California
Armament Development & Test Center, Eglin AFB, Florida
Air Defense Weapons Center, Tyndall AFB, Florida
Special Weapons Center, Kirtland AFB, New Mexico
Tactical Fighter Weapons Center, Nellis AFB, Nevada

*Designated National Ranges for Joint Access.

It is also appropriate to point out that although most of the elements of the TEFB are highly structured and well instrumented they often do not lend themselves to supporting specific operational tests. This could be because of size limitations, non-representative terrain and natural features, heavy DT&E workloads and/or other factors. In such cases, other locations must be identified. For example, the classical COMBAT HUNTER tests were accomplished at Fort Riley, Kansas. In this case, the test support instrumentation had to be taken to the location.

The key point is that the appropriate location for the planned operational tests is selected by a host of criteria such as availability, size, representative natural features, economy, logistics, support instrumentation and others. If an element of the TEFB can meet the majority of the criteria it should be used. In essentially all cases such selection assures less problem for the Test Director.

2. NON-TECHNICAL RANGE SERVICES

In general and in addition to range instrumentation, each range provides a variety of services to the range user to support his test program. For example, crane facilities, fork lifts, warehouses (including some with climatic control), ammunition bunkers, ground target areas, launch pads, airbase support, impact areas, drop towers, photographic laboratories, communications, machine shop, repair facilities and other services are usually provided. Further, such physical accommodations as food and housing services for visiting test personnel are available.

Range Safety is a vital service provided for protection of not only range and test personnel, but also surveillance of areas adjacent to the range in the event of malfunctioning and thus errant test devices posing a danger to these areas.

3. RANGE INSTRUMENTATION

Test Range Support Instrumentation can be viewed as falling within one of the following system categories:

a. Metric Data (also termed "TSPI" or Time Space Position Information) Systems.

b. Engineering Sequential or Documentary Systems.

c. Telemetry Systems

d. Range Timing Systems

e. Range Support Systems (including range Communications, calibration, meteorological, recovery, frequency monitoring and geodetic control).

Metric data measurement, or TSPI is normally accomplished by either optical tracking instrumentation, radar or multilateral position measurement systems.

A Metric System (or Tracking System or TSPI System to introduce alternate terms) acquires a target (or multiple targets), locks on automatically and proceeds to "track" it. Data outputs from the system subsequently plot the trajectory or path of the test object whether it be a missile, projectile or aircraft. The derived track in terms of range, azimuth and elevation measurements, when coupled with timing pulses, will provide recorded data that can be analyzed after the test flight or operation. For example, a missile/target miss distance or "score" could be computed from appropriate data derived from a Metric System.

Various types of tracking radars are often available, from the World War II SCR 584 to the AN/FPS-16, AN/FPQ-6 and the AN/MPS-36. The AN/FPQ-6 family provides the greatest accuracy.

Multilateral position measurement devices are generally used to provide aircraft position data when more than one target is to be tracked. A co-operative beacon or transponder is installed in the aircraft with ground sensors installed along the aircraft or target path. The sensor outputs are routed to a central control area for processing and display (including real time). The number of ground sensors and of course their inherent system capability determine the accuracy of the metric data obtained for the range user.

Optical devices are used when a clear line of sight exists to the target. These devices are available from basic manually aimed telescopes on mounts to remotely controlled optical tracking systems capable of being slaved from other devices such as a tracking radar or another optical system.

Cameras are generally installed on the optical instrument for data collection. Timing pulses are recorded on the film for time correlation of each photographic frame of data. Each data frame also records the azimuth and elevation angles of the tracked object as measured from the instrument.

Optical instrumentation may be placed in two broad classifications: metric optics and engineering sequential/documentary photography. Metric optics are concerned with acquisition of precise, measurable information regarding flight characteristics of missiles, ordnance and aircraft, such as position, and attitude (yaw, pitch and roll). Engineering sequential and documentary photography includes the recording of significant test events (e.g., bomb release) and for the preparation of training films and production of other film sequences.

The optical instrument generally used for metric data is the cinetheodolite. This device records test flight "pictures" and azimuth and elevation data and timing codes on a single film so that velocity and acceleration can be accurately computed. As mentioned previously, optical instruments in this category can be slaved to (and thus "steered" by) a tracking radar, that is, they can be remotely aimed at the target by the positioning of the radar antenna.

A telemetry system consists of a sensor or group of sensors which are coupled to a telemetry transmitter on the aircraft, missile, etc., under test. The sensors are arranged to measure temperature, pressure, voltage, current, and other physical parameters pertinent to the planned test. These outputs serve to modulate the telemetry transmitter. This equipment is installed in the aircraft, projectile or other object under test whether airborne or on the ground.

The ground based telemetry reception equipment, installed at the various instrumentation sites on the range, receives the transmitted data, and records the information on magnetic tape or paper chart along with range time codes. These recordings are made available for further data reduction and analysis by the range user. The final product to the user is the parameter (temperature, control surface position, etc.) plotted and/or tabulated as a function of elapsed test time, i.e., time of actual occurrence.

Timing is generally a central system supporting all instrumentation. A central time standard calibrated to a master clock or National Bureau of Standards Radio Timing Signal transmitted by WWV/WWVH, is used for local calibration of the central timing system.

Timing signals are routed throughout the range via telephone lines and/or microwave links. In addition, timing information can be transmitted to test aircraft for correlation of data being obtained from the airborne instrumentation.

4. RANGE SUPPORT SYSTEMS

Range Support Systems are those used to facilitate proper operation of range instrumentation and to assist range users in a circuitous manner that does not yield direct test data. For example, many diverse radio frequency signals are present at and near most test ranges despite their remote locations. These signals must be categorized and continually monitored to prevent RF interference and, if present, to rapidly establish the source of interference and correct it without excessive loss of critical and expensive range time. This process uses the range's Frequency Monitoring and Interference Control System. A second example would be the highly important range communication network that serves to correlate the many support elements of any given test operation.

In addition to the instrumentation described, most ranges utilize TV cameras and associated components, portable recorders for both video and digital data. Airborne instrumentation packages are also available for use in test and supporting aircraft.

Another support function directly related to various tests is the Meteorological Group. This system provides data from surface and upper air observations of atmospheric conditions to range users and the range data reduction facility as required.

One or more surveillance radar systems, depending upon range size, are generally utilized to monitor air space over the range and for certain off-range areas. This surveillance, in addition to providing vectoring information to test and support aircraft, is of value to Range Safety in detection of any outside "intruder" civilian or military aircraft in restricted areas during tests.

Communications, scheduling, planning, maintenance, power distribution, instrument calibration and safety are but a few of the functions and services which are not specified by range users, but which are essential to the operation of the range. Internal requirements encompass all areas of support necessary to testing. While the basic function of a range is to provide data and other support to meet range user requirements, there is also a definite obligation to provide this support safely and economically. Additionally, there are secondary requirements which must be met in obtaining the user's data. In general, the users do not specify requirements in these areas, but they are nontheless essential to the operation of the range.

5. USER REQUIREMENTS SUBMISSION PROCESS

In the case of established ranges, users are normally requested to submit program descriptions, schedules, and requirements to the range at the earliest possible date to insure adequate lead time for preparation of support. Because the lead time for an instrumentation system acquisition program is quite long, short lead time documented workload must be supported with then existing range capabilities. For this reason, the ranges generally encourage the users to establish their requirements as far in advance of the test project start date as possible to aid in economical range support planning.

Discussions with range users are normally formalized in the Range Commanders Council's (RCC) System (UDS) in which requests for support from the users, and statements of supportability from the range are exchanged. The Operations Directive, the final document of the series, becomes the actual support directive. Conferences are held with the range users to ascertain the validity of program requirements and to inform them of support capabilities available to them. The various submitted documents of the series showing who prepares them and in what sequence is shown in Figure 10-1. Not all test ranges (even those of the TEFB) use the RCC UDS, per se. Some use modified versions specifically tailored to their own unique test support capability. The important point is that all ranges of any consequence must necessarily employ

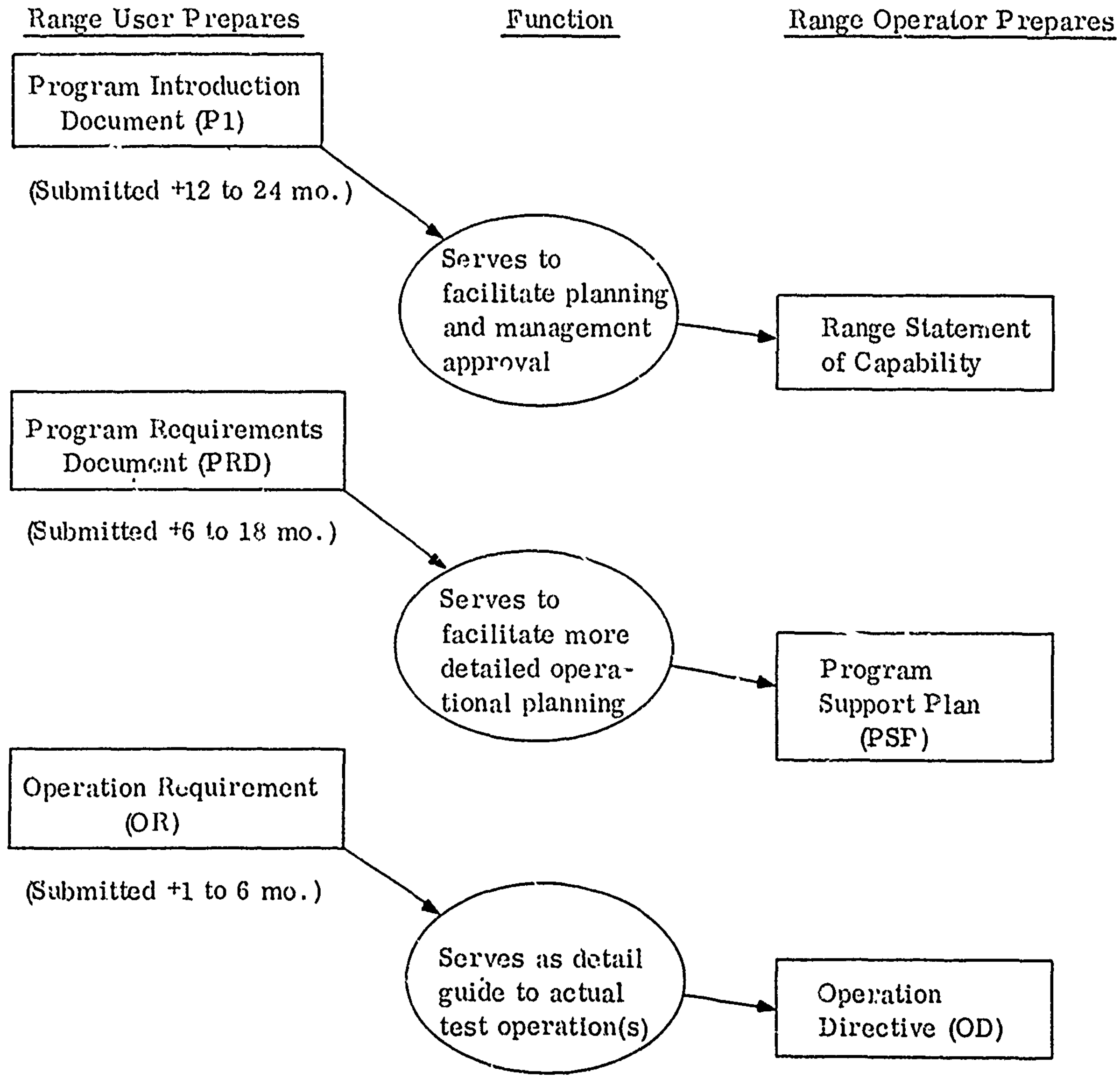


Figure 10-1. Summary of required range documentation.

some method of structured, disciplined written "communication" with their users. All Test Directors, once a given range (or ranges) has been selected for his tests, should immediately familiarize himself (and his applicable staff members) with the methods used at that range for levying the test support (i.e., range user) requirements. This is, of course, best accomplished by a visit to the range and obtaining a copy of their "Range Users Handbook" which, again, most structured ranges maintain and provide to users.

6. TEST RANGE ORGANIZATIONAL STRUCTURE

A generalized structure of range instrumentation can be represented by Figure 10-2. It is shown to demonstrate the basic support elements of any reasonably equipped range.

"Instrumentation Systems Structure", represents the total capability. This in turn, can be subdivided into three elements, called Data Collection Systems, Data Control and Instrumentation Support System, and Test Support Systems. These three elements can be further subdivided into different System Levels, with each system in a particular level contributing its unique functional capability to the next higher level system of which it is a part.

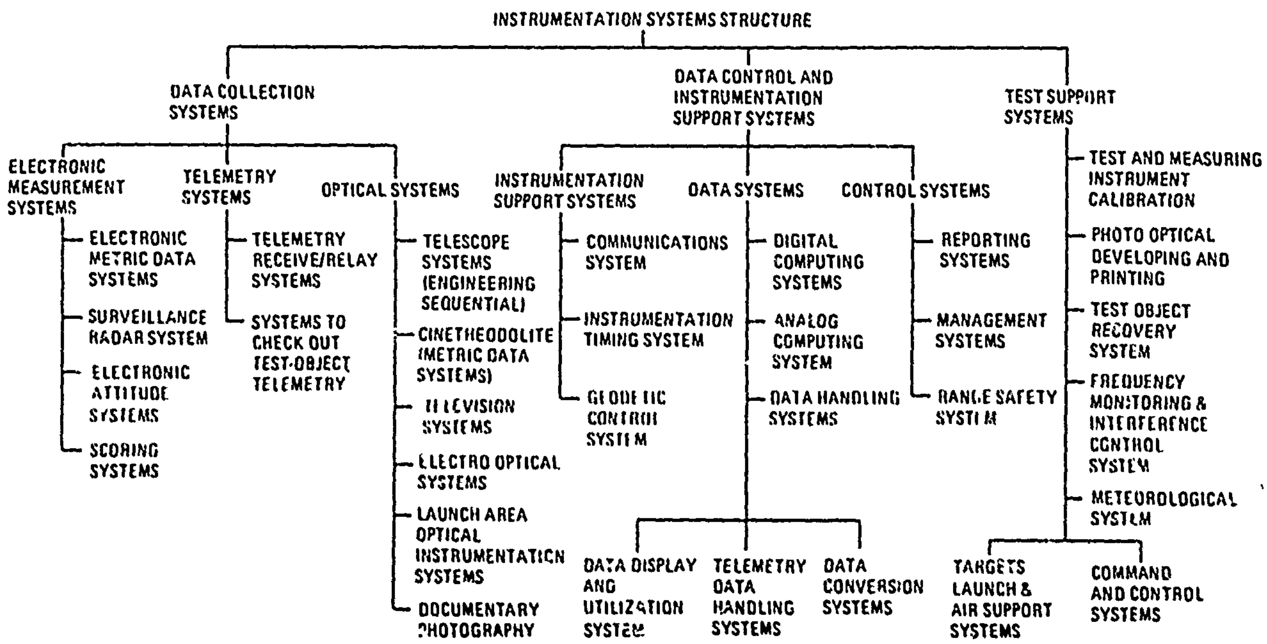


Figure 10-2. Instrumentation systems structure.

The Data Collection Systems sub-structure represents the ultimate objective of any range's effort, the collection of data, or more generally the gathering of information. The Data Control and Instrumentation Support Systems sub-structure represents the mechanism for applying the collection resources to the workload, and the Test Support Systems sub-structure represents additional items of support capability necessary for the range to operate the data collection systems and to function as a range.

Sub-elements of particular interest are discussed below.

Metric (or TSPI) Measurements

Again, it is noted that "Metric Data", "Time Space Position Information Data", or in some cases "Electronic (or optical) Trajectory Measurement Data" Systems are, for all practical purposes the same thing. These are all terms used by the test range community with "Metric Data Systems" being the most common.

Metric measurements include (a) Trajectory Measurements (i.e., the time history of the flight path of a missile, bomb, aircraft, etc.) and (b) Attitude and Event Measurements.

Trajectory data are time-correlated position, velocity, and acceleration data which describe the physical motion of the center of mass of the vehicle under test. In practice, trajectory measurements are almost always angle and/or range measurements from which position data are computed. Velocity and acceleration are usually obtained as derivative data from the position information. Thus, velocity and acceleration measurements are usually not measurements at all but are mathematical derivations. In general, they are limited by the precision of the position data from which they are derived. It follows that the requirements for position data are most representative of the total trajectory measurement requirement.

Attitude measurements are time-correlated angle and angular rate measurements of the orientation of a set of body-fixed axes with respect to the velocity vector or a set of earth-fixed axes. For convenience, event time measurements and miss distance measurements are usually included in this category. These measurements are made either with ground-base instruments (almost always optical instruments) or with vehicle-borne sensors, the outputs of which are telemetered to the ground. Real-time missions usually require real-time attitude data. On-board sensors are the only practical source of real-time attitude information.

Figure 10-3 illustrates metric requirements for several typical USAF tactical system tests.

Representative Category of USAF Tactical System Testing	Trajectory			Attitude (Pitch, Yaw, Roll)	
	Position	Velocity	Acceleration	Attitude	Attitude Rate
Air-to-air and Surface to Surface	High-Medium (Launch, Booster Intercept)	Very High-High (Boost, sustained flight)	Very High-High (Launch, Boost)	Medium (Boost, Intercept)	Low (Boost)
Dispenser and Bomb Drops	Medium (Launch, Terminal)	Medium (Launch Terminal)	Usually not required	Usually not required	Usually not required
Target Systems Drone	High-Medium (Boost, Sustained Flight, Terminal)	Medium-Low (Sustained Flight)	Low (Boost	Low (Sustained Flight)	Usually not required
Equipment Components or Subsystem	High-Medium (Sustained Flight)	High-Medium (Sustained Flight)	High (Launch-Boost)	High (Sustained Flight)	Usually not required

LEGEND FOR MEASUREMENT PRECISION REQUIREMENTS (Nominal)

	Very High (VH)	High (H)	Medium (M)	Low (L)
Position (ft)	<1	1-10	10-50	>50
Velocity (ft/sec)	<0.5	0.5-2	2-20	>20
Acceleration (ft/sec^2)	<0.5	0.5-2	2-20	>20
Attitude (degrees)	<0.5	0.5-2	2-10	>10
Attitude Rate (degrees/sec)	<1	1-10	10-45	>45

Figure 10-3. Typical metric (TSPI) requirements.

Surveillance Radar Systems

An Airspace Surveillance System provides airspace coverage for range and certain off-range areas. Under ideal circumstances, the system can provide manual plotting information on multiple targets.

Television System

Closed circuit Television System provides viewing of remote and hazardous areas in addition to providing an instrumentation operation aid. The majority of the systems are utilized in climatic chambers, propellant handling areas, test stands, and other hazardous or remote areas.

Communications and Data Transmission

The Communications and Data Transmission System, using all standard modes, provides for transmission of voice, test data, timing, and all other types of signals to required locations on the range (instrumentation sites, launch areas, aircraft, etc.).

Instrumentation Timing

The Instrumentation Timing System generates and distributes coded time formats to define a time base for every range instrument, thus providing correlation of all collected data. Timing accuracy is affected by the generating means, the distributing means (wire line, radio, cable, etc.) and by the recording means. Systems generally provide an on-range timing correlation accuracy in the order of ±50 microseconds under optimum conditions for select instrumentation sites.

Geodetic Control System

The Geodetic System locates and orients range launch points, impact points, and instrument sites/stations with respect to the reference geoid and local coordinate systems including range user requirements for such services.

Data Systems

Both real-time and non-real-time data analysis and processing capabilities are normally provided. Systems include real-time data inputs, real-time data outputs, real-time telemetry data input, teletype input and output, and remote terminal input/outputs. Machine processing of real-time inputs is unique according to the specific real-time mission. Non-real-time processing requires performance of the following functions:

Computation of final solutions.

Generation of tapes and listings for subsequent review and analysis.

Printing and distribution of final data reports to range users.

Data Conversion Systems

Raw field data must be corrected for system errors and converted to engineering units to be meaningful to the range user. A range's capability can include both real-time and non-real-time, data correction, conversion and formatting for computer entry.

Test and Measuring Instrument Calibration

Most ranges operate a standards and calibration laboratory which not only executes a local calibration program in the interest of quality assurance, but which also provides services to Range users. Special services provide for integrity of measurement and transduction in areas of optics and gauging, mechanics and dynamics, photometry and radiometry, electricity and magnetism, radio and microwave technique, materials testing and radio-logical survey.

Photo-Optical Developing and Printing

The range produces a great volume of data records on photographic film or photosensitive paper which must be processed within a few days after exposure. The facility processes film and provides other services such as loading, issue, storage, reproduction, etc.

Test Object Recovery System

Most recovery operations are dependent upon visual search by ground and/or airborne observers. Radar vectoring, or data from other trajectory measuring systems, is used to position ground and/or airborne observers in the approximate impact areas. Once located and recorded, the test vehicle (or debris) is returned to range user or disposed of as required.

Frequency Monitoring and Interference Control System

The FMIC activity monitors all frequencies used in connection with range testing. Spectrum utilization by all transmitters, receivers, and antennas, as well as other propagation phenomena are evaluated. Control measures are applied when undesirable radiation could interfere with testing.

Meteorological System

The Meteorological System provides data from surface and upper air observations of atmospheric conditions to range users and the range data reduction facility as required. A network of sites are maintained on range to gather meteorological data off-range testing.

Targets, Launch and Air Support Systems

Rockets, aircraft, and drones are supplied either by the user or the range (depending upon pre-established agreements) to provide aerial targets for range users. Targets may carry augmentation in the form of radar beacons, optical lenses, corner reflectors, infrared generators, and smoke. Ground targets are usually project-peculiar facilities and are sited in various range impact areas.

Command and Control System

The command and control system provides the means to implement decisions made during tests. These decisions concern aircraft vectoring and control, missile guidance, and missile flight safety commands.. By its nature, the system must operate in real-time and must receive real-time data from the measurement systems.

7. SUMMARY

It can be seen from the foregoing that significant assistance to a Test Director is available from structured test support ranges. Further, a method exists to allow him to communicate his support requirements to the range and for the range to specifically respond as to what it can provide.

This overall subject has been treated only generally, the most important point to be made is to establish early communication with the specific range that is selected to support any test or test series.

Chapter 11

TEST PLAN

1. INTRODUCTION

The Test Plan is the vehicle which translates a test concept, statistical/analytical test design, and Test Directive into "real world" resources, procedures, and responsibilities. It reflects the Test Officer's method of accomplishing the test objectives within constraints of time, budget, manpower and material resources. The size and complexity of the test program is determined by the nature of the system being tested and the type of testing that is to be accomplished. OT&E of major weapons systems may require large numbers of separate tests to satisfy test objectives and the associated Test Plan may therefore be several volumes in length, while other testing may be well defined by a relatively brief Test Plan. The Test Plan format described herein is intended to be minimum, basic, and applicable to a very wide spectrum of Air Force operational testing. It calls out topics which must be considered during the planning process, and which must be contained in the Test Report. Every effort has been made to have the Test Report format "track" the Test Plan format so that a straightforward relationship can be seen between test planning and test reporting.

Existing Major Command Regulations, IOT&E and FOT&E Test Plans, and AF and DOD level requirements were researched during the preparation of the format described in the following pages. These inputs, plus deficiencies noted during a survey of OT&E Reports, formed the basis for the format. It achieves the requirement for standardization, indicates potential for producing a better-quality Test Plan, and causes minimum disruption to formats now in use.

2. TEST PLAN OUTLINE

The Test Plan format comprises three major subsections:

a. The Main Body containing thirteen numbered headings.

b. The Basic Annexes containing detailed information supporting the Main Body of the Test Plan.

c. Detailed Test Procedures Annexes containing definitive instructions for performing sub-tests within the overall test program.

The Main Body and Basic Annexes comprise the Basic Plan. The size and complexity of the overall test program determines the number of Detailed Test Procedures Annexes which will be required. Programs requiring the simultaneous testing of numerous subsystems, and programs requiring separate time-phased tests may identify appropriate subsystems or test phases which will have dedicated Detailed Test Procedures Annexes.

The Basic Plan contains all support requirements, overall schedules, and descriptions of the general methods of test for the overall program. When possible, detailed test procedures should be included as annexes at the time the Basic Plan is distributed. For programs of extended duration where it may be impractical to define detailed, step-by-step procedures months in advance, Detailed test procedures should be written and distributed prior to each individual test in the program. The distribution should be made sufficiently in advance of the test start date to allow for coordination, briefing of all participants, and resource acquisition.

The Test Plan format (see Figure 11-1), with instructions for completing the plan elements, are presented on succeeding pages. The following general rules apply for successful use of this format:

a. The numbering system should be kept intact. If any part of the format is not applicable to the test in question, it should be so stated. If non-applicability is not readily obvious, a brief explanation of why the section is not being addressed should be included.

b. This is intended to be a basic, minimum format. If a particular test requires additional information, it should be included under the appropriate major heading. The sequence and content of subsections is not constrained.

c. Provision is made for classified annexes. These should be used when most of the planning information is unclassified, with a limited number of agencies having a need-to-know classified details. If classified information is called for in any section of the format, the classification should be stated in that section, and the actual information placed in an identified classified annex. Within the classified annexes, references should be made back to the appropriate format section. The agency originating the Test Plan should have control over distribution of the classified annexes. They should not be used when not required, or when classification of the entire Test Plan is more practical.

d. Six annexes (Test Design, Data Requirements, Instrumentation, Logistics Support Requirements, Reliability and Maintainability, and Intelligence/ Threat Information) are called out as part of the format. These should be used in every Test Plan, and lettered A through F, as indicated. Other annexes may be used as required to provide detailed information pertaining to any section within the Main Body. These Annexes should be lettered successively, beginning with "G".

e. Detailed Test Procedures Annexes should be listed within the Table of Contents after the Basic Plan Annexes (those described above), and should be designated Annex 1, 2, 3, etc.

f. The term "test item" refers to the subject of the test, i.e., that which is being tested. It may be hardware, tactics, procedures, or doctrine. For simplicity, it is called the test item.

PRELIMINARY PAGES

i. Title Page
ii. Abstract
iii. Table of Contents
iv. Terms and Abbreviations
v.* Related Documents

*The actual number of these pages will be determined by the length of preliminary elements (e.g., Table of Contents, Terms and Abbreviations, etc.).

MAIN BODY

1. Introduction
2. Test Purpose and Objectives
3. Concept of Test Operations
4. Method of Accomplishment
5. Test Schedule
6. Test Management and Organization
7. Responsibilities/Support
8. Personnel
9. Required Test Reports
10. Safety
11. Security
12. Information
13. Environmental Protection

ANNEXES

A. Test Design
B. Data Requirements
C. Instrumentation Plan
D. Logistics Support Requirements
E. Reliability and Maintainability Data Plan
F. Intelligence/Threat Information
G.-Z. as Required

1, 2, 3, etc. Detailed Test Procedures (Name of Test)

DISTRIBUTION:

Figure 11-1. The Test Plan format

3. INSTRUCTIONS FOR USING TEST PLAN FORMAT

The following instructions for use of the Test Plan format are keyed to the format by paragraph number.

PRELIMINARY PAGES

i. Title Page: Existing cover and title page formats in use by the major commands should be retained. Typically, the title page contains:

   a. The Security Classification. State the security classification of the Test Plan document. Include appropriate downgrading and distribution limitation statements.

   b. The Title. State the title of the plan. If the plan is prepared in more than one volume, state the overall title, and subtitle the individual volumes appropriately.

   c. The Number. State the alphanumeric designation by which the document is most commonly referred to by the originating agency.

   d. The Date. State the distribution date of the Test Plan, including at least the month and the year.

   e. The Command. Identify the major command releasing the Test Plan and the operating agency which prepared it.

   f. Approval. Identify who the Test Plan was (1) prepared by, submitted by, and (2) approved by (provide a signature block for approval).

ii. Abstract: The abstract should be written in narrative form and may contain the following information:

   a. The word "ABSTRACT" in capital letters at the top of the page.

   b. The nomenclature of the test item.

   c. The purpose of the test.

   d. A brief explanation of the mission of the test item.

e. The test agency, major support agencies, and test locations.

f. The anticipated time frame of the tests. Indicate optimum start and stop dates, weeks, or months, as appropriate.

g. A brief description of the test approach.

h. A brief description of the data to be gathered.

iii. TABLE OF CONTENTS: The table of contents is seldom used in a plan of eight pages or less. If used, list numbered headings, with the page numbers on which the headings occur. Start the table of contents on a new righthand page.

Separate lists of illustrations and tables should be included if considered essential. The list of illustrations should include the figure number, legend, and page number for each illustration. Abbreviate lengthy legends.

iv. TERMS AND ABBREVIATIONS: Define unusual terms the first time they are used either in the text of the plan or as a footnote. When more than five such terms are used, list them in alphabetical order with definitions in this section of the plan.

Define abbreviations, acronyms, and symbols when first introduced in the text. When more than five are used, include a list (separate from the list of terms described above) of definitions in this section of the plan.

v. RELATED DOCUMENTS: Identify prior or concurrent Test Plans, reports, technical orders, manuals, and other appropriate literature related to the test. Include titles, identifying numbers (including DDC accession numbers when possible), publication dates, security classification, and originating major commands. EXAMPLES: ROC, PMD, TOA, DCP, PMP, PM, Test Directive.

MAIN BODY

1. Introduction.

The introduction should contain:

a. Background: Include details relevant to the origin of the test requirement, explanations of prior or anticipated phases of the test program, and other information which may help management and other participants in their understanding of the test. If applicable, identify system deficiencies and/or problem areas, including unresolved test objectives revealed by prior testing. State actions taken to correct these deficiencies/problem areas and identify retest requirements, if any.

b. Authority For Test: Identify the authorization for the test.

c. Management Agency: State the name and location of the agency having overall responsibility for administration, planning and coordination of the test.

d. Priority: State the assigned command priority and USAF precedence rating.

e. Mission: Summarize the mission and concept of operations of the test item. If the test item is a subsystem, state its function and the mission and concept of operations of the total system. If appropriate, reference other documents containing detailed descriptions of the mission and OPS concepts.

f. System Description: Describe the functional relationships of the major subsystems of the test item. If the test item is a subsystem, describe the functional relationships of the total system at the subsystem level. Use block diagrams if necessary. If the test item is non-hardware, describe flight profiles, formations, existing procedures being modified, and other information to afford a clear understanding of the test item. List significant physical parameters of the test item (e.g., size, weight, frequency, power, packaging, etc.). Identify the differences between the test item and anticipated production hardware. State the anticipated effect of these differences upon testing.

2. Test Purpose and Objectives

State the overall reason/requirement/purpose of the test. This should be a short, concise statement, normally contained in one sentence.

For IOT&E, identify critical issues relevant to the test. These critical issues are contained in the Program Management Directive (PMD), Command Supplements to the PMD, the Program Management Plan (PMP) and the Development Concept Paper (DCP).

Where testing of effectiveness is cited as a test objective, identify those data elements of the system's Availability, Dependability, and Capability that will be measured, analyzed and reported upon.[1]

State the test objectives; the Test Officer will receive the test objectives contained within a Test Directive, Test Order, or Project Order. Depending upon the size of the program, complexity, and degree of command interest, these objectives may be stated in varying degrees of detail. The Test Officer may be given the latitude and responsibility to formulate specific test objectives. These specific objectives must be relevant to resolving the critical issues and feasible, within time and resource constraints. The Test Officer should assign importance factors to each objective. If resources and support are likely to be "bumped" by higher priority programs, the most important objectives should be accomplished first when feasible. Primary and secondary objectives may be designated. Primary objectives are those essential to answering the critical issues of the test. Secondary objectives are those which will yield additional, useful data or training on the system, but are not critical to successful accomplishment of the test.

Typical objectives may be:

a. To determine the capability to accomplish a penetration mission in the presence of -------

[1]See Chapter 6

b. To determine the bounds on employment of ---

c. To determine the compatibility of the test item with interfacing systems----

Test Objectives (Logistics).

Assessments of logistics supportability, reliability, and maintainability are cited as typical major objectives of OT&E in AFR 80-14. Depending upon the latitude of the test objectives given him and the degree of AFLC participation in the tests, the Test Officer may be given the option and responsibility to formulate specific logistics test objectives. Typical objectives may be to:

a. Verify maintenance concepts and procedures.

b. Assess achievement of maintainability requirements and study individual items of equipment for ease of removal or repair.

c. Assess and predict meeting reliability requirements. Comparison of estimated failures to actual failures.

d. Verify AGE adequacy and compatibility including software. This should include an evaluation of Integrated Data Systems if applicable.

e. Examine Technical Orders for accuracy, completeness, and usability.

f. Observe and analyze supply procedures.

g. Analyze throw away/level of repair recommendations.

h. Verify and update estimation of life cycle Cost/Logistic Support Cost Model parameters listed below, if applicable.

(1) Meantime between maintenance action

(2) Maintenance man-hour per action

(3) Percentage of repair action per location (flightline, field shop, etc.

(4) Other items required to be quantified to operate the appropriate Life Cycle Cost/Logistic Support Cost Model.

i Analyze safety aspects of support.

j. Observe corrosion effects.

k. Predict supportability under simulated deployment condition.

l. Analyze configuration management and quality control.

m. Evaluate transportation, handling, and packaging procedures.

n. Evaluate the program/project management organizations' verification and update of Aircraft Structural Integrity Program (ASIP) tests and analyses required by AFR 80-13, and as reflected in the ASIP Master Plan.

o. Evaluate the adequacy of the Non-Destructive Inspection (NDI) program and equipment.

Test Objectives (Training).

Refinement of operational requirements for personnel training is cited as a typical major objective of follow-on OT&E. Depending upon the latitude of the test objectives given him and the degree of ATC participation in the tests, the Test Officer may be given the option and responsibility to formulate specific training assessment test objectives. Typical objectives will:

a. Assess personnel subsystems vs. T.O.'s, manuals, and training

b. Address the determination of, or the adequacy of training requirements for Air Force personnel.

c Determine that selection of personnel to support the program is made from the proper technical career speciality code(s).

3. Concept of Test Operations

Define those aspects of the complete system that will be tested. The complete system will include:

a, The test item, all possible modes of operation, and all environments.

b. Test item operators and training requirements.

c. Maintenance concepts and procedures.

d. Maintenance personnel and training requirements.

e. Logistics requirements and supply procedures.

f. Two-sided scenarios, evaluating the test item in one or more threat environments.

Tests of limited scope.

Many tests will have directed or self-imposed limitations to scope. Not all aspects of the complete system, as described above, will be tested.

Indicate, as appropriate:

a. The type of test (combined DT&E/IOT&E, IOT&E, FOT&E, joint, etc.)

b. The number of test articles to be employed.

c. Overall tactics and techniques to be evaluated.

d. The maintenance concept (i.e., degree of contractor maintenance involved).

e. The logistic support concept.

f. The expected test duration.

g. Special supporting personnel and equipment required (i.e., special intelligence or other unique personnel needs, special sensor readout, unusual data, instrumentation or scoring equipment requirements.)

4. Method of Accomplishment

If applicable, list separate subtests or test phases. Criteria for separating the test program into individual phases may include: completion of data collection to satisfy a test objective, commitment of different resources to support the test, limited availability of specific resources, etc.

For each subtest listed above, identify the objectives to be addressed during that test. If practical, this may be done in matrix form, with subtests listed on one axis and objectives on the other.

Describe the general method of each test, i.e., actions to be taken and procedures to be followed during the actual conduct of the test. See Annexes for a Detailed Test Procedures Format.

Summarize raw data, and the source of that data, to be taken to address each objective. Describe the reduction to be performed on each raw data item and describe the final form of the reduced data.

Include, as appropriate, a summary of missions/sorties/flying hours by material category required to accomplish each test, and profiles describing each mission.

Include a trials/variables matrix which summarizes the combinations of test conditions or variables to be tested during each trial.

5. Test Schedule

Indicate nominal program milestone dates. When practical, use a line chart with calendar days across the top axis and different lines showing the duration and time phasing of the various activities listed below it. Milestone dates may include, but not be limited to the following:

a. Receipt of Test Order, Test Directive.

b. Distribution of Draft Test Plan for review and coordination.

c. Distribution of Final Test Plan.

d. Finalization of support agreements.

e. Procure equipment and instrumentation.

f. Receive equipment and instrumentation.

g. Planning briefings and meeting.

h. Distribution of Detailed Test Procedures.

i. Deployment of supporting units.

j. Commence testing.

k. Periodic reviews.

l. Complete testing.

m. Ship equipment and instrumentation to next test location.

6. Test Management and Organization

Identify the organization having overall responsibility for planning, coordination, and conduct of the test.

Identify key project personnel, to include the OT&E Staff Officer, OT&E Test Director, and OT&E Test Officer.[2] State the organization, office symbol/code, and telephone number (autovon and commercial) for each.

Describe the major command relationships, including AFTEC participation. Include an organization chart showing reporting channels. If appropriate describe the composition of the Joint Test Force, including military departments, in the organization chart.

Identify primary and alternate points of contact for each organization supporting the test. Include the office symbol/code and telephone number (autovon and commercial) for each.

Identify the organization or individual to exercise administrative control over units deployed to support the test. This section may include reporting-in procedures, disciplinary authority, court-martial jurisdiction, etc.

If base support or tenancy agreements exist for all deployed units, they should be referenced but not repeated.

2. See Chapter 1

7. Responsibilities/Support

Identify each participating organization and list the responsibilities, services or equipment to be accomplished or provided by them in support of the tests. The responsibilities of key project personnel may also be listed separately, if required. This section should reflect the degree of AFTEC, implementing command, supporting command, user command and contractor participation in the tests. If necessary, detailed listing of services, instrumentation, etc., should be placed in an annex and referenced here.

Identify procedures and communications necessary to return unused funds, test instrumentation, test items, supporting aircraft, and personnel to the agencies which committed them to the test.

8. Personnel

Identify personnel required to support the test program. This may be done in tabular form, with column headings such as: Job Title, Grade, Air Force Specialty Code, Organization Providing, Period of Utilization, Desired Skill Level, and Number Required. When possible, identify the individuals by name. Personnel accomplishing the test should be of the type and experience similar to those who will employ, maintain, and support the test item if acquired and deployed. If this is not possible, identify differences between the test personnel and operational personnel.

Identify training requirements for test participants. Personnel requiring specialized training may include data collectors, aircrews, test range controllers, system operators, and maintenance crews. Integrated training, including practice runs with all participants playing an active role, may also be required. Identify training schedules and outline the content of the instruction to be provided each group requiring specialized training.

9. Required test reports

This section of the plan should state the required frequency of test reports, where they will be sent, special content/format instructions, and preparation responsibilities. The basic format and content of test reports is described in Chapter 15.

10. Safety

The amount of safety planning required for a test varies with its inherent hazards. Typically, this section of the plan will:

a. Designate a Safety Officer and outline his responsibilities which typically include: (1) briefing all test personnel on safety critical aspects of the test, (2) implementing all prescribed safety measures, (3) monitoring the conduct of the test, (4) investigating and reporting accidents occurring during the test, (5) drafting the Safety section of the Test Plan, and (6) coordinating it with the cognizant Ground Safety Office.

b. Designate a Flying Safety Officer and outline his responsibilities, which are similar to those of the Safety Officer, in flying operations related to the test.

c. Identify safety critical aspects of the test and procedures to be followed with relation to them.

d. Establish responsibilities for investigating and reporting aircraft accidents, in accordance with AFR 127-4.

e. Establish an emergency reaction plan.

11. Security

This section should be written in a manner which does not reveal classified information. Identify, - but do not reveal - classified aspects of the test, to include but not be limited to:

a. Data. Identify classified data that will be produced during the test. State storage requirements, handling procedures, and restrictions.

b. Photography. Identify classified non-data documentary type photography to be produced during the test. Indicate reasons for classification, e.g., external views of classified hardware, assessment of weapons effects, etc.

c. Hardware. List classified hardware items and identify responsibilities for storage and handling.

d. Events. Identify classified events to occur during the test and state the reasons for classification, e.g., bomb release altitude, target acquisition range, etc.

e. Communications. State requirements for secure communications lines and telemetry links.

If all supporting agencies have received a DD Form 254 containing the above information, it should be referenced, but not repeated.

12. Information

State restrictions upon release of information concerning the test. Identify individuals or offices with authority to make news releases. Identify the cognizant Office of Information. If appropriate, identify responsibilities for providing photographic support and transportation for information coverage of the test.

13. Environmental Protection

If an environmental assessment has shown the test to have potential impact upon the environment, identify actions and procedures planned to minimize, mitigate, or neutralize any adverse effects of the test. These actions and procedures should comply with AFR 19-1 and 19-2. Examples of projects that could adversely affect the environment include projects involving chaff, radiation, lasers, smoke, chemical or biological weapons, and noise. including sonic booms.

Identify data that will be collected to assess the impact of the test item upon the environment.

The Test Plan should contain one of the following two statements, as appropriate:

a. "AFTEC has determined that the conduct of this evaluation will not have an adverse effect on the environment."

b. "All pertinent environmental factors have been considered, and AFTEC has determined that the designed conduct of this evaluation will inflict minimal hazards to the environment."

A. Test Design

Recorded information on the test design is important for several reasons:

1. A written record is useful for retrospective analysis of many tests to determine type and quality of test design used and how to improve test design.

2. Also test design information is essential to future users of the test data in order to know the quality and limitations of the data and thus insure that the data is not misused.

3. Review and approval of the Test Plan may require establishment of confidence in the validity of the test design, especially as regards the handling of different operational variables in the test.

Therefore although it is essential that a part of the Test Plan adequately describe the test design, it is not necessary that all parts of the Test Plan be distributed to all recipients.

Describe the analytical/statistical approach being used to answer questions posed by the test objectives. This description should include (as appropriate) but not be limited to:

1. Listing of independent and dependent variables for each trial or group of trials (as applicable) and description of the statistical handling of each independent operational variable (i.e., investigated at different levels, blocked, randomized, or held constant; factorial or nonfactorial arrangement).

2. Basis for establishing scope of the test.

3. Rationale for selection and description of scenarios to be used.

4. Factors driving degree of operational realism (including physical environment, operating and maintaining personnel, test hardware).

5. Statistical basis for replication.

6. Prediction of uncertainty in results.

7. Effect of resource limitations on test design.

8. Use of modeling and simulation.

B. Data Requirements Summary

This annex should include, but not be limited to:

1. Sample data sheets, questionnaires, debriefing forms, etc., that will be used to collect data during the tests. If standard forms are to be used, and if their inclusion in the plan would be impractical, their numbers may be listed.

2. A summary of all raw data items to be produced during the test. This summary should include:

   a. An identification of data requirements for each subtest or trial. This may be done in matrix form by listing data items (e.g., documentary photography, TSPI for specified aircraft, debriefings, etc.) down one axis, and listing subtests or trials across the other axis. X's would be used to indicate requirements. Detailed data requirements for each subtest should be identified in the detailed Test Procedures Annexes.

   b. An identification of the agency responsible for providing each data item listed above. This may be done in matrix form by listing data items down one axis and supporting agencies across the other axis. X's would be used to indicate responsibilities.

C. Instrumentation Plan[3]

1. List Project Unique and/or Project Furnished instrumentation required to support the test. Identify:
   a. The nomenclature, manufacturer, and model number.
   b. The number required.
   c. Location where required.
   d. The source providing it.
   e. Responsibilities for repair and calibration.
   f. Need date and duration of its requirement.
   g. The utilization (e.g., data recording, telemetry package, system checkout, etc.)
2. If appropriate, include a block diagram explaining the instrumentation setup.
3. If range support is required, identify:
   a. The Test Range supporting the test
   b. Range instrumentation that will support the tests. Include a map or diagram defining the coordinate system to be used for the data, vehicle positions, ground station sites, and flight paths.

---

3. Test instrumentation may have limited application in some types of OT&E. Instrumentation is commonly thought of as a DT&E associated function, and many testers are loathe to "hang wires" on a system undergoing OT&E. However, instrumentation not physically connected to the test item is involved in many types of OT&E. Data recording devices, communications, and TSPI instrumentation must all be considered by the Test Officer. Even though he may have little or no role in their selection, acquisition, or operation, he must be aware of their limitations and capabilities to meet his requirements. See Chapter 10 for a discussion of range instrumentation.

c. Test range coordination documentation (CG, PI, PRD, etc.) defining Range participation and support.

d. If a test is to be run on an area other than an established range, this fact should be stated and arrangements which have been made or will be made should be explained.

List radiating sources (and frequencies) associated with the test item and project peculiar instrumentation. Identify frequency management responsibilities for each test location. Reference approved assignments and clearances.

D. Logistics Support Requirements

Logistics requirements vary considerably according to the test and test item. This annex may include, as appropriate:

1. Supply procedures, including organization codes, delivery destination codes, requisitioning, bench stock issue, Not Operationally Ready Supply Support (NORS), special tools availability, Due-in From Maintenance (DIFM) items, listings of critical items, and supply requirements data collection forms.

2. Maintenance support requirements including data collection forms for operational readiness, reliability/maintainability, material performance, phase inspections, etc.

3. Munitions maintenance support requirements.

4. Explosive Ordnance Disposal (EOD) support requirements.

5. Base support requirements, including billeting, messing, transportation, postal services, accounting and finance, security, TDY funding policies, medical, POL, etc.

6. Shipping and transportation, including schedules, listings of materials to be shipped, estimated cubage and weight, departure points and destinations, methods of shipment, numbers of personnel, etc.

If base support/tenancy agreements or range documentation contain the above information, they should be referenced here but not repeated.

E. Reliability and Maintainability Data Plan

List R&M data elements, identified by AFLC and/or AFSC, to be recorded during the test.

Identify the data collection system, e.g., AFM 66-1, System Effectiveness Data System (SEDS), etc., and/or data collection forms to be used for collection of these data elements.

F. Intelligence/Threat Information

Describe the threat which generated the requirement for the test and/or the hostile environment in which the test item is to function when operationally deployed. If this information is classified, this Annex should be bound and controlled separately, with distribution limited to those with the required Security clearance and need-to-know. If the test item is to counter a particular weapons system, describe that system in terms of quantitative performance parameters. Describe the hostile environment, to include (as applicable):

1. Types and intensity of hostile ground fire
2. Electronic counter measures
3. Ground-to air missile defenses
4. Air-to-air missile defenses
5. Enemy interceptor capabilities
6. Radar defenses
7. Air-to ground ordnance
8. Enemy ground forces.

DISTRIBUTION

Annexes G-Z should be used as required to provide additional information relating to any section within the main body of the test plan. List agencies and offices to receive the plan. Identify the number of copies that each will receive.

4. DETAILED TEST PROCEDURES FORMAT

Separate detailed test procedures should be written for each test called out in Section 4 of the Test Plan. Once distributed and approved, they become Annexes to the Test Plan, and are numbered consecutively according to the order in which they occur chronologically. These procedures should be written comprehensively. They must contain all the information needed to perform and repeat the test. To the maximum extent practical, every significant action should be anticipated and listed. It is understood that routine actions and procedures involved in flying an aircraft, operating support radars, performing maintenance, etc., cannot be prescribed in the procedures. Every action affecting measured and recorded test variables and parameters should, however, be included in the detailed test procedures. The Test Officer should consider implementing the following practices:

a. Establish a review and concurrence cycle for these detailed test procedures. Copies should be distributed to each individual or organization participating in the test sufficiently in advance to allow for required revisions. These revisions, if any, should be minor in nature since support and method of test have already been agreed to in the Test Plan.

b. Establish an "action item" and open issues file. Preprinted forms, with spaces for the date assigned, assigner, individual responsible for performing the action, nature of the action to be performed, and due date, should be present at all planning and coordination meetings. This is the best way to insure that problems do not "fall through the cracks" and surface during the test. The Test Officer should maintain a master file of action items. One agenda item for the pre-test briefing should be a review of action items to insure that all have been successfully completed.

c. Establish a technical assistance panel of Test Officers and/or engineers for the duration of the test program, each with responsibility in their area. Areas of responsibility are assigned as needed and appropriate. Typical areas may include:

Test instrumentation

Test item engineering

Test item maintenance

Ground safety

Flying safety

Operations analysis

Data collection

Data processing

Test range operations

Information

d. Establish test log books for the duration of the test program. Typically, a dedicated log is assigned to each System Project Officer and to major equipment and instrumentation systems involved in the test. All actions affecting the system are entered in the log. The contents may include:

Changes to detailed test procedures

Procurement and transfer of accountability actions

Maintenance actions

Hours of operations

Packaging and shipping actions

Etc.

e. Insure that the same project peculiar/furnished equipment, instrumentation, and cables (manufacturer, model no., and serial no.) are used for the duration of the test program. Use of unfamiliar instrumentation may introduce unknown variables into the test system in the form of slightly different impedance characteristics, unknown stabilities and regulation, different nonlinear gain and loss characteristics, etc.

f. Insure that control is maintained over personnel used in the test program. Uncontrolled shuttling of test participants tends to destroy test program continuity and makes assessment of training requirements difficult.

The format (i.e., major elements and their ordering) of the detailed test procedures is shown in Figure 11-2.

TITLE

A. TEST PURPOSE AND OBJECTIVES

B. PARTICIPANTS

C. SYSTEM CONFIGURATION

D. DATA

E. PRE-TEST PROCEDURES

F. TEST CONDUCT

G. POST TEST PROCEDURES AND DATA REDUCTION

H. REPORTING REQUIREMENTS

I. SCHEDULE

APPENDICES

Figure 11-2. Detailed test procedures format.

The following instructions for use of the Detailed Test Procedures format are keyed to the format by paragraph numbers as shown in the Figure 11-2.

TITLE

State the name of the subtest as listed in Section 4 of the Test Plan.

NAME OF SUBTEST

A. TEST PURPOSE AND OBJECTIVES

1. State the purpose of the subtest.

2. Identify which objectives of the overall test program are to be addressed during the subtest.

3. Identify particular objectives of the subtest, if different or more specific than above.

B. PARTICIPANTS

1. List the personnel actively participating in the test. Identify their organization and state their job title or function (e.g., Test Director, Data Collector, etc.)

2. Identify responsibilities for notifying all participants and concerned agencies of the upcoming test.

C. SYSTEM CONFIGURATION

1. State the location of the subtest.

2. Describe the configuration of the test system. Do not repeat the System Description called for in Section 1 of the Test Plan, but define, as appropriate:

   a. The scope of the test item. Is this a system test, or a limited test of subsystems?

b. The physical configuration of the test system (test item, cameras, tape recorders, mounts and fixtures, etc.) Include diagrams as required.

c. The instrumentation configuration, including, as appropriate, block diagrams, cabling and wiring diagrams, transducer locations, etc.

3. List all instrumentation to be used in the test. Include model numbers, manufacturers,. and serial numbers.

4. List all AGE to be used in the test. Identify test item dedicated AGE. Include model numbers, manufacturers, and serial numbers.

D. DATA

1. Exclusive of logistics data elements, identify measurements and other data that will be taken to address each objective. For each measurement shown identify:

a. Real-time display requirements, including location and method of display, if applicable.

b. The recording medium, including channel assignment, if applicable.

c. The source of the data, e.g., Range FPS-16 radar, Range KTH-53 cinetheodolite, magnetic tape voice track, etc.

d. The expected, or acceptable range of values for the measurement.

e. How verification will be accomplished.

2. Identify logistics data elements to be collected, and the data collection system and/or forms that will be used.

E. PRE-TEST PROCEDURES

1. Define equipment installation and instrumentation set-up procedures and responsibilities. Indicate timing of each action, if significant, and timing

of the entire procedure with relation to commencement of testing. Include a description of equipment and instrumentation readiness checks to be performed prior to test item turn-on. This description should be accomplished as shown below in Section F of this format.

2. A pre-test briefing should be held, with all, or selected test participants in attendance. This briefing should be scheduled to occur as close to the start of the test as is practical. The optimum time would be between instrumentation set up and readiness checks, and commencement of testing, thus allowing system engineers and technicians to report on the status of their areas of responsibility. The Test Director normally conducts this briefing and works from a checklist of systems' status and support readiness. This section should state the time and location of the briefing, a brief description of the agenda, and required attendees.

F. TEST CONDUCT

Several methods may be used to describe actions and procedures occurring during the conduct of the test:

1. If time phasing of the actions is not critical to the test, as may be true in testing electronics and communications gear, the following checklist method may be used:

   a. State the initial condition of all switches and controls on the test item and test instrumentation.

   b. List all following steps in this manner:

STEP NO.	EVENT OR ACTION (EXAMPLE)	REMARKS
1	Turn radar scope power switch "ON"	Allow a 5 minute warm-up and stabilization before proceeding
2	Adjust gain vernier until display appears etc.	Display should be 2 cm high on scope graticule.

2. If time-phasing of events during the test is critical, a test countdown, or "timeline" may be used. Basically, a countdown may be constructed by deleting the "step no." column in the checklist example above, and substituting "T" + or - time references in hours, minutes, or seconds indicating the timing of events and actions. The T-O event or action is usually chosen to mark the start of the test, with T-times indicating system readiness checks, calibrations, aircraft arrival times on station, etc. An indication of timing criticality for the events should be included. If the validity of test data will suffer from non-precise timing, it should be so stated. Some events, however, may be less time critical, for example, equipment turn-on times, calibrations, take-off times, etc. Include an indication of "Go-No Go" decision points and critical events. These are points where the test will be terminated or put in a "hold" mode if the preceding events have not occurred as planned.

3. Many variations of the time line are possible. A recommended approach is shown in Figure 11-3.

Parameters/ Attributes*	Subtest Profile			
	Taxi	Take Off	Climb	Cruise
ALTITUDE, OBJ A				
Range		0-200 ft	200-10K ft	5K-30K ft
Accuracy		10 ft	50 ft.	100 ft
Precision		5 ft	10 ft.	10 ft
FOC		Continuous	Continuous	
Coverage	None			
UF		LOW	LOW	
SPEED, OBJ A				
Range	0-40K	0-2		
Accuracy	2K			
Precision	1K			
FOC	Continuous			
Coverage				
UF				

*See Chapter 9

Figure 11-3. Time line display

4. If procedures defined in Technical Orders, Preventive Maintenance Instruction (PMI) cards, etc., are to be used during the test, reference them, but do not repeat them here.

5. Include flight profiles and a missions/sorties/flying hours summary, by aircraft type.

6. Identify the number and location of voice communications terminals that will be used in the test. State call signs, frequencies (primary and alternate), and requirements for secure links.

7. Define steps to be followed in the event system malfunctions or other circumstances occur that cause deviations from these Detailed Test Procedures. The Test Director must decide whether to terminate the trial in progress, interrupt the trial for replacement or repair of equipment, or to continue the trial in a partially degraded mode. A "Go-No-Go" decision table may be formulated before the test.

   Possible malfunctions are anticipated and the course of action is decided beforehand. Factors contributing to this decision table may include:

   a. Does the malfunction affect a primary or secondary test objective?

   b. Is there a back-up system available?

   c. How long would repair take?

   d. Are resources and support available to perform the trial at a later time?

   In the event a malfunction occurs that was not anticipated by the decision table, a system of decision responsibility should be established before the test. Define which individuals shall have the authority to recommend a course of action to the Test Director, who shall have final authority. These individuals are usually the Test Officers or Engineers comprising the "Technical Assistance Panel." All changes occurring during the test should be entered in appropriate test logs.

8. Exclusive of equipment failure or cancellation by support agencies, identify:

   a. Meteorological conditions which will cause cancellation of the test, e.g., surface or upper wind velocities, visibility, clouds, etc.

   b. Periods (time of day) during which the test must be completed or be cancelled.

   c. Other criteria which will cause cancellation of the test.

G. POST TEST PROCEDURES AND DATA REDUCTION

1. A post-test debriefing should be conducted as soon after the test as practical. The Test Officer may decide to limit attendance at this debriefing to certain key participants. This section should identify those participants. Include the agenda and/or sample questionnaires and debriefing forms.

2. Summarize all data items (tapes, strip charts, films, etc.) to be produced during the test.

3. State the disposition and handling of each data item (i.e., identify who assumes responsibility for it at the conclusion of the test). For raw data items requiring an intermediate processing step (e.g., film processing, data tape processing etc.), state the required turn-around time, responsible agency, and agency or individual to receive the data.

4. Identify the processing, reduction, and analysis to be performed on each raw data item. State the form in which the final, reduced data will be presented in the test report, and group the final, reduced data by test objectives. A Data Processing Plan should be prepared prior to the test and coordinated with the computation center that is to perform the reduction and analysis of the data produced in the test. Typically, the Data Processing Plan will include (for each data channel):

a. Start and stop times of calibrations and trials.

b. The desired sampling rates for data reduction.

c. Identifications of statistical analysis programs to be applied to the data.

d. The desired form of presentation of the processed, analyzed data.

e. The desired number of copies of processed, analyzed data.

f. The security classification of the data.

g. The coordinate system that data points are being compared to.

h. Identification of channels to be stripped out. [4]

i. Desired stripout paper speeds. [4]

j. Desired maximum pen excursions for stripped data. [4]

5. Identify storage requirements and responsibilities for raw and processed data produced in the test.

H. REPORTING REQUIREMENTS

Identify required test reports, including preparation responsibilities, content, approval cycles, and agencies to receive the reports. Include sample report outlines.

I. SCHEDULE

Describe a nominal test schedule, which may include, but not be limited to days on which the following events will occur:

1. Review and concurrence with these Detailed Test Procedures.

2. Package and ship equipment and instrumentation.

---

[4] "Stripped Out" - Converted from the original raw data tape and re-recorded on strip charts displaying one or more channels of data per chart.

3. Commence installation and checkout of equipment and instrumentation.

4. Complete installation and checkout of equipment and instrumentation.

5. Conduct pre-test briefing.

6. Commence testing.

7. Complete testing.

8. Receive data.

9. Begin data reduction and analysis.

10. Complete data reduction and analysis.

11. Submittal dates of reports.

12. Package and ship equipment and instrumentation.

TEST PLAN CHECKLIST

1. Is the plan consistent with the directed test objectives?
2. Is the plan consistent with program objectives expressed in the ROC, DCP, and PMD?
3. Does the plan define a test which will yield the required information?
4. Does each section of the plan give adequate detail to allow the desired actions to be accomplished?
5. Does each section of the plan give adequate detail to cause the desired actions to be accomplished?
6. Does each section of the plan give adequate detail to control test conduct to the degree desired?
7. Does each section of the plan give adequate detail to properly inform authorities who will review and approve/disapprove the plan?
8. Does the plan call for schedules or expenditures of resources which cannot be met?
9. Does the plan follow the Standard OT&E Test Plan Format?
10. Are the plan sections numbered/lettered properly?
11. Have all suggestions in Chapter 11 for content of each section been considered?
12. Is classification of the plan properly marked?
13. Is the material clearly presented?
14. Is information unnecessarily repeated in different places?

Chapter 12

CONDUCT OF THE TEST

1. INTRODUCTION

The culmination of test planning and design activity is realized during the conduct of the test. The progress and ultimate success of a test will greatly depend upon the planning which preceded it, the support provided it, and the proficiency of the participants; and will be a direct reflection on the guidance, capability, and leadership of the test director.

The very nature of testing is involved with exploring the unknown or uncertain and always carries a degree of risk. With thorough planning and attention to detail, the risk can be reduced and controlled. However, when experimenting with unknowns, even with low level risk in complex engagements and scenarios requiring high degrees of operational freedom, the probability of fault is magnified. It is therefore necessary to include in the plan contingencies in case of failure. The best test results will normally be obtained by the Test Director who has the knowledge and versatility to react to the unexpected.

Testing is dynamic in nature. Tests are subject to mechanical and human failures, priorities, weather, and other factors which can be only partially controlled. When tests involve complex systems, often with man as a part of the system as well as the evaluator of its capability, such tests are likely to deviate from the Test Plan. In addition, test findings may include discovery of unforeseen capabilities or limitations which make the original design of the tests somewhat less than optimum. Due to the vicissitudes of testing, it may be necessary to make changes in the test plan during the course of testing. However, do not change the plan unless absolutely necessary, and then only after careful consideration of the disruption that will be caused. In-progress changes to the test plan will often have profound effects in the requirements levied upon support agencies for range operation, instrumentation, data collection, reduction, and analysis. In some instances, desirable changes will not be possible under the constraints of test resources and priorities. In order to cope with such situations, test personnel can benefit from a thorough understanding of the methods and techniques of testing which this document presents.

A test of only moderate complexity normally requires that the efforts of many individuals and items of equipment with different jobs and capabilities be combined to achieve a common goal, i.e., obtaining valid information in a form which permits analysis and the drawing of scientifically sound conclusions. The preparation of a Test Plan will involve meetings from which there will emerge a common understanding of what is required and how it is to be obtained.

Since the Air Force OT&E programs vary widely in objective, scope, complexity, and urgency, it is not possible to describe precisely how the Test Director will go about day-to-day test direction. There are, however, several precepts which have general application to all AF OT&E's which are discussed below.

2. REHEARSALS OR DRY RUNS

Perhaps the most common error in running tests is to assume that when the Test Plan is coordinated, approved and published, often with considerable difficulty, that agreement as to what is to be done is equated with the capability to do it. Since each agency is confident that it can fulfill its commitments, inadequate consideration may be given to the overall accomplishment. At this point, the Test Director must rehearse or dry run the test down to the last detail before valuable test items and resources are expended. No matter how carefully conceived the Test Plan has been to foresee problem areas, "dry runs" will reveal others. The alternative to rehearsal is to utilize the first few replications of a test for that purpose, thereby compromising test design and increasing the duration and cost of physical testing.

3. MAINTAINING RECORDS

Tests are conducted to obtain data/information for subsequent analysis. Test data/information may be collected in many ways: electronic, electro-mechanical, direct observation, etc. When processed for analysis, the information will normally consist of relationship of numbers (quantitative) and words (qualitative). It is very important that information be correlated, so that observations relating to an event are identified with it, time tagged and cued to each action. Identification will require a system which labels each bit of information accurately.

There is often a temptation to discard data which are obviously faulty or gathered in the course of a test phase which was incomplete and has to be redone. Keep all data, no matter how suspect they are. Frequently such data become invaluable in analyzing overall test results.

4. PERIODIC TEST PROGRESS CHECKS & EVALUATIONS

Only in the most simple tests are the test completed prior to examination and evaluation of the collected data. When tests involve multiple objectives and considerable replication, it is essential to pause periodically to examine and evaluate what has been done and consider the impact of testing to date upon the testing that remains. Milestones for such pauses are usually self-evident, but need to be planned. For example, in a test involving aircraft, a check point would be planned after each mission. A mission is ordinarily planned to fulfill more than one specific objective, and the data/information relating to each objective should be examined carefully to see whether subsequent missions require modification. Collecting data and assembling them in a form for quick evaluation takes time and may cause impatience on the part of test operations personnel. Under certain conditions, it may be reasonable to pause after two or more similar missions. There is a point, however, beyond which it is inappropriate to collect more data without evaluation.

The necessity for periodic evaluation has been widely recognized by the Air Force. Even with the most advanced electronic data processing, the ability of instrumentation to gather data/information has far outstripped the capability to process and evaluate it. Consequently, provision is ordinarily made to extract carefully selected data for "quick look" evaluation which may lead to changes in the course of continued testing.

5. CHECK TEST DATA FOR CONSISTENCY WITH TEST OBJECTIVES

One test execution pitfall which may have damaged or ruined more tests than any other is the tendency to reach the end of a test with a varying amount of information about each of the objectives but insufficient information to say anything with confidence about any of them. Tests are ordinarily designed to produce the required amount of information, based upon a predetermined number of replications, for each test objective. The test plan should arrange the objectives in order of priority. They may range from "essential" information, through "highly desirable," to other information which would be "by product". It is essential, as the test proceeds and changes are made, to adhere firmly to these priorities. Some lesser objectives may have to be dropped during testing if these interfere with obtaining "essential" information. For example, it would be poor test management to conclude a test with low-confidence information on a series of eight objectives, when by discarding the lowest five it would have been possible to gather high-confidence information on the objectives which ranked 1, 2, and 3. The corollary of this guidance is to avoid including low-ranking objectives in the Test Plan if obtaining information about them will obviously endanger achievement of the "essential objectives." It is not always possible to fulfill the test plan completely, but it is inexcusable to finish a test with no valid information.

6. MAKING CHANGES DURING TESTING

AF OT&E is a dynamic process. The Test Director must be alert to the desirability and even necessity of making changes in the test plan on short notice. The most drastic change that can be made is to suspend testing for a period-or indefinitely-or to bring an end to the test before the test plan has been fulfilled. Paradoxically, these are the easiest actions to take and to defend. It is common sense to recommend suspension of testing if meaningful results are not being obtained or if data recovery is so meager that there is no reasonable chance of fulfilling the essential test objectives. Sometimes there is no alternative to stopping the test. One major reason for interrupting a test is the discovery that the data collection effort (white forces) is not producing realistic data which can be adequately used to fulfill test objectives. Corrective actions are then required to improve the data collection process. If such actions are not successful, the next option may be a restatement of more realistic test objectives. The test may also be concluded if the objectives have been fulfilled earlier than planned. Although this is not common, it has happened often enough to be cited as a possibility.

Perhaps the most difficult change to make is the addition of a test objective during the course of testing. Historically, some of the most valuable findings of tests were not anticipated but became apparent in the course of testing. Such findings are essentially by-products of the planned testing process. Unexpected test findings sometimes arouse unusual top level interest and test modification to investigate them is actively supported to the extent that the original objectives of testing assume secondary status. Such a change should be made with an awareness that it must be very important to justify the trouble it will cause. In such cases, it is recommended that a separate test - with all the urgency the situation justifies, be considered.

Before making changes in the test plan for any reason, test integrity must be kept in mind, i.e., the original test design and the comparability and validity of test results may be affected. Unless it can be demonstrated that valid results can still be obtained and that useful conclusions can still be drawn, it will probably be inadvisable to make substantive changes during the course of testing.

7. IMPORTANCE OF THE TEST DIRECTOR

A special study of a relatively large number of operational tests performed over a period of years, with the tests ranked according to the success with which they had been conducted, revealed the names of certain Test Directors which appeared repeatedly. Since the number of Test Directors at this particular test agency, during the period studied was also several times the number of tests which made the "most successful" list, there is considerable evidence that good Test Directors produce good tests. While there are other contributing reasons to the situation noted, this general conclusion has held consistent and remains valid.

8. CHARACTERISTICS OF A GOOD TEST OFFICER

Leadership. Test Director must be capable of leading a dedicated team of individuals whose background, experience and interests will be quite different. Teamwork is essential in all operational tests, in major OT&E programs the team assembled to conduct it will be very large. There will be disagreements and misunderstandings. The Test Director must clearly establish his authority and responsibility very early in his relationship with his team members. He must be decisive, for while he will have access to expert advice, he is ultimately responsible for making the decisions.

Understanding the Test as a Whole. The Test Director must acquire a comprehensive knowledge of the system or equipment being tested, the design of the test, and the test support which he will use. He should know the capabilities and limitations of instrumentation. He must have a good grasp of data collection and reduction procedures and problems and the techniques and goals of analysis of test results. Knowing as much as possible about the test has some very practical results. For example, the Test Director will know what demands he may reasonably make upon those people and organizations assisting him, and can, therefore, avoid unrealistic requirements. He will also know when his test program is being improperly supported and can take appropriate corrective action. He must become a part of the test: operate the equipment, fly a sortie, collect some of the data, participate in the analysis, and keep mobile while keeping in contact with control.

Objectivity. The Test Director must develop and maintain an unbiased attitude toward the test, and he must be sure that all test personnel under his direction maintain the same attitude. There will inevitably be preconceived opinions about test progress and what it will prove. However, without such preconceived ideas, test concept development and design would be nearly impossible to achieve and, certainly, inefficient. Furthermore, operational tests of Air Force systems will usually be conducted to validate pretest calculations of performance and capabilities, often subjected to prior study and/or simulation. There will nearly always be a desired result expressed in quantitative form.

When the test begins, however, the Test Director must deal only in facts, insofar as they can be determined by testing. He must avoid influencing the course of testing to achieve a desired result, and he must also be constantly on guard against being influenced by others.

Test Management Ability

The Test Director must above all, be a good manager. He must maintain test program orientation in relation to the critical issues and objectives planned to be satisfied. He must have control of his resources and constant awareness of their disposition. Lead times for resource acquisition and their availability are equally as important as the quantity required. Critical areas requiring attention include but are not limited to equipment, instrumentation facilities, manning, scheduling and priorities. As a note of caution, the dynamics of OT&E program changes will normally always have an adverse effect on resource management. In order to maintain test program continuity, all management efforts and decisions must be exercised in context with the desired successful test accomplishment expressed in terms of success criteria established in the planning phase. The success of a "test" is not directly keyed to the successful performance of a system or object under test, but it is measured in relation to the achievement of test objectives through the collection of valid data/information designed to be gathered to meet the success criteria.

CHECKLIST

CONDUCT OF TEST

OT&E PROGRAM MANAGEMENT MASTER CHECKLIST

Carefully study assignment responsibilities and develop clear test management approach. Study the mission and employment methodology for the test item(s). Define the practical limits of the test in general terms.

1. Get an assistant - Your future Operations Officer's a good one.

2. Acquire appropriate background and directive documentation - Test Directive, ROC, PMD, DSARC Reports, DT&E Plans and Reports, other related test documentation, and test item descriptions, tech manuals, operating/employment instructions, intelligence estimates, etc.

   -Develop a clear understanding of the problem(s) and critical issues driving the test.

   -Initiate a preliminary planning and coordination conference to include all responsible commands and agencies involved in operation and support of the test item.

3. Become intimate with the test item whether it is hardware, software, doctrine, tactics, or operational procedure, etc.

4. Take action to have assigned a cadre planning staff. Administration, Operations, Engineering, Maintenance and Logistics, Tech Support, and intra command/agency liaison personnel as required.

   -Carefully review the qualifications and experience of proposed team members.

5. Provide for staff office space, administrative support, and living accommodations.

6. Initiate action in other participating commands/agencies for required Test Plan Support Annexes.

7. Provide for and properly schedule any required special schooling required for the staff and/or test team members, particularly for any training required prior to reporting for duty with the test team.

8. Establish interfaces with other commands/agencies as required.

9. Become more intimate with the test item and scout test location and test support resources while the cadre staff is assembling.

10. Assign specific responsibilities and projects to arriving staff members, with particular emphasis on tasks prerequisite to designing the testing program and writing the Test Plan.

-Define the test concept and scope the test.

11. Complete the test design (Refer to Chapter 7) coordinate and publish

12. Complete, coordinate, and publish the Test Plan (Refer to Chapter 11).

13. Establish, if program size and complexity warrants, a Program Review Council and meeting schedule comprised of test team staff, representatives of participating units, and supporting organizations.

Establish:

Problem resolution policies
Communication channels (internal)
Status reporting procedures
Detailed milestones and schedules
Working rapport within the test force

14. Program, schedule, and complete formal acquisition of required test resources and support.

15. Insure that required test resources, team members, instrumentation and plans are in hand and functional when test inventory/item is received, or scheduled start date is reached.

16. Make provisions to have visitors briefed without disturbing key personnel.

17. Start active testing.
Provide procedures for quick-look analysis for checks on test problems and progress.

18. Start Test Report (Refer to Chapter 15).

19. With infinite aplomb and composure sit back and watch your perfectly contrived plans executed with magnificant precision and professionalism, to an impressive and timely conclusion, and an exquisite Test Report -- like Hell! Grow eyes in the back of your head; manage, supervise, drive, finagle, cajole, scheme and drive some more, and with any luck at all, everything will come out reasonably well, and someone will, hopefully, appreciate what you and your troops have done.

20. Insure that data management and analysis is accomplished as planned (Refer to Chapters 9 and 13) and that conclusions and recommendations are sound and substantiated by valid data.

21. Insure that all required special, interim, and progress reports are submitted as planned; and that the final Test Report is in process concurrently with active testing.

22. Plan for the disposition of accountable test resources and disbanding of the test team on a timely and programmed basis.

23. Write letters of commendation and appreciation as appropriate - they do good things for everybody and are particularly helpful for subsequent test programs.

24. Host a final appreciation party before the test team is physically disbanded.

25. Review, coordinate, edit, polish, and publish the Test Report. Command and AFTEC approval is required.

26. Have another party, a promotion one hopefully.

Chapter 13

DATA ANALYSIS

Data analysis, like test design, is an area sometimes considered the sacrosanct kingdom of the statistician and properly of interest only to the statistician. This is a mistake because a Test Officer can take greatest advantage of his professional staff only by understanding the types of things they can do and the variety of ways to draw conclusions from test data. This section is designed to give the Test Officer an introduction to the field.

In spite of the fact that no two operational tests ever seem to be the same, the Test Officer will learn that most of the data analysis problems can be boiled down to a fairly short list. Each test has its own problems, to be sure, and at times the unique nature of the test design or the test data requires the statistician to employ a specialized and seldom-used analysis tool, but when abstracted from the nouns used in a particular problem, there is a limited number of fundamentally different questions being asked. What is the value of some parameter believed to be? In what range will the value of some parameter probably be? What is the relationship between values of one variable and values of another? What are the sources of the observed variability among trials? Do two (or more) groups appear to be the same? Does a parameter have the value claimed by the advocate or developer of the system?

At times these same questions get turned around to ask about the probability of realizing certain conditions in the population of inference. What is the probability that a parameter is between two given values? What is the probability that the parameter is less than a given value? What is the probability that some percent of the population values for a variable will be less than a given value?

The application of techniques to answer these and other questions will be illustrated by example below. The list is not exhaustive; it illustrates most of the classes of problems that will be encountered, and the understanding of other specific techniques will not require a great deal of reorientation.

1. STATISTICAL TECHNIQUES

Estimation

(a) Point Estimates

(1) Central Value. Estimation of the population (arithmetic) mean or median are the most common examples of this category. The population mean μ, which is estimated by the sample mean $\bar{X}$, is a good parameter to use in describing symmetrical distributions. For the data of Figure 13-1, the sample mean is 5.5 hours and the estimated population mean is 5.5 hours.

1.0, 1.1, 3.0, 3.6, 5.2, 6.6, 6.7, 8.9, 9.2, 9.7

Figure 13-1. Time-to-repair data (hours)

If the distribution is skewed the population median (a measure of the counted middle position of a series of numbers, unweighted by the value of each number) may be more useful. The population median is estimated by the sample median.

0, 1, 1, 2, 2, 2, 2, 3, 5, 8, 15

Figure 13-2. Miss distance data (meters)

The estimated population median for the data from Figure 13-2 is 2 meters. Half the shots are expected to miss by more than 2 meters; half by less.

(2) Dispersion - One Measure - Two of the most useful measures of dispersion, or spread, are the range and the variance. Population range is the total spread in value between the highest value and the lowest value of the population; it is estimated by a tabulated constant times the sample range. For the data of Figure 13-1, the estimated population range is 8.7 hours. For the data of Figure 13-2, the estimated range is 15 meters. The population variance is the arithmetic mean of the sum of the squares of the differences between the individual observation and their

arithmetic mean. An unbiased estimate is made by summing the squares of the differences between individual sample observations and the sample mean and dividing the sum by one less than the total number of observations in the sample. For Figure 13-1 data, the estimate of the population variance is $[(1.0-5.5)^2+(1.1-5.5)^2+ \ldots + (9.7-5.5)^2]/9$ or 10.57 hours. Since unsigned miss distances (Figure 13-2) are measures of dispersion in themselves, the variance of this set of data would measure "dispersion of the dispersion" -- probably not the most useful bit of data for the operational tester. Consider instead the variance of the signed data (Figure 13-3). The estimated population variance is 29.29 meters^2. Another commonly-encountered population dispersion measure is the standard deviation. The population standard deviation is the square root of the population variance but because of the skewness of the sampling distribution of standard deviation estimates, the square root of the estimate used for variance is only moderately useful in estimating the population standard deviation. On the average, it gives an estimate which is too low. It is in the units of the measured data, however. Unbiased estimates of the standard deviation can be made.

0, -1, -1, 2, -2, -2, 2, -3, 5, 8, 15

Figure 13-3. Signed miss distance data (meters)

(3) Dispersion-Variance Components. It may be that the test scenario has several sources of trial-to-trial variability in outcome or response. In a properly designed test it is possible to divide the total variance into several components, each attributable to one of the controlled independent variables or to the trial-to-trial variability when all independent variables are (supposedly) held constant. This last is usually referred to as experimental error and indicates the extreme lower limit on the uncertainty with which population event outcomes can be predicted. The technique used is known as analysis of variance.

From the data of Figure 13-4, the population variance from trial to trial for any single operator is estimated to be 0.38 seconds^2 while the population variance among the (mean) average performances for different operators is estimated to be 2.09 seconds^2.

	TRIAL UNDER SIMILAR CONDITIONS				
OPERATOR	1	2	3	4	5
1	10	11	10	11	12
2	13	12	11	12	11
3	15	14	14	13	13
4	11	11	12	11	11

Figure 13-4. Time for radar operator to detect approaching aircraft after aircraft comes within specified range (seconds)

(4) Relationships. Another type of estimate is the relationship between values of an independent variable and the values of a dependent variable. Then for any value of the independent variable a point estimate of the dependent variable can be made. If there is a clear relationship, and the independent variable is controllable in operational employment, the relationship can be used to predict or control the dependent variable to some degree. An example would be the relationship between aircraft velocity on a bombing run and miss distance. The aircraft is assumed to have released one bomb at a 45° dive angle on each trial, and three trials were conducted at each of several release altitudes. The data is plotted in Figure 13-5 and because it plots well on logarithmic scales, these graduations are shown on the top and right sides and were used to simplify calculations. A useful fit to this data is the straight line

$$\log_{10} (\text{miss distance}) = -1.86+0.953 \log_{10} (\text{release altitude}).$$

$\log_{10}$ is the base-10 logarithm in each case. Another way to express the same line is

$$\text{miss distance} = 0.00139 (\text{release altitude})^{0.953}.$$

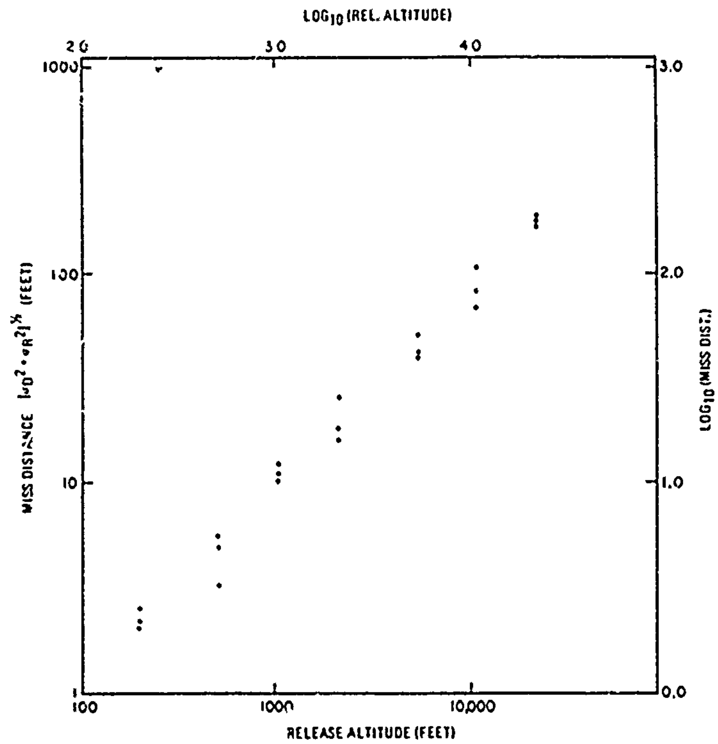


Figure 13-5. Release altitude - miss distance relationship data

The statistician would say that this line is the least square regression of miss distance on release altitude. See Figure 13-6.

(b) Interval Estimates of Parameters.

In most cases the point estimate of a parameter does not carry enough information to give a useful picture of test results. For instance, whether the data are

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

or

1, 1, 1, 1, 1, 1, 12, 12, 12, 12, 12, 12

the arithmetic mean is 6.5. Interval estimates give an interval (or for more than one dimension, region) within which the true population parameter is with a specified probability. Now in fact, the parameter either does or does not lie in the stated interval. The probability spoken

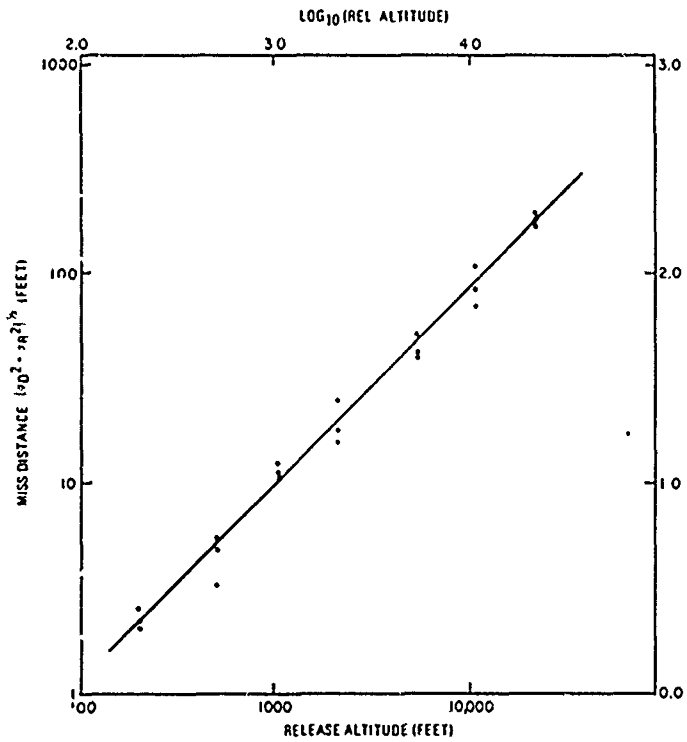


Figure 13-6. Estimated line of regression of miss distance on release altitude

of is the probability that the statistician is constructing an interval which actually does contain the population parameter. The higher the probability of containment, the longer the interval will be. Interval estimates contain more information than point estimates because the former take into consideration the variability of the sample data as well as the sample size (A larger sample allows a shorter interval to be used).

(1) Central Value. The arithmetic mean of the population of times-to-repair from which the sample of Figure 13-1 was drawn will be between 1.7 hours and 9.3 hours with probability of 99%. In other words, based on the statistician's knowledge of sampling, there is only one chance in 100 that the population arithmetic mean is not between 1.7 hours and 9.3 hours. An alternative interval estimate for the same data is

that the population mean is less than 8.8 hours with probability 99%. Interval estimates for measures of central value other than the arithmetic mean tend to give longer intervals for the same level of confidence in their correctness.

(2) Dispersion. Interval estimates for measures of dispersion can be presented and interpreted in just the same way that those for measures of central value are. For example, a 95% probability interval estimate of the population variance for the sample of Figure 13-1 is the interval from 3.17 hours^2 to 22.33 hours^2. A 90% probability interval for the same estimate is the interval 3.91 hours^2 to 19.87 hours^2. This puts probable "bounds" on the variability of repair times.

(3) Relationships. Two different interval estimates related to a regression line may be useful. One gives an interval for the population mean of the predicted variable at a single value of the predicting variable while the other gives an interval for the predicted variable at any value of the predicting variable -- in other words, a "band" estimate for the line as a whole. Both of these will be illustrated for the example used in 1.a. (1). (d). The interval estimate for the single release altitude 3000 feet is shown in Figure 13-7 as a vertical line. Notice that it is shorter than the band estimate for the entire regression line (Figure 13-8) at 3000 feet.

(c) Interval Estimates of Content.

Knowing where a population mean is likely to be might not be very helpful to the Air Force. Who cares, for instance, that the estimated (arithmetic) mean miss of all bombs dropped under the conditions of the data in Figure 13-9 is 0±2.8 meters with about 95% probability (point and interval estimates combined)? This says that all aiming biases have probably been removed, but does not predict how

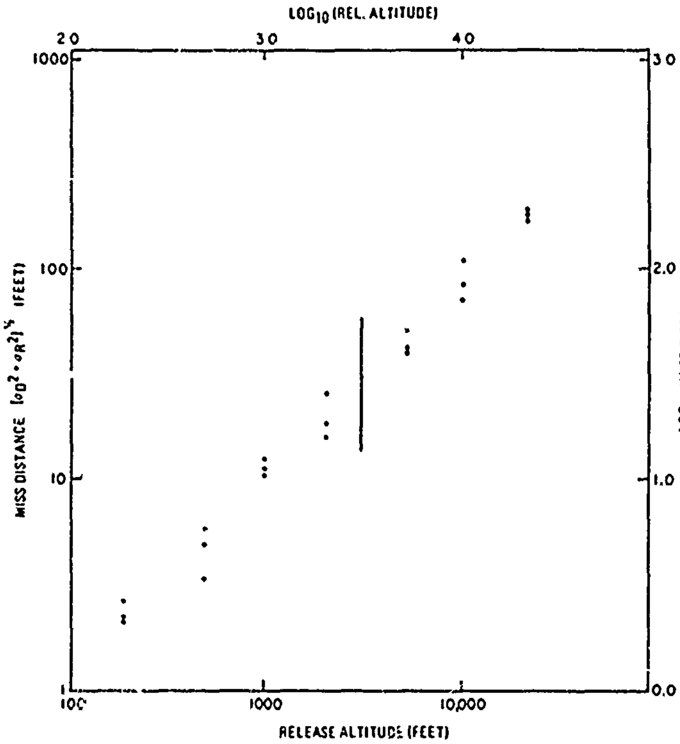


Figure 13-7. Interval estimate of regression of miss distance on release altitude 3000 feet (95%)

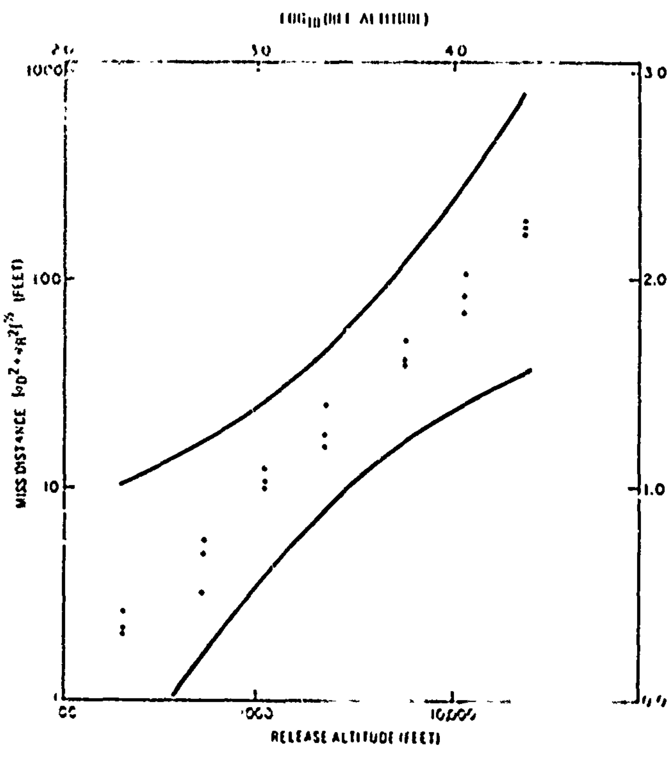


Figure 13-8. Band estimate of regression of miss distance on release altitude (95%)

-10, -9, -8, -7, -3, -3, -2, -1, 0, 0, 0, 2, 2, 4, 4, 6, 7, 9, 9

Figure 13-9. Signed miss distance data (meters)

much scatter there will be in the operational bomb drops. An estimate of the variance would help some, but population variance and standard deviation do not always have an obvious significance to the non-statistician. A useful device may be the estimated tolerance interval, which includes a specified fraction of the population with a stated probability. For the data of Figure 13-9, as an example, it may be stated that the interval -9.5 meters to +9.5 meters will contain 75% or more of the (population) hits with 95% probability. This says that there is only one chance in twenty that more than 25% of the hits will be further than 9.5 meters away from the target. This problem has been for a one-dimensional distribution of miss data only; tolerance regions for two and three dimensions are also useful.

b. Comparisons

(a) One Sample vs. Assumed Population.

An aircraft development firm claims that there is no built-in bias in the navigation system of a new prototype. Yet, in operational testing the Air Force collects the error data in Figure 13-10 for repeated sorties over similar routes. Is there a bias in the system? Although eight out of the ten trials gave a positive range error, it can be shown that more than five percent of the samples of size ten from a population with zero bias will give this much of an appearance of bias in one direction or the other as long as there is a certain level of variability. Fewer than ten percent will, however. The statistician will say that the results are statistically significant at the .1 level, though not at the .05 level. Without further

1.0, 0.7, -1.1, 0.4, 0.8, 0.6, -0.8, 0.6, 0.9, 0.2

Figure 3-10. Navigation error-range (kilometers)

interpolation, the observed significance level is said to be .05; actually it is only slightly higher. The lower the significance level, the more likely it is that the observed difference indicates a real difference. For a different illustration of the idea of statistical significance (never to be confused with operational significance), consider the test of a new flight suit in which the wearer was asked to give an indication of preference for either the old suit or the new, based on comfort. If the two were actually indistinguishable, the tester would expect just about half of the wearers to prefer the new suit and half to prefer the old. Perhaps the data from the test shows 15 of the 25 participants preferring the new suit while 10 prefer the old. A statistical test on the proportion (actual vs. expected) reveals that for sample size 25 the result is significant at the .32 level, and hence we would say that no clear preference is initiated by this data.

(b) Two Samples. Often the operational tester will want to find out whether one test item is significantly better than another. The meaning of "significant" is ambiguous here; it was used to point out that very problem. First the Air Force must define an operationally significant difference, and then the tester can find out if the test data indicate that such a difference has been observed or if there is a statistically significant deviation from the expected (hoped for) difference. Perhaps two dogfighting techniques are being evaluated and the criterion for decision is a 25% increase in the maximum time a friendly aircraft can pull lead on an enemy aircraft at any one time in a particular game. If technique A maximum time is about 20 seconds, a 25% increase with technique B would give a total of 25 seconds. A test gives the data in Figure 13-11.

The sample (arithmetic mean) averaged differ by 4 seconds. The statistician cannot say that the hoped-for improvement was realized. An estimate of the true difference based on the results of this test would be +4.0±2.0 seconds with 95% confidence; this includes the hoped-for

GAME NO.

TECHNIQUE	1	2	3	4	5	6	7	8	9	10	11	12
A	20	19	18	15	24	24	22	21	19	12	20	26
B	28	26	24	25	23	23	16	22	24	19	30	28

Figure 13-11. Maximum time pulling lead

operationally significant difference so the statistician certainly would not say the difference does not exist. The most direct reply to the question of an operationally significant (+5 seconds) improvement is that the observed (sample) difference is a statistically significant deviation from the operationally significant improvement only at the .20 (approximately) level. The observed (sample) overall improvement of 4 seconds with technique B is a statistically significant deviation from the possibility of no improvement at the .0025 (approx) level.

(c) Several Samples. At times, it will be important to decide whether there is any difference among several groups of data. Consider as an example a bombing accuracy test run with six separate aircraft. The tester wants to check an earlier assumption that there is no change in bomb delivery accuracy determined by the particular aircraft used. A first step is to see if the dispersion (as measured by the variance) differs from aircraft to aircraft. For the data of Figure 13-12 it can be shown that the dispersions (variances) are all equal for a .005 significance level. Another test can now be used to determine whether any one of the aircraft gives an unusually high or low mean radial error, or if the apparent differences only reflect the dispersion of the bomb drops. The technique known as analysis of variance (see 1.(a).(3)) is also used in detecting differences among means and together with a proper test of significance it indicates that there is no probable real difference (.25 significance level) in mean radial error among the six aircraft. This result comes in spite of the fact that miss distances for one aircraft (#651) appear to be quite a bit higher than for the others. The reason

TRIAL	AIRCRAFT TAIL NUMBER					
	643	243	651	652	496	314
1	83	10	118	113	0	46
2	105	176	148	65	20	139
3	2	150	138	37	53	33
4	57	43	115	43	88	113
5	61	29	86	179	110	4
6	29	71	31	145	204	193
7	197	37	189	113	166	121
8	93	108	172	9	111	95
9	120	--	209	--	--	--
mean	83	78	134	88	94	93

Figure 13-12. Bombing radial error (meters)

#651 is not obviously different is the large dispersion among hits for any one of the aircraft.

(d) Independence. The tester may have observed what appears to be a correlation between pilot rank and the scoring of direct hits in an operational test of a new laser-guided glide bomb. The data being examined is in Figure 13-13, and the tester wants to know if bombing accuracy and pilot rank are independent or if they are correlated as it appears.

PILOT RANK	NUMBER OF BOMBS	
	DIRECT HITS	MISSES
CAPTAIN	40	12
MAJOR	33	15
LT. COL.	24	17

Figure 13-13. Pilot scoring records

A statistical test for independence shows that these results are statistically significant at the .2 level but not at the .1 level. Now does the tester conclude that the younger pilots deliver bombs with greater accuracy? This example is a good illustration of the difference between making conclusions (which the science of statistics does not do) and quantifying the conclusiveness of the test data.

2. GRAPHIC TECHNIQUES

A great deal can be learned from test data when it is layed out on paper in "picture" form. As the amount of data collected grows, the process of digesting the data becomes more difficult. A large table of numbers may be inconclusive (sometimes the meaning to be attached to just a few numbers is not clear.) and the statistical descriptors chosen to summarize the data may not tell the whole story. A good visual presentation of the data not totally dependent on numbers may suggest trends and exceptions not immediately obvious from the raw data or picked up in statistical tests. It can be helpful both to the data analyst and to the user of a test report.

Histograms.

The histogram is a graph of frequency of occurrence against level of a variable for a limited number of discrete levels or groups of continuous levels. Figures 13-14, 13-15, and 13-16 are examples.

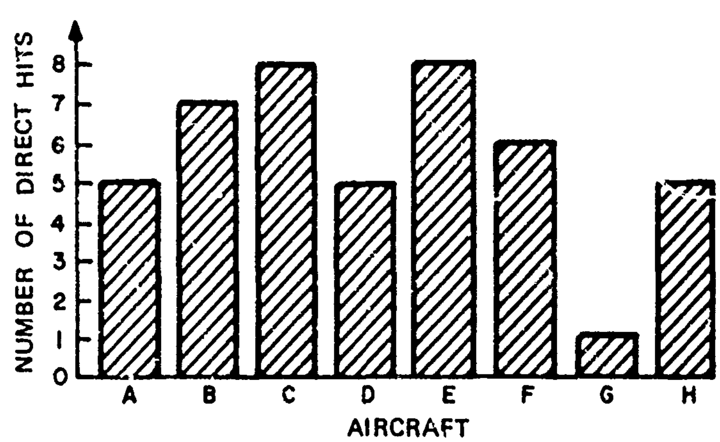


Figure 13-14. Sample histogram

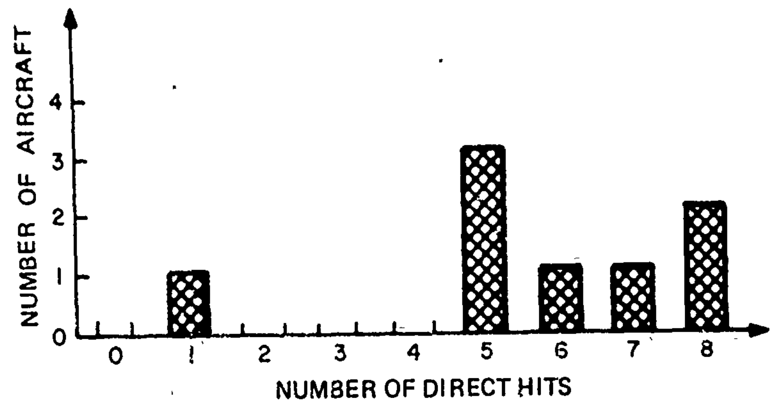


Figure 13-15. Sample histogram

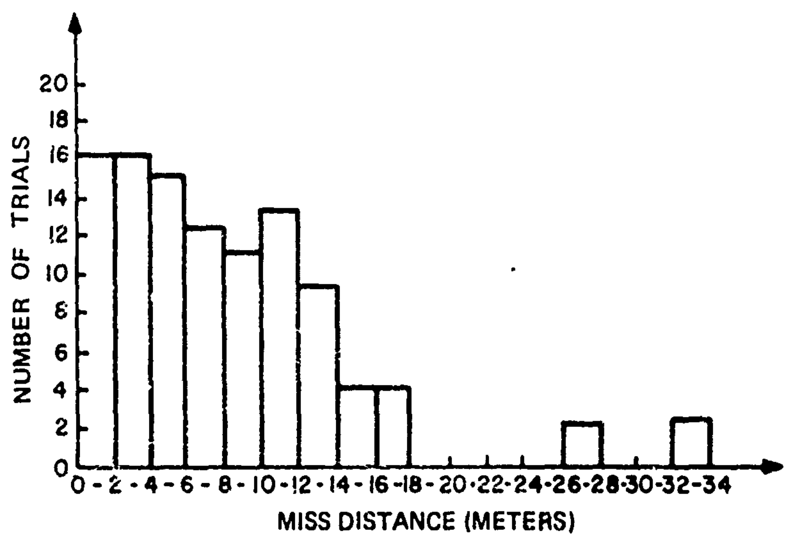


Figure 13-16. Sample histogram

Continuous Frequency Functions

If the number of intervals into which the data is naturally divided is large and the intervals can be ordered on a scale of values, or if the data is taken at a finite number of points along a continuous scale, a smooth density distribution curve or cumulative distribution curve can be useful. Figures 13-17 and 13-18 are examples.

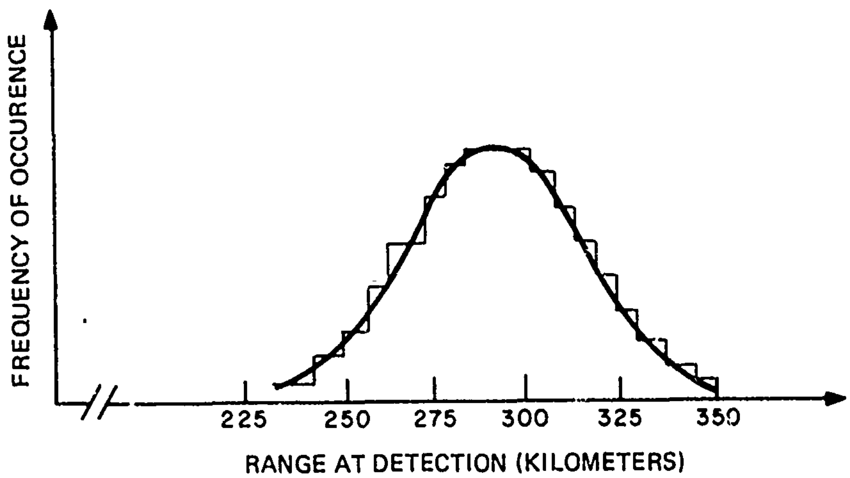


Figure 13-17. Density distribution function

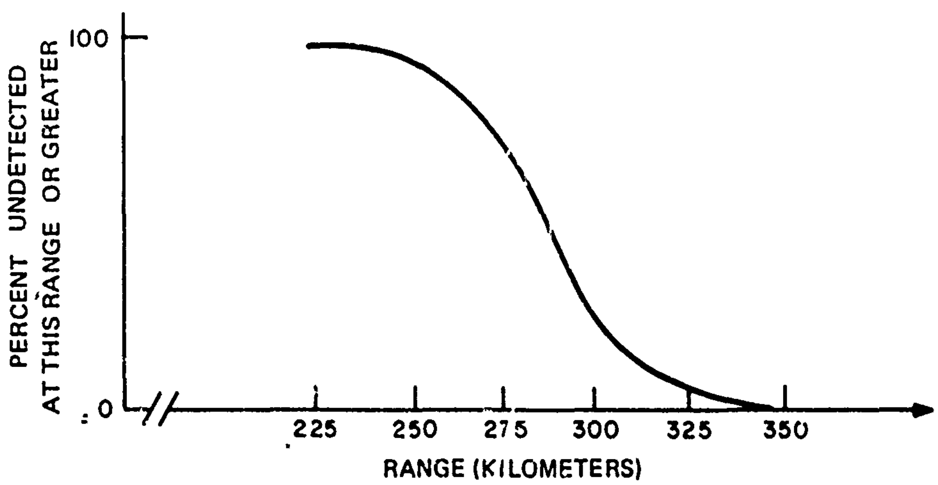


Figure 13-18. Cumulative distribution function

Scatter Diagrams

Understanding of two-dimensional data and possible relationships between values of two different variables are both facilitated by presenting the data as a series of points on a set of perpendicular axes. Examples are shown in Figures 13-19 and 13-20.

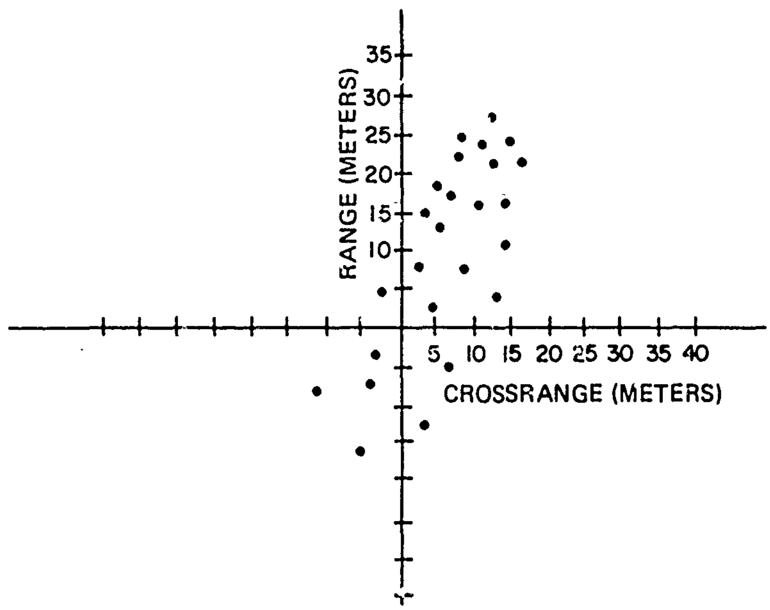


Figure 13-19. Miss distance scatter diagram

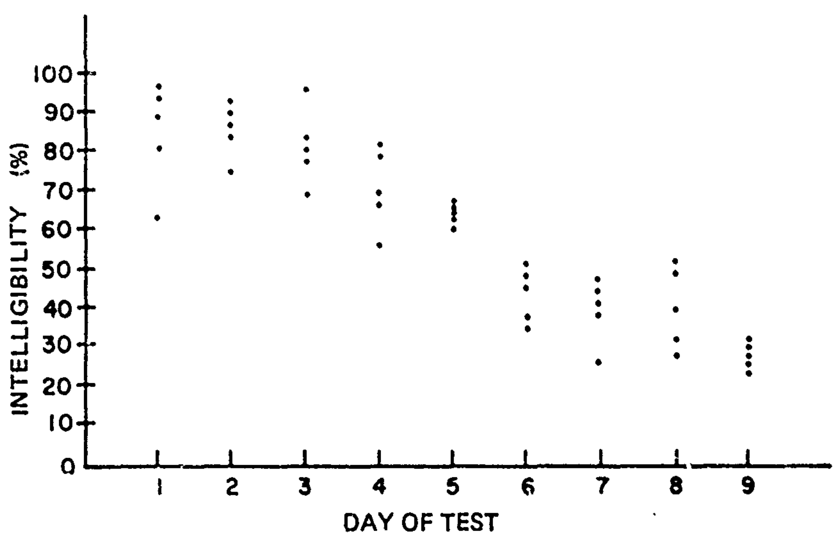


Figure 13-20. Intelligibility score scatter diagram

Time Line Diagrams

At times it will be desirable to make a "big picture" review of the events in a dynamic scenario, and trends in data not easily quantified may be obvious by showing graphically what happened, when it happened and (perhaps) where it happened. In time line diagrams one can show momentary events, continuing events, interactions between participants, and insertion or removal of participants. An air-to-air engagement involving several aircraft is the basis for the time line diagram of Figure 13-21.

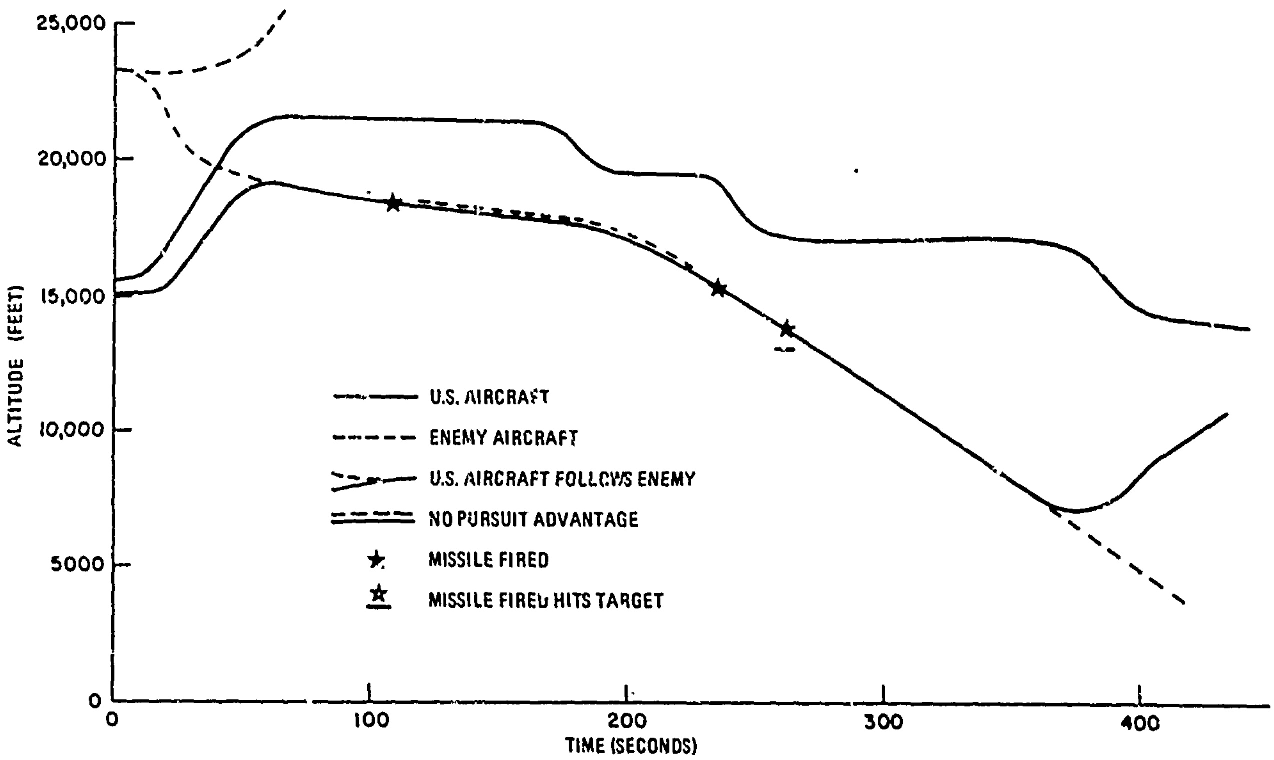


Figure 13-21. Time line diagram air-air-game

Data analysis has been divided into two parts -- the statistical and the graphical techniques. Although the statistical techniques are more quantitative in nature, the usefulness of being able to grasp large amounts of data in a pictorial summary cannot be overstated. It is often important for the tester to follow progress closely and be able to draw tentative conclusions (or at least summarize the results of a test) at any time. Graphical analysis techniques, together with the current availability of hand-held calculators and high-speed computers, make this task easier for the Test Officer. It is a much bigger problem to draw supportable non-statistical inferences from the data than it is to derive factual information.

REFERENCES FOR DATA ANALYSIS

A good library will have a large number of related books for the test officer who wishes to learn more about any of the topics discussed in Chapter 13. The problem is to find one that suits the educational background and style of the individual user. Some of the best references for the non-statistician are:

Davies, O. L., and Goldsmith, P. L. (ed); Statistical Methods in Research and Production; Hafner Publishing; 1972. Helpful discussion of regression and analysis of variance.

Dixon, W. J., and Massey, F. J., Jr., Introduction to Statistical Analysis (3rd ed); Good on estimation and hypothesis testing. Includes a large number of tables.

Natrella, M. G.; Experimental Statistics (National Bureau of Standards Handbook No. 91); U. S. Government Printing Office; 1966. Also available as Engineering Design Handbook -Experimental Statistics (U. S. Army Material Command Pamphlets 706-110, 706-111, 706-112, 706-113, and 706-114). Written in near cookbook style. Includes a large number of tables (706-114).

Snedecor, G. W.; Statistical Methods (4th ed); Iowa State University Press; 1946. Later editions are more readily available, but this one has reputation for clarity. Helpful analysis of variance discussion.

There are several books of statistical tables available. Some of the most extensive are:

Abramowitz, M., and Stegun, I. A.; Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables (National Bureau of Standards Applied Mathematics Series No. 55); U.S. Government Printing Office; 1970.

Beyer, W. H.; Handbook of Tables for Probability and Statistics (2nd ed); The Chemical Rubber Company; 1968.

Owen, D. B.; Handbook of Statistical Tables; Addison-Wesley; 1962.

Chapter 14

FORMULATION OF THE CONCLUSIONS AND RECOMMENDATIONS

1. IMPORTANCE OF CONCLUSIONS AND RECOMMENDATIONS

The real impact of any operational test is determined by the statement of conclusions and recommendations in the test report. These are the grounds for taking future action and they are among the limited pieces of a final report that are incorporated, verbatim, in an executive summary for quick review by senior officers. Any inaccurate or unjustified statement in these sections or any omission of relevant information could have a major impact on a multimillion dollar acquisition decision or on the safety, defense, and/or effectiveness of operational troops. These conclusions and recommendations are supposed to contain, in effect, all that has been learned in a test, and they must not be considered so lightly that they are formulated and written down fifteen minutes before the final report goes to press.

2. THE MEANING OF CONCLUSION

Before striking off blindly to tell everything he has observed in the course of a test, the tester should make sure he is aware of the type of information that is supposed to be conveyed in a statement of conclusions and recommendations. A conclusion, when used in this context, is not simply a closing statement but is rather an inference or a factual statement that the tester believes he has evidence to support. It is not a decision. Nor is it a re-statement of the results of a test. A conclusion is a statement of what has been shown and embodies a prediction arrived at by a logical reasoning process. "The mean radial miss distance was 20 meters" is not a conclusion; "The mean radial miss distance for all Mk 1 bombs dropped in this way is estimated to be 20 meters" is a conclusion. "Bombs released from higher altitudes missed the target more often" is not a conclusion; "Releasing Mk 1 bombs from higher altitudes causes them to miss the target more often" is a conclusion.

Two types of logic may be used in arriving at conclusion. Deductive logic is used by the tester who says, "This missile consistently fails to guide properly; therefore, the air-to-air engagement capability of the launch aircraft will be low." Inductive logic is used by the tester who says, "The test missiles failed to guide properly in nine out of ten launches; therefore, this type of missile consistently fails to guide properly."

A conclusion is derived from a particular source (or sources) of information. This means that the tester should make clear the basis for the conclusion and differentiate among conclusions based on: (a) the test results alone, (b) the test results and the results of other evaluations, (c) the test results and knowledge of operational practices, (d) the test results and the knowledge of operational requirements, and (e) some combination of the above.

3. THE MEANING OF RECOMMENDATION

The meaning of "recommendation" is not so apt to be misunderstood; it is simply a statement of advice. Although recommendations may be poorly formulated, their purpose is seldom misconstrued. Conclusions may turn out to be recommendations in disguise (e. g., it is concluded that the AN/ABB-X should be purchased as an anti-personnel weapon system), but the practice should be avoided. The Test Officer can check to see that (1) all recommendations appear in the Recommendations section of the report and (2) the Conclusions and Recommendations sections are not redundant.

4. DRAWING CONCLUSIONS

How does the Test Officer go about drawing correct conclusions? How does he go about reaching all the conclusions that he could? How does he limit this activity to conclusions which are relevant? Some general guidelines can be suggested for drawing conclusions: (1) to address test objectives and (2) summarize the test data in other significant ways.

The test was conducted to provide answers for questions posed by the objectives. (There may be objectives aimed simply at the collection of data for use by someone else — such as when the Air Force Test and Evaluation Center collects data for Air Force Logistics Command —, but in these cases there will be no "answer"). The test design was formulated to make sure data would be collected as required to answer those questions. Therefore, if the test was properly conducted the data analyst should be able to draw conclusions quite mechanically. In theory this is completely correct; in practice it only outlines a more involved process of analyzing the test data and making inferences. See Figure 14-1. The tester draws his conclusions in four steps.

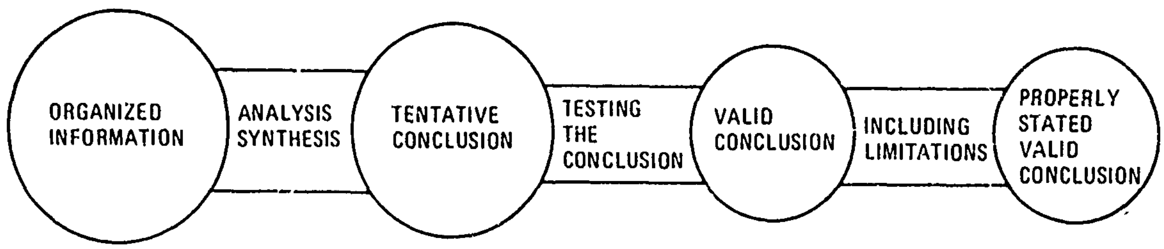


Figure 14-1. Derivation of conclusions.

Organization.

All information appearing relevant to a single objective, whether from the current operational test or from other sources of information being exploited, is brought together in a logical format. At times this can be done mentally, but there are also times when it will require tables and matrices of data or even stacks of formal and informal reports. The nature of the organization will depend on the stated objective and is not important to anyone other than the tester. What is important is that extraneous material be temporarily set aside and that the format be useful to the tester.

Analysis/Synthesis.

From the data available, the tester attempts to draw logical conclusions — possibly about the outcome of the single test but more likely about a population of inference. He compares generally similar events to see what might be different and he compares generally dissimilar events to see what they might have in common. In analysis he resolves events into finer levels of detail to seek out reasons for observed differences while in synthesis he builds classes of events on the basis of potential causes for observed similarities. These processes are closely related to the inductive and deductive inferences discussed above. Conclusions are drawn in this step but they are only tentative. The next step is to scrutinize the tentative conclusion and see if it will really "hold water."

Testing the Conclusion.

This is not something that happens only at a single time in the derivation of a properly-stated, valid conclusion. There has already been some informal screening as the tester tries to draw tentative conclusions and retains only a few while discarding the majority as unreasonable. Now the testing or screening becomes more formal. A valid conclusion has two characteristics: it is supportable and it is appropriate. Consider these one at a time.

Supportable. The conclusion that is not supportable is in a very tenuous position. The tester cannot explain why he believes it to be true, the information-seeker does not know whether he should consider it as evidence in making a decision, and anyone who attaches any importance to it compounds the risk that it will be used as the basis for an incorrect decision. If a conclusion cannot be supported by test and evaluation data, it would be better left unsaid. In the long run it will be more useful to decide to buy the new AN/DBD-66 because SDB Corporation has an outstanding record in product service than to overlook poor product service while making a decision to buy on the basis of one unsupportable (and possibly incorrect) conclusion. Use the following checklist to determine whether tentative conclusions are supportable or not.

a. Is the conclusion supported by the weight of the data? In other words, does the data fit the conclusion? Both quantitative and qualitative information can be used to support a conclusion; a qualitative description of an event may be more useful than large amounts of quantitative data on it. If the data is not clearly in support of the tentative conclusion, the conclusion may be discarded outright or other variables may be investigated to see if the conclusion should be modified.

b. What is the impact of any missing data? Is the credibility of the (tentative) conclusion hurt by a lack of data over a certain part of the range of some variable?

c. Was the test sufficiently controlled? A big drawback of freeplay exercises is that they cannot be recorded in enough detail to permit a later repetition of the same exercise. If the subject test is not adequately controlled, the tester may have to admit he does not really know the conditions under which the data was collected.

d. Are effects confounded? The tester must be certain (and able to prove) that he has identified the cause of any effects he reports on and is not attributing an observed effect to the wrong cause or multiple causes.

e. Is the correct population of inference being reported? A review should be made to see if the population about which inferences are being drawn was correctly sampled or if, perhaps, a smaller population was sampled.

f. Was the test realistic? Was a prototype hardware item tested instead of a production item? Was the test accomplished by operational and support personnel of the type and qualifications of those expected to use and maintain the system when deployed? Was the system evaluated in the context of anticipated operational missions? Was the operational environment realistic?

g. Was the test design adequate? The design for primary factors should have permitted all the necessary comparisons to be made. The design for controlled background factors should have permitted proper sampling of the population of inference while maintaining control over the test conditions.

h. Did deviations from the test plan invalidate any of the data? If trials were rescheduled, the randomization plan may have been affected. If certain replications were not run, the precision was probably affected.

i. Have the simulations used been validated? Computer programmers are fond of saying, "garbage in — garbage out." The same holds true for any simulation, whether it is conducted on a computer or not.

Appropriate. A tentative conclusion could be supportable yet not proper for the question at hand. Use the following checklist to screen out inappropriate conclusions.

a. Is the conclusion relevant? The issue must not be cluttered by statements — true or not — that have no direct bearing on it.

b. Is the conclusion too broad? The tester should be addressing only the stated objective.

c. Is the conclusion trivial? Was this conclusion already recognized? Is it only repeating a widely-accepted idea?

The tester should not immediately discard conclusions that have been found supportable but not appropriate. They may be useful in addressing other objectives or they may be included without reference to any particular test objective.

If a tentative conclusion has passed the above tests, it may be considered valid for inclusion in a report on the test results. The statement of the conclusion can now be put in its final form.

Including Limitations.

The report reader should be left with no uncertainty about the meaning of a written conclusion. Two cautions are in order.

State the Population of Inference Clearly. The tester does not want to make claims for populations he has not really evaluated. Therefore there should be no opportunity for the report user to interpret the conclusions. Applicability should be quite clear.

State the Uncertainty. Conclusions, almost by definition, are never statements of an absolute fact. Consequently, the tester should report the certainty/uncertainty of his conclusions. Uncertainty comes from an inability to measure some characteristics more accurately or precisely, whether this is caused by limitations in the measurement process itself or limitations in sample size.

Uncertainty also comes from a lack of intent to measure some characteristic more accurately or precisely, simply because the information would not be operationally useful.

Further Data Organization.

The process described for arriving at a properly-stated, valid conclusion is repeated for each test objective that seeks more than raw data or for each question posed by a test objective. Then the tester should try to derive other information from the data by organizing the available data in other ways and repeating the process further. By taking a different view of the data he may see things that were not apparent before. Several ways of organizing the data are:

a. By System Characteristic. If not already covered in a test objective, what can be stated with regard to overall system effectiveness, supportability, suitability, compatibility, etc. ?

b. By System Function. What has been learned about the system's ability to fulfill different roles ?

c. By Mission Segment. What has been learned about the aircraft takeoffs, landings, refueling, ordnance delivery, etc. ?

d. By Each Measured Variable. What has been learned about system behavior with respect to each of the measured variables ?

e. By Interface. What can be stated about any machine-machine or man-machine interface ? Has the test uncovered any human engineering problems ?

f. By Population of Inference. If generalizations cannot be made about one population, can they be made about a smaller one ?

Before sending the report to press, the tester should insure that conclusions are not redundant and that they are consistent with each other. This is not the time to be too mechanical in approach. It will be helpful to sit back, put things in perspective, and take as broad a view of all the available information as possible.

5. FORMULATING RECOMMENDATIONS

A Test Officer formulates recommendations in much the same way he formulates conclusions. Wherever some sort of action beyond his responsibility seems indicated by the conclusions drawn or by the inability to draw conclusions, he may compare tentative recommendations. Then he subjects them to tests of supportability and appropriateness (only this time support should come from the conclusions or lack of conclusions) and makes their statement as unambiguous as possible. Six types of recommendations are commonly made:

To purchase/not to purchase.

To employ for a particular purpose.

To employ with certain techniques or within a certain envelope.

To make modifications.

To conduct further investigations into an apparent deficiency.

To conduct further tests to resolve uncertainty.

At this time it becomes apparent that a fourth test of appropriateness should be added. Does the tester have the type of information required to make recommendations in this area? Recommendations involving the acquisition of a system should be very carefully considered so that the operational tester does not imply that he has full knowledge of more areas (urgency, cost, alternatives, etc.) than he actually does.

6. CONCLUSION

As suggested at the beginning of this section, the conclusions and recommendations following a test can easily receive altogether too little attention. They will contain, however, whatever justification there is for the time and money spent in a test. If the conclusions and recommendations are not justified or simply wrong, the test will have been counter-productive. If the data is not combed for every inference that can be drawn, the Air Force has received less than its money's worth.

It will help insure action on the recommendations if the pertinent action organization is identified.

CHECKLIST FOR DRAWING CONCLUSIONS

1. Have all questions posed by the test objectives been addressed?
2. Have you tried to reach other conclusions by organizing the data according to:
    a. System characteristic?
    b. System function?
    c. Mission segment?
    d. Measured variable?
    e. Interface?
    f. Population of inference?
3. Is each conclusion relevant?
4. Is each conclusion a statement of inference, rather than a repetition of the test data or a recommendation for action?
5. Is each conclusion supported by data?
6. Is each conclusion drawn about the correct population of inference?
7. Is any data inconsistent with any conclusions?
8. Was the test realistic?
9. Was the test design adequate?
10. Was the test controlled sufficiently?
11. Are any effects confounded?
12. Has the impact of missing data been checked?
13. Did deviations from the test plan invalidate any data?
14. Have simulations been validated?
15. Is any conclusion trivial?
16. Is any conclusion too broad?

17. Is the degree of uncertainty for each conclusion clear?

18. Is the basis for each conclusion clear?

19. Is the population of inference for each conclusion clear?

20. Are the conclusions consistent with each other?

21. Are any conclusions redundant?

22. Do the conclusions cover all that was learned?

CHECKLIST FOR MAKING RECOMMENDATIONS

1. Do you wish to make recommendations regarding:

   a. Purchase decisions?

   b. Employment for a particular purpose?

   c. Employment techniques/envelopes?

   d. Modifications?

   e. Further investigations into apparent deficiencies?

   f. Further investigations to resolve uncertainty?

2. Do you have the type of information required to make each recommendation?

3. Is each recommendation relevant?

4. Is each recommendation supported by the conclusions?

5. Is any conclusion inconsistent with any recommendation?

6. Is any recommendation trivial?

7. Is any recommendation too broad?

8. Is any recommendation too narrow?

9. Is any recommendation ambiguous?

10. Are the recommendations consistent with each other?

11. Are any recommendations redundant?

12. Should any additional recommendations be made?

13. Is the organization to whom the recommendation applies identified?

Chapter 15

TEST REPORT

1. INTRODUCTION

The purpose of this section is to present a standardized format for the preparation of the Test Report to be required at the conclusion of any formally directed and planned Initial Operational Test and Evaluation (IOT&E) and/or Follow-on OT&E. The format and outline recommended here is as close as possible to a composite format of those used and developed within the Air Force Commands that carry out OT&E programs. The basic format has been progressively refined over the years and proved its effectiveness; concurrently a myriad of less effective formats have been tried and discarded. This proposed format will not be unfamiliar to either the Air Force testing communities or the readers/users of such reports. It has, however, not only been welded together from the best of the past and present, but also has been expanded to insure the inclusion of all inputs required by Headquarters USAF, and DOD Directives pertaining to OT&E and DSARC requirements. In addition to detailed guidance on format, and a wide range of pertinent areas the report may be required to address, this chapter also includes some recommended philosophy and practical pointers. These are directed at inspiring enthusiasm and professionalism in the preparation of the final Test Report, upon which will be based costly production decisions.

2. RESPONSIBILITY

In addition to the other demands made upon him, the Test Director is solely responsible for insuring that all reports required on a test program are prepared and distributed, in accordance with this manual, the Test Directive, Test Plan, and AFR 23-36. The Test Report will normally be the most complete and comprehensive document of the required reports; as a consequence it will also be the most difficult to plan, organize, write, and edit.

The Test Director should find this guide and checklist of considerable value in preparing any test report required during the conduct of the test program. Since the background and experience generally looked for in selecting a Test Director does not usually equate to his ability to write technically, the report writing details may be quite arduous and trying to him. As a consequence, the Test Director should make a determined effort to have someone assigned to the test team who is an expert in and enjoys technical writing, and insure that he is assigned the task of preparing specific reports, particularly the Test Report, starting from the inception of the testing program. By having the demanding details of this burden in competent hands, the Test Director will have significantly more time for running the best test program possible, with the maximum confidence that it will be properly and professionally reported.

3. PURPOSE OF THE TEST REPORT

The sole purpose of any test report is to communicate the results of the accomplished test(s) to the offices and agencies who require the information to carry out their functions and missions of planning, operating, and managing the Air Force. After the testing is completed, the data analyzed and calculated, and the conclusions reached, the test report is the terminal effort that must bring all of its preceding direction, planning, testing, and analysis to fruition. The most well planned, capably conducted, and comprehensively analyzed test is of virtually no value until it has been accurately and understandably reported to those in position to translate the test findings into sound and effective action.

Objectives. There are two primary objectives to be accomplished with the Test Report:

a. To disseminate information required by decision makers, planners, and operators.

b. To establish a formal, professional document of the test results as a future and permanent reference for the Air Force; and as appropriate, for the entire scientific and technical community.

Comprehensiveness. All formal reports of tests have certain common requirements in that each must state what was done, how it was done, how the results were observed or acquired, what they were, how they were analyzed, the conclusions reached, and usually the testers recommendations. The Test Directive authorizing the test being documented, and the Test Plan detailing its accomplishment will state the requirements and usually the desired contents of all required reports, which can run the gamut, from oral/phone reports, through wire, letter, preliminary interim and progress reports, to the final report. All of the written reports, however, should follow the same standard format as required by this manual, appropriately abbreviated or limited to be comprehensive and concise to the degree directed. The recipients of the reports and the use to which they will be put will quite logically establish the level of comprehensiveness required.

Timeliness. In the dynamic environment of the Air Force the value of test information may diminish rapidly with time. This fact makes the timeliness of test reports extremely important, and the tradeoff between completeness and timeliness becomes one of the Test Director's biggest problems, particularly so with the Test Report. The problem of timeliness can be solved, but it takes continuous and energetic action and planning to do it. History has shown that two of the most successful methods for reducing the time required to publish the report are:

a. Writing the initial draft of the report concurrently with the conduct of the test. This also avoids having to recall and remember important facts after the test is completed.

b. Delegate the responsibility for writing different parts of the report to different and knowledgeable test team members. This is particularly effective on the more extensive and complex OT&E's, because it permits the concurrent writing of the major portions of the report.

With a program of progressive report writing, it should be possible to produce the final test report promptly because the data and test circumstances will have been properly recorded while they were fresh in the minds of the participants, and writers. Also, in this case, final editing provides a more accurate review of the original test results and should enhance the quality of the report.

4. OUTLINING AND ORGANIZING

The format for the Test Report is shown and described in detail below. Although the format shown applies to the final report written at the conclusion of a test, it should also be used for written preliminary, interim, and progress reports, but abbreviated to the extent necessary to provide only the information required in the specified reports by the appropriate test directive and Test Plan. The reason for using the prescribed format are many, but the important reasons are:

a. Many years of Air Force Testing and hundreds of tests have shown this to be the most effective and expedient format for formally documenting test results. A myriad of less accommodating formats have been proposed, tried and failed through the years.

b. Experienced personnel and officials requiring the results of such tests are at home with the format, which greatly facilitates their extracting the needed information in the degree of detail that meets their requirements quickly, and easily.

c. It provides a practical, well organized, and easy to follow guide to assembling and writing a comprehensive, coherent, and logical report, which may be quickly reviewed as a unit from its basic framework.

d. It reduces report writing to little more than translating the outline into prose, with appropriate tables, charts, graphs, and pictures that will be effective in communicating the results to the readers.

e. When the report is being written jointly by a number of participants, the outline framework is the most effective way to show each the bounds and areas he is to cover, and how his efforts contribute to the whole report.

Although it can be anticipated that specific unique types of tests might be accommodated more appropriately with a different format, for reasons of readability and familiarity a determined effort should be made to use the standard format. This format also enjoys the advantages of being able to accommodate unusual test programs without requiring extensive changes to its general framework.

Format. The general format for documenting test results in written test reports is shown in Figure 15-1 (see "Checklist" for expanded version of this format).

Organizing. The format establishes the general framework of the report. Within this framework, however, there is considerable latitude as to the nature and quantity of information that is to be included in the report. Ordinarily the objectives and scope of the tests addressed in the Test Plan will provide appropriate guidance for a large part of the Report organization. For many test programs, particularly IOT&E's associated with major acquisition programs and there are a number of issues and specific items that must be addressed and included in the report to comply with directives and/or regulations promulgated by the Secretary of Defense, the Department of Defense, and Air Force Headquarters. Carefully addressing each of the items enumerated in the expanded outline detailed below will insure that all of the specific items to be reported on by higher echelon directives are given consideration in the organization of the test report. All items in the outline which fall in this "required" category are identified with an asterisk following the itemized paragraph number (e.g., iii.b*, 3.a*, or C.3.*).

Organizing the report includes another important aspect, in addition to structuring it in a logical manner: deciding what material is to be included in the report. Generally, during the course of testing, the accumulated data and information exceeds that which can, or should be, effectively included in the report, and consequently it must be evaluated and assessed for the merit of including it in the report. Usually this information will fall into three categories: (1) essential, and must be included in the

Preliminary Pages

- i. Cover/Title Page
- ii. Formal Review and Approval of Report
- iii. Foreword
- iv. Abstract
- v. Table of Contents
  - Lists of Tables, Graphs, and Figures/Illustrations
- vi. List of Abbreviations and Symbols
- vii. Glossary of Terms
- viii.* List of Related Previous Tests and Reports

*The actual numbering of these pages will be determined by the length of the preliminary elements (e.g., Table of Contents, Glossary of terms, etc.)

Report of Test Activity

1. Introduction
2. Purpose of OT&E
3. Method of Accomplishment
4. Discussion and Analysis
5. Summary

Annexes

- A. Detailed Description of Test Item/Tactics/Doctrine
- B. OT&E Environment
- C. OT&E Test Methodology/Design
- D. Supporting Data and Analysis
- E. Pre Testing Required/Accomplished
- F. Test Organization
- G. Test Operations
- H. Maintenance
- I. Logistics and Supply
- J. Personnel
- K. Training
- L. Safety
- M. Security
- N. Others

Completion Pages

Reference List and Selected Bibliography (if appropriate).
Distribution List.
DD-1473, Report Documentation Form.

Figure 15-1. Test report format

main body of the report; (2) relevant and supporting, which should be included but usually in appropriate annexes; and (3) superfluous, which is not significant enough or sufficiently relevant to be included in the report, but which generally must and should be saved for the period of time specified by command policy.

Outlines. This section provides the details and considerations relative to the Checklist. The Checklist, with these descriptions and directions should be used by test report writers as an outline guide and memory jogger to insure that each report is submitted in the proper form, and that none of the required information is omitted. It should also serve as a reminder of all the significant areas and items that the author is expected to cover. Bear in mind that there may be circumstances or special OT&E items which could require some alteration to the outline, or which may not be referred to in the present format. Responsible project officers and report writers should stay alert to this latter possibility to prevent the possibility of omitting some significant element in their report because it was not addressed here. The degree of detail required and those outline items which are not needed in a particular report will be a function of the nature, scope, and objective(s) of the test itself, and the previous guidance received from the Test Directive and the basic Test Plan for the OT&E being reported.

5. WRITING THE REPORT

This section is not intended to instruct in the art of writing per se; its purpose is to provide some useful practical guidance on how to communicate the results of tests promptly and effectively to the people that need them.

Report writers can communicate and present test results in several ways — with words, numbers, graphs, charts, and illustrations. A truly professional report is usually composed of combinations of these which present the desired information in the most concise manner possible. Some "musts" to be observed are:

a. The report must be written in the third person. Never use I or we; the report is to convey what was done, not who did it.

b. The report must not be so stylized as to detract from the importance of the subject and the technical presentation. Also avoid the use of innuendo, humor, sarcasm, emotion, melodramatics, and the like.

c. Jargon must not be used. Although it might improve communication within a given profession, technical reports are rarely confined to one segment of the technical/procurement community and jargon generally will tend to inhibit accurate communications with all readers.

For the most part, the information used in the Test Report will have been acquired during testing. Usually this information and data will become available as the testing progresses. Trends in testing will usually be discernable, and a preview of the content and emphasis of the final report can be established before the actual testing is completed. Report writers should take advantage of this early information and data, and have the preliminary draft of the report written prior to the completion of the actual testing. This progressive and concurrent report writing has some very significant advantages. The most obvious advantage is that it should permit the final test report, and any interim reports, to be published promptly. Another important advantage is that the data and related circumstances are documented immediately, which is more effective than attempting to remember and reconstruct the tests from scant notes and raw data considerably after the fact. Still another plus is that it gives the writer an opportunity to review his initial writing, with probably a considerably broader view and seasoned background, to evaluate it in a more knowledgeable and meaningful manner than when it was originally done. The result will be enhancement of the professionalism and value of the Test Report.

6. EDITING THE REPORT

Some final pointers for polishing the final product before formal publication are:

a. It is virtually impossible to write a cohesive report directly from an outline when a group of team members has each authored a different section of the report. The resulting variations in style, technique, completeness, and clarity will necessitate very conscientious and probably extensive editing.

Generally the larger and more complex the test program being reported upon, the longer and more difficult the Test Report will be, and the more extensive the editing required to publish a really good report — one worthy of all the work and effort that has gone into planning, executing, supporting and reporting the whole test program.

b. The pressure to publish a timely report to meet a compulsory deadline may curtail editing time and effort. These circumstances should be anticipated and planned for to insure that efforts of the Test Director and the test team are not degraded at the last minute by producing a hurried, unpolished, and inadequately edited report.

c. It is usually a good idea, particularly if professional technical writers are available, to have someone rather than the original author edit his writing after the author has edited his own efforts. Final editing of one's writing is usually difficult because everyone has a tendency to read what he intended to write as opposed to what he may have actually written. Editing by someone who can assume the objective view of the intended reader can be an immense help.

d. A technical editor, if available, can also save a lot of time toward publishing the report by more expediently correcting grammatical errors, checking numbering of graphs, tables, illustrations, etc., checking proper cross-referencing, and insuring that proper paragraphing and form is maintained throughout. Any changes he does propose should always be concurred in by the Test Director or team members to insure that no change in meaning is inadvertently made and published.

e. Remember when the first draft is finished, most of the hard work is done. The subsequent revisions and polishing take time and patience, but will produce a more rewarding report. The extra effort will be appreciated all the way up the line, as well as throughout the contemporary technical community.

7. INSTRUCTIONS FOR USING REPORT FORMAT

The following instructions for use of the Test Report format are keyed to the format by paragraph numbers.

PRELIMINARY PAGES

i. Cover/Title Page.

Each of the Major Commands (MAJCOM) has developed a specific and standardized cover format that contains a characteristic and artistic command emblem and color that identifies the originating command. The remainder of the information required on the cover must be provided to properly and specifically identify the report and the test program it documents.

Should there be any Command restrictions on the distribution of the report, this information and identification of the office having distribution authority for the report should be specified on the cover.

ii. Formal Review and Approval of the Report.

This page is required to authenticate the report. It must show the assigned project number, project title, the organization responsible for the project, and that the report has been reviewed and approved by the AFTEC Commander or the major command headquarters prior to its distribution outside AFTEC or the Command as required by AFR 23-36. This will be a single page following the Cover/Title Page, and is not numbered.

iii. Foreword

The foreword should consist of a short paragraph that states the authority for conducting the test, the type of test conducted, where the test was conducted, and what organization(s) actually accomplished the testing. The initial paragraph should be followed by a list of the key participating personnel to show their responsibility for the Test Program, and should indicate their rank and organization. Any kudos for organizations or personnel significantly supporting the test organizations may be briefly added to this section.

The first page of the foreword should start the lower case Roman numbering of the preliminary pages; i.e., "i", which should be carried on in sequence.

It should be noted that the identification of the type of test and test organization(s) must be specified in accordance with DOD directives.

iv. Abstract

The Abstract should provide a very abbreviated narrative summary of the tests conducted, critical issues addressed, and significant or major conclusions reached.

v. Table of Contents through vii. Glossary of Terms

These lists are self explanatory; however, it should be emphasized that they are very important in saving the reader much time and lost motion if he is interested in only certain areas of the report. Also it can be time consuming and frustrating to search for the meaning of an unfamiliar abbreviation "seen before someplace in the body of the report." Hence, the List of Abbreviations and Symbols, and the Glossary of Terms must be complete and easy to find. Refer to AF Manual 11-2 for widely used Air Force abbreviations, and guidance in forming new ones.

viii. List of Related Previous Tests and Reports.

This list must include any tests and reports of tests accomplished or in progress, which have been or are being conducted in conjunction with the test being reported on. Also included in the list should be any test known to the report writers or testers that are related to the present test(s); for example, tests of the same equipment in different test environments or with different objectives.

Any other tests known to the writer(s) that they feel might be of significant interest to readers/users of their report may also be included. The list should include the test title, dates of the tests, the report number and date, and the project number.

REPORT OF TEST ACTIVITY

This is a title only and indicates the beginning of the basic test report. The tenor of this report should be a concise, factual reporting of critical issues addressed, significant factors, objectives, and results of the test(s). Details, supplementary information and other than primary and essential results and data should be relegated to the Annexes. Liberal references to the details in the Annexes should be made in this portion of the report, so that interested readers may be accurately directed to more details and specifics should they be so interested.

Arabic numerals starting with page 1 are used to number the pages in this section of the report.

1. INTRODUCTION

The introduction contains a succinct description of the tests that were conducted and the directives that authorized them. It should include:

a. Background of the Test Requirement which may encompass the Required Operational Capability (ROC); include a problem area, or the rationale that established the requirement for the tests conducted. A short history of what has been accomplished or transpired preceding the current test program may be appropriately documented.

b. Description of the Test Item, Tactic, or Doctrine that was tested and evaluated should be given in sufficient detail to adequately identify what was tested, and to permit proper differentiation of similar or related items; or to delineate the changes to a specific item which has made it a candidate or subject for testing. References to Annex A should be made for a complete and comprehensive detailed description. Although generally referred to as "test item" herein, the subjects of OT&E may well be something other than hardware, e.g., software, tactics, doctrine, operational procedures.

c. Scope of Test(s) is a broad brush statement covering the original test limits, and any limitations observed during the course of the subject testing program.

2. PURPOSE AND OBJECTIVES

This section contains an abbreviated presentation of why the test program was undertaken and its major objectives, which are amplified upon in the following subparagraphs.

a. Overview of Test Requirements is a restatement of the purpose of the test from the Test Plan for the project, expanded to include any subsequent changes to the plan and test requirements during the test program so that all requirements actually addressed during the test program are identified.

b. Specific Objectives are also a reiteration of the critical issues and test objectives addressed, as well as any subsequently added changes or deleted issues and objectives occurring during the conduct of the test program.

c. Known or Anticipated Discrepancies Addressed permits the report writer to acknowledge and specifically respond to Test Plan direction to investigate and evaluate any known or suspected discrepancy or problem area during the test program. If no such specific requirements are stated in the Test Plan this subparagraph should be omitted in the report.

3. METHOD OF ACCOMPLISHMENT

This section discusses how the OT&E was carried out. It should generalize and summarize but nevertheless present a reasonably complete picture of how the testing was done, including the effects of the areas discussed below on test accomplishment. There should be ample references to appropriate Annexes for details and specifics. The areas to be considered in this section include:

a. Test Schedule should summarize when certain tests were accomplished, precedence/priorities assigned to test series, the time span covered in the test and/or specified test series, and the impact on the overall test program. Any failure to meet an established schedule or milestones must be noted and explained.

b. Test Procedures should address how data was gathered and recorded, types of missions (profiles) or tactics flown, or tests accomplished on hardware. The effects or consequences of data collection on the authenticity of operations, procedures, or function should be noted in general terms, with adequate references to appropriate Annexes for details.

c. Test Personnel must be identified as to experience and background. Specialists specifically selected to enhance testing, or specially trained technicians used for testing are to be highlighted. If random selection of "average" operational personnel is used, the consequences of their use, as well as the effects of specialists on test results, should be addressed. Again, specific and lengthy details should be presented in Annex J.

d. Measure of Effectiveness components are Availability, Dependability, and Capability. If effectiveness is an objective of the test, all of these components must be addressed. How the components were obtained, derived and evaluated should be discussed in sufficient detail that the results of the test(s) discussed in Paragraph 4 below are meaningful to the maximum extent and their significance is completely understood (see Chapter 6).

e. Test Environment should be summarized so that the essentials and salient features of the testing environment, particularly in relation to the anticipated, or known, operational environment are recognized. All of the significant details of the test environment and the consequences of them on the test program and test results will be addressed in Annex B, and it should be referred to as required to effectively augment this subparagraph.

f. Changes/Deviations from Test Plan or Test Directive must be identified and explained in this subparagraph if they effect the purpose and objectives of the test program. This summary should refer to Annex D-4 for details of changes, deviations, omissions, or additions to the programmed and planned tests, and for information on all other changes to the plan that do not necessarily or directly impact upon the test objectives.

4. DISCUSSION AND ANALYSIS

This paragraph is the heart of the entire test report because it presents the results of the test conducted and the consequences of the evaluation of those results. This is the reason the tests were conducted. This is what all the time, money, manpower and test resources were expended to produce. It must be comprehensive to the extent that all results affecting critical issues to be addressed and significant test objectives are discussed logically, accurately and completely; and that all conclusions are amply substantiated by properly validated or technically sound data.

Each subparagraph (1), (2), etc., should address as a unit, the results, conclusions, and recommendations appropriate to each critical issue and test objective. This will provide related continuity to each objective and save the reader from trying to track through several pages of results followed by several pages of conclusions and then another several pages of recommendations. Also, as a part of each subparagraph, any actions taken during the course of the test program to correct or remedy any known or discovered discrepancies concerned with the issue or objective should be identified. Any known or substantiated result of such recommended or accomplished actions should also be described and evaluated.

Issues, objectives and questions which could not be properly evaluated because incomplete, inconclusive, or unsubstantiated data prevented the acquisition of significant or adequate test results, should be reported in these subparagraphs. Any valid results, conclusions, recommendations, or information that would be of value to the report reader should be included here, as well, even if it is not related to a specific objective.

5. SUMMARY

The Summary provides the medium whereby the significant management aspects of the test program can be elaborated upon, when reporting upon them has been directed, or when the Test Director feels that such reporting may be productive, valuable, or historically significant. The following subparagraphs indicate those areas frequently of interest or generally requiring expansion by DOD or higher headquarters directives or regulations. Obviously, as the existing directives and regulations are changed and modified to meet future requirements these areas of interest may change also. The major areas to be considered in this section include:

a. Milestones/Schedules not Met must be specifically identified. Valid significant reasons for failing to meet prescribed schedules should be identified and explained as an aid to more effective future planning. These reasons can be many and varied, anticipated or unexpected, technical, fiscal, political, administrative, etc. Examples might be an unanticipated need for additional training, or technical capability; inadequate funding, manpower, or test resources; lack of logistic support, or unexpected high hardware failure rates; weather; strikes.

b. Resources Overruns will be identified with the reaons and extent of the overrun(s) explained. The impact and effect on the test program of encountered overruns should be addressed.

c. Evaluation of Tradeoff Studies should be made if such studies were accomplished during the test program. If, as a consequence of the test conducted, the Test Director, or Test Staff feel that there may be merit in conducting additional tradeoff studies, they should be recommended in this subparagraph.

d. Plans for Future Testing and their relationship to the test program being reported upon, should be identified as they become known to the test staff.

e. Management Summary provides the Test Director with the opportunity to document significant problem areas encountered during the conduct of the test program and the execution of the Test Plan. He should also make any appropriate recommendations that he feels could improve the effectiveness and expediency of future test programs.

ANNEXES

This is a title page only and indicates the beginning of the detailed annexes. The annex identifying letter (A, B, etc.) and its specified subject matter are to be maintained as indicated in the Checklist.

Those subjects which are sufficiently explained in the body of the report, or which are not appropriate to the test program conducted may be omitted, but its title letter is to be omitted as well. In this manner each lettered annex (A thru M) will have the same subject in all reports, which will facilitate both reading the report and finding specific details.

Should the nature of the test program, the test item, or special requirements or direction, necessitate additional and/or supplementary annexes, annexes N and beyond may be used as appropriate to accommodate those unique requirements.

Each annex is to be a unit within itself with outlining and paragraphing as required by the length, quantity, and complexity of the details to be included. Information included in one annex should not be repeated in other annexes, although adequate cross referencing may be used as needed.

Page numbering in the annexes will be with arabic numerals in sequence starting with 1 and running the length of each separate annex, the number on each page must be proceeded by its appropriate annex letter (e.g., A-1, A-2, A-3, B-1, B-2, etc.). To provide neat and obvious separation of the annexes, each annex should have its own title page which will be blank except for the annex title centered and its subject directly under it, approximately one third of the way down the page. (e.g., ANNEX B, OT&E ENVIRONMENT) (see Checklist at end of chapter).

Further explanation and breakdown of the annexes than that provided in the Checklist will not be provided here. The outlining provided will show the nature of material, information and details required and the general format desired. The author(s) should bear in mind that the annexes are the places to include all the information, procedures, details, and rationalizations, etc., that are in any way significant and appropriate to test program accomplished, and particularly those not discussed or noted in the Report of Test Activity section of the report. Remember that the annexes are for the report users who need to know all the significant details about one, a few, several, or all areas of the OT&E Test Program; as opposed to the users whose need is only to know generally what was done and what resulted.

REFERENCE LIST AND SELECTED BIBLIOGRAPHY

References should be numbered sequentially as they appear in the text of the report, and thus will be listed in this section in the same order, as is standard referencing procedure.

Probably few reports will require this section; however, it is not an unlikely requirement if the testing conducted is follow-on or related to previous tests whose reports are used to avoid repeating subtests or to extract appropriate data.

If studies are accomplished in conjunction with or as a prerequisite to the test conducted any reference material should be acknowledged and listed.

Should special or unique test procedures, data reduction or acquisition methods, etc., be used during the test program, the Technical Reports, operating instructions or similar documents used should be referenced and listed. In many cases, the use of such references can substantially reduce the amount of detail necessary to define, describe, or explain methods and procedures used in the tests being reported upon.

DISTRIBUTION LIST

The basic and required distribution list will be provided in the Test Plan and Test Directive, as well as any Command or higher headquarters restrictions to the distribution of the Test Report.

The author(s) should, however, keep in mind that frequently during the conduct of the test program technical assistance or other unscheduled support may be required which will probably add new units or agencies to the planned distribution list.

Also, the Defense Documentation Center (DDC) is required by AFR 80-44 to receive twelve copies of each unclassified and unlimited report, and two copies of each classified or limited distribution report.

DD-1473, REPORT DOCUMENTATION FORM

DD Form-1473 is required to be completed in detail as outlined by MIL-STD 847-A. It will be accomplished and included in all technical reports prepared by or for Department of Defense organizations.

CHECK LIST FOR PRELIMINARY PAGES

i. Cover/Title Page

   a. Major Air Command Sponsoring or Responsible

   b.* Organization Publishing the Report

   c. Type of Test and Name of Test Item

   d. Type of Report

   e. Project No.

   f. Date

   g. Security Classification and Downgrading Information, if Classified

ii. Formal Review and Approval of Report

iii. Foreword

   a. Authority for Test

   b.* Type of Test (Independent, Combined, Joint)

   c.* Test Location(s)

   d.* Test Organization(s)

   e. Key Participating Personnel

iv. Abstract

   a. Narrative Summary of Tests

   b. Significant Results

v. Table of Contents

   a. List of Contents

   b. List of Tables, Charts, Graphs and Figures

vi. List of Abbreviations and Symbols

vii. Glossary of Terms

viii. List of Related Previous Tests and Reports

CHECKLIST FOR REPORT OF TEST ACTIVITY

1. INTRODUCTION
   - a.* Background of the Test Requirement
   - b.* Description of Test Item, Tactic, or Doctrine
   - c.* Scope of Test
2. PURPOSE AND OBJECTIVES
   - a.* Overview of Test Requirements
   - b.* Specific Objectives
   - c.* Known or Anticipated Discrepancies Addressed
3. METHOD OF ACCOMPLISHMENT
   - a.* Test Schedule - Summary
   - b.* Test Procedures
   - c.* Test Personnel
   - d. Measure of Effectiveness
   - e.* Test Environment
   - f.* Changes/Deviations from Test Plan
4. DISCUSSION AND ANALYSIS
   - a.* Results (by objective and critical issues)
     - (1)* Conclusions
     - (2)* Recommendations
     - (3)* Actions Taken to Correct Discrepancies
     - (4)* Known Results of Actions on Recommendations
   - b. Issues, Objectives, and Questions not Completely Evaluated
5. SUMMARY
   - a.* Milestones/Schedules Not Met
   - b.* Resources Overruns
   - c.* Evaluation of Trade-off Studies
   - d.* Plans for Future Testing

     Relationship With Test Program Being Reported
   - e. Management Summary

CHECKLIST FOR ANNEXES

A. DETAILED DESCRIPTION OF TEST ITEM/TACTICS/DOCTRINE

1. Description of Hardware/Tactics/Doctrine Tested
   a. Special Installation Requirements
   b. Problems and Significant Difficulties Encountered
   c.* Major Subsystems Not Tested
2. Deviations from Operational Configuration

   Major Subsystems Modified for Test
3. Configuration Changes During Test Program
4. Anticipated Configuration Changes That Will Impact Test Results

B. OT&E ENVIRONMENT

1. Authenticity of Physical Environment
2. Authenticity of Tactical Environment
3. Environmental Effects on Validity of Results
4. Constraints and Limitations Imposed/Encountered

C. OT&E TEST METHODOLOGY/DESIGN

1. Test Procedures Used
2.* Simulations
   a. Physical
   b. Computer
   c. Authenticity of Threat Simulations
3.* Models/Modeling Employed
4. Data Collection Methods

5. Data Reduction Procedures

6.* Data Validation Procedures

7. Instrumentation Constraints and Limitations

8. Describe the analytical/statistical approach being used to answer questions posed by the test objectives. This description shall include:

(a) Listing of independent and dependent variables for each trial or group of trials (as applicable) and description of the statis cal handling of each independent operational variable (i.e., investigated at different levels, blocked, randomized, or held const t; factorial or nonfactorial arrangement;

(b) Basis for establishing scope of the test;

(c) Rationale for selection and description of scenarios to be used;

(d) Factors driving degree of operational realism (including physical environment, operating and maintaining personnel, test hardware);

(e) Statistical basis for replication;

(f) Prediction of uncertainty in results;

(g) Effect of resource limitations on test design;

D. SUPPORTING DATA

1. Test Results, MOE, Procedures

2. Specific Mission Profiles

3. Special or Significant Equipment Settings

4. Deviations from Test Plan or Test Directive

5. Copies of all Specialized Forms and Questionnaires Used

6. Data Storage and Retrieval (DMIS Interface)

E. PRE TESTING REQUIRED/ACCOMPLISHED

1. Concepts or Opinions Confirmed or Verified

2. Questionable Variables or Parameters Identified

3. Novel or Unique Test Methods Confirmed

F. TEST ORGANIZATION

1. Deviations from Normal Organizational Structure
   a. Rationale for change
   b. Effects and consequences of changes

2. Evaluate Probable Performance of Standard Organization

G. TEST OPERATIONS

1. Test Operational Environment vs. Normal Operations
   a. Rationale for change
   b. Effects and consequences of changes

2. Evaluate Probable Performance with Standard Operational Procedures

3. T.O. and Operational Procedures Validation/Evaluation

H. MAINTENANCE

1. Reliability

2. Maintainability

3. Availability

4. Compatibility

5. T.O. Validation

6. Significant Problem Areas or Difficulties Encountered

I. LOGISTICS & SUPPLY

1. Supportability

   a. Peculiar Problem Areas and Difficulties Encountered

2. Deviations from Standard Logistics Procedures

3. Life Cycle Cost Analysis

   a. Results

   b. Determination Methodology

J. PERSONNEL

1. Manning Deviations from Standard UMD

   a. Rationale for Change

   b. Effects and Consequences of Changes

2. Evaluate Probable Performance With Normal UMD

K. TRAINING

1. Training Accomplished Prior to Testing

2. Training Acquired during Testing

3. Effect on Test Results

4. Recommended Training Requirements

L. SAFETY

1. Safety Regulations Not Complied With

2. Unique Safety Procedures Employed

3. Effect on Test Results

   a. Environmental

   b. Operational

   c. Data Acquisition

4. Recommended Safety Requirements

M. SECURITY

1. Special Security Procedures Required

2. Effect of Security on Test Results

   a. Environmental

   b. Operational

   c. Data Acquisition

3. Recommended Security Procedures

N. OTHERS

As needed.

CHECKLIST FOR COMPLETION PAGES

Reference List and Selected Bibliography (if appropriate)

Distribution List (Include 20 copies to DDC)

DD-1473, Report Documentation Page (Required by MIL-STD 847-A)

Chapter 16

JOINT OPERATIONAL TEST AND EVALUATION

1. ESTABLISHING A NEED FOR JOT&E

Joint Operational Test and Evaluation has several meanings that are best understood with reference to several authoritative sources.

The Blue Ribbon Defense Panel Report of July 1970 says that "JOT&E is operational testing where the new system is tried alongside of, or against, capabilities of another military service to estimate more clearly the system's joint operational worth."

DOD Directive 5000.1 "Acquisition of Major Defense Systems" (July 1971) states:

> "For programs involving two or more Components (Military Departments), the Component having dominant interest shall designate the program manager . . . . test and evaluation shall commence as early as possible. A determination of operational suitability, including logistics support requirements, will be made prior to large scale production commitments, making use of the most realistic test environment possible and the best representation of the future operational system available."

DOD Directive 5000.3 "Test and Evaluation" (Jan. 1973) describes joint testing as follows:

> ". . . . For those systems which have a natural interface with equipment of another Component, or may be acquired by two or more Components, joint OT&E will be conducted where required. Such joint testing will include participation and support by all affected Components as appropriate . . . . The Deputy Director of Research and Engineering, Test and Evaluation, (DD(T&E) is assigned across-the-board responsibility for OSD in Test and Evaluation matters. (His stated responsibilities include) . . . Monitoring closely such joint testing as is accomplished by the DOD Components in connection with their planned acquisition of specific systems. In addition, initiating and coordinating the accomplishment of

such additional joint testing* as is necessary, with specific delegation to an appropriate Component (or Components) of all practical aspects of the joint test . . . . Monitoring, only to the extent required to determine the applicability of results to weapon system acquisition or modification, that test and evaluation which is (1) Directed by the Joint Chiefs of Staff which relates to the Single Integrated Operational Plan (SIOP) operational factors; (2) Conducted primarily for development or investigation of organizational or doctrinal concepts."

Joint operational test and evaluation in its actual conduct differs little from any OT&E conducted by a single Military Department. Therefore, the test planning, test design, measures of effectiveness, test reporting, and other related chapters of this manual are equally applicable to the USAF member assigned and involved in JOT&E.

There are, however, some very important differences. These include the following:

Policies regarding the need for JOT&E

Validation of the need

Directive authority for JOT&E

Military Department OT&E structure

Organization of a JOT&E team (or force)

Resources source

Approval chain

Political arena

Policies regarding the need for JOT&E basically stem from the Office of the Secretary of Defense. The actual statement of need can originate in OSD or from any DOD component and often does.

Although there are no specific restrictions placed on where and how to seek and identify needs for joint test and evaluation, at least we have the general guidance of DODD 5000.1.

*Several of these DD(T&E) initiated Joint Tests are noted in Appendix B.

and .3 as cited above. This guidance suggests the evaluation of major and selected less-than-major systems. Within this category, systems exhibiting the following characteristics are of primary interest:

Systems which are developed by one DOD Component for use by two or more DOD Components.

Systems evaluations which can be significantly improved or be more realistically evaluated through employment of two-sided testing techniques.

Systems which have interoperability, or inter-supportability relationships involving other DOD Components.

Systems which embody employment concepts, doctrine and tactics not fully assessed, supported or adopted on a joint service basis. In most cases these involve critical issues which are of joint interest in the OSD, DCP coordination and DSARC decision process.

It is the general policy of OSD to require joint test and evaluation of major and selected non-major Defense Materiel Items in the acquisition process, whenever:

New or competing military employment concepts are associated with critical issues of joint interest in the OSD decision making process; and, technically valid and cost effective ways of reducing decision uncertainty through test and evaluation are identified.

Defense Materiel Items are developed by one DOD Component for use by two or more DOD Components.

Defense Materiel Items have clearly identified major inter-operability or inter-supportability relationships with operational mission elements of other DOD Components.

Materiel Item acquisition decisions require additional data which can be optimized through employment of "two-sided" test and evaluation techniques involving two or more DOD Components.

ıL'? ΛC

or, whenever it is necessary to:

Validate the role of the system in specific military situations.

Resolve or aid resolution of DSARC questions and critical issues.

Measure systems operational performance parameters in a realistic field and ILS environment.

Develop or validate engagement models and factors used for operational planning, force development, doctrine and tactics.

Evaluate systems modifications and improvements.

When it is determined that a specific Defense Materiel Item in the acquisition process meets one or more of these criteria, the DOD Component advocating acquisition of the item will identify joint test and evaluation options in draft DCPs submitted for OSD action. Generally, one of the following circumstances will govern such action:

The required operational capability was generated by another DOD Component or in parallel with another DOD Component.

Harmonization actions (see AFR 57-1) arising through the DCP coordination process identify joint concerns or considerations which can be quantified through test and evaluation.

The Defense Materiel Item to be acquired is related through "in progress" development or modification of a materiel item of another DOD Component.

2. VALIDATING THE NEED

Having a basic need for JOT&E identified, it is now necessary to move through a validation step. Reasons are obvious — in most cases significant resources will be committed to the action; further, not all JOT&E needs lend themselves to a reasonable technical solution within a required time frame.

Normally this validation is made by a feasibility study working group formed at the DD (T&E) level. The objective of this group is to develop and examine test design concepts for technical adequacy and efficiency relative to the stated test objectives, schedules and resources. The feasibility study working group is chaired by a member of the DD(T&E) staff who is responsible for the organization and selection of WSEG/IDA and other working group participants as necessary to fulfill the above stated objectives. The feasibility study conducted by this group provides the baseline input to subsequent planning. Thus, the group's output obviously becomes highly important to the JOT&E Test Director and his team. This assumes, of course, that the need was validated and the JOT&E will take place.

For the USAF officer who might find himself involved in a JOT&E action, it is recommended that he not use the feasibility study working group's output without proper study. This documentation must be carefully assessed for practicality, completeness and technical adequacy. Some questions which will enlighten this area are:

Who were the members of the Joint Test Feasibility Working Group? What is their background and experience in OT&E? How many operating command people participated in the study or review of its recommendations?

Who was the principal author(s) of the study report? Will he (they) be assigned to the JTD for follow-on planning and joint test program implementation?

Are the stated purposes and objectives for the test program general or specific? Examples of typical over-generalized statements which signal trouble are: Resolve or aid resolution of (unspecified) critical issues; or validate the role of the xyz system in combat situations (unspecified).

How does the test concept compare with current operating command and intelligence estimates of the threat, doctrine and tactics from both red and blue viewpoints? Are deviations identified, explained and justified?

How were the requirements for type and number test items determined? Are units specifically defined? For example: "Strike sized element" is not a specifically defined unit.

Has test design planning been a part of the feasibility study? Was an easy or difficult area selected? How far did they carry it? What are the characteristics of the related back-up documentation not appearing in the study report?

Are environmental factors involved in the test? Do they occur in usable testing areas? Has the frequency of occurrence been checked by authoritative sources?

Is the quality and quantity of data specified? Can it be practically related to the amount of time and funds provided for the evaluation?

3. AUTHORIZING TESTING

The authority to direct a specific JOT&E resides with the DD(T&E) as noted above in the excerpt from DOD Directive 5000.3. Joint testing which is determined to be feasible is implemented through DD(T&E) designation of an Executive Agency (DOD Component or other DOD agency) which provides the Joint Test Director and exercises responsibilities for the accomplishment of all of the practical aspects of the joint test and evaluation. DD(T&E) exercises overall management monitoring of the joint test and evaluation activities and provides planning, programming and budgeting support for unique test costs.

4. ORGANIZING FOR THE TEST

Now that we have a validated need and a directive to proceed, it is next important that we get organized.

Obviously, in order to effectively organize a test team or force using personnel (and procedures) from other than ones own parent department, a thorough understanding of the OT&E structure and procedures used by each of the Military Departments is necessary to anyone who finds himself involved in Joint OT&E actions. This information is provided in Appendices B, C, D, and E. A summary of the major interfaces is shown below for ready reference.

Army	Navy	Air Force	Marines
ACS FOR*	Asst. Dir. RDT&E (OT&E) OP-98C	DC/S (O&P) AFXOOW	Spl Asst./OT&E DC/S (RD&S)
OTEA Ft. Belvoir, Va.	OPTEVFOR Norfolk, Va.	AFTEC Kirtland AFB, N. M.	Same as Navy

HQ & field JOT&E focal points

This participation ranges from comments and inputs to draft DCP's, maintenance of five-year OT&E plans and programs, to test management and conduct of independent evaluation and reporting of test results. The identified Headquarters staff offices maintain liaison with DDR&E, DD(T&E) as necessary to fulfill requirements for coordinated OT&E actions.

The organizations of joint test forces are in most cases an extension of "ad hoc" test feasibility working groups formed in the "verification of need" phase. The alignment of these working groups is usually along lines that permit assignment of responsibilities to each participating agency or service. For example, in the joint Electronic Warfare test there were functional areas as follows: Simulation, Instrumentation, Integrated Air Defense System (Red Force), Data Reduction, Analysis, Operations, Communications/Electronics, and Plans and Programs.

The organizational phase begins when an executive agency has been designated and the Joint Test Director is identified. Experience has shown that the transition to the organization phase has been most smoothly accomplished when the selected JTD has been extensively involved in the earlier test feasibility study working group activites.

An initial organizational consideration is the characteristics of the implementing directive issued by DD(T&E) establishing the joint T&E program. Of equal importance are the charter which is drawn by the Joint Test Director for approval of DD(T&E), and memorandum agreements or Letters of Instruction concluded between the participating DoD Components.

5. ROLE OF THE JTD

In the development of a charter the JTD should clearly and appropriately define the scope, responsibility, and authority necessary to implement and conduct the program.

Sound plans and effective management are only possible through positive action to establish a clear baseline for their development. Responses to the initial questions (such as what do you want and what do you need?) must be carefully thought out. The following checklist will assist in insuring thorough consideration of these matters since they are usually confronted before adequate thought has been given to planning and organizational requirements.

The Situation Driving the Requirement

Is a logical basis for T&E action provided and is it supported by complete information in the sources of background information? How was it developed and by whom?

Are the fundamental objectives clearly stated and are they consistent with the basis for T&E action?

How are T&E concepts and solution options defined? Are they directive or otherwise restrictive? Do they involve untried, high risk or unique planning characteristics or technical factors?

Have all major participants and related development agencies been identified? What are their official views concerning support, resources, schedules and priorities?

Resource Provisions Cited

What financial planning, programming and budgeting actions have been taken relative to the Joint Test program? What are the funding targets and limits?

Are responsibilities for Planning, Programming and Budgeting System (PPBS) actions identified? What program elements, funding categories, accounts, etc. are involved? Can they be adjusted? On what authority?, Within what limits? And on what schedule?

Have test staff manpower estimates been made? What standards and precedents were used? What latitude is provided for change?

What provisions have been made for test items, instrumentation support equipment and facilities? Are existing contracts involved? What is the contracting officer's and system program manager's positions with respect to change if it should be required?

Will installations within the DOD Test & Evaluation Facility Base be involved, e.g. National Ranges? If so, have Range Commanders' Statements of Capability been issued? What is their position relative to funding needs and other resource capabilities?

Will the test program involve military construction? What PPBS provisions have been made?

Schedules & Milestones

Does the directive document contain information on phases and completion dates? If so, are they adjustable based on detailed planning actions? Under what conditions?

What is the ratio of time allowed for planning versus test operations? Note: For relatively straightforward tests not requiring design and development of specialized instrumentation, equipment or facilities the ratio should be at least 3 to 1.

Do schedules provide for a planning review period? Note: The planning review period should be at least half the length of the basic (initial) plan development period.

Are the schedules consistent with availability of test items? Is adequate time allowed for test item modification? Has this been verified by the Development Agency, Logistic Agency and contractor, if involved? Are allowances made for contingencies, e.g., damage in transit, missing parts, or field reliability factors.

The foregoing discussion on organizing for JOT&E assumes a directive has been issued and that the directive deals with a perfectly feasible test series. This conclusion should not be without some question. If you are involved in a JOT&E assignment examine the scope of the test yourself. Asking the following questions will prove helpful:

How does the test force size and composition compare with that planned for a real military engagement?

What are the boundaries of the system to be tested?

Are the number of test samples selected considered sufficient to realistically exercise the threat spectrum described in the system's design objectives?

Is the testing appropriately spread throughout the system's performance envelope?

Are simulations identified and can they be realistically related to the elements of the test model?

Does the testing involve development of new or unproven techniques, instrumentation or unusual test item modifications?

Does the test involve live firings and if so have the safety and environmental considerations been assessed?

Are specialized training or other ILS factors involved in the testing concept?

Is an environmental impact statement required?

6. RESOURCES FOR TESTING

Assuming need, validation, directive, some initial organization and planning, and a well thought out charter (agreed to by the directive authority) the topic of resources must be considered. Two aspects of resources are important: (1) what do you need, and (2) where are you going to get it?

In JOT&E, you have to depend on several sources including your parent department, other participating departments, and the DD(T&E). Take careful note of this in the following discussion.

The task of determining requirements for resources is an essential early planning step. These are addressed in terms of the four established DOD Appropriations categories; i.e., Manpower, Items of Equipment (procurement), Operations and Maintenance; and Military Construction.

The Five-Year Defense Program is the heart of the DOD Planning, Programming and Budgeting System. Understanding of the PPBS and its objectives and implications is important to the JTD because any program which cannot be realistically accommodated within it is going nowhere. There are substantial procedural hurdles to be passed in gaining approval in this "survival of the fittest" competition for resources. The JTD must give particular consideration to four of these problems in his initial planning phase.

Manpower. Organizations are people. The characteristics and combinations of talents found in the selected people are a primary factor in determining an organizational approach. Finding competent and appropriately experienced test and evaluation professionals who can be made available to the joint test force staff is very difficult, particularly when relatively scarce OT&E experience is emphasized and DT&E experience is de-emphasized.

The key to the identification of the right type, numbers and mix of personnel is the statement and analysis of the testing goals or objectives. If the objectives stated in the concept and plan are conflicting or incomplete, the organization will become a conglomeration of incompatible, separate enterprises, each following its own aim or interest. Any doubt concerning the clarity or consistency of the objectives should be removed by reconvening the test feasibility working groups and rectifying these deficiencies. Well formulated objectives encourage synergistic planning in all of the creative parts of the organization and result in unity of action that cannot be achieved in any other way.

A feasibility study is not a plan although many such studies contain the elements of a good plan. Adequate manpower and organization development depends heavily on the plan, therefore, before attampting to man the test staff, the plan should be examined in terms of the following general characteristics of a good plan:

Based on objective thinking and free of undue emphasis on peripheral objectives with logical development of the basis of action.

Accurate forecast of the nature and requirements of test events and their sequencing.

Flexibility for adjustment of contingencies without serious loss of economy or effectiveness.

Comprehensiveness to cover all actions that will be required of participants and test force organizational elements for satisfactory accomplishment of the objectives.

Simple and unambiguous so that it can be carried out readily without confusion.

Economical in its use of assigned test resources.

Considers and assesses alternate methods for attaining objectives by identifying the most important factors and relating tangible factors separately.

Each member of the test staff will be a product of his experiences and have some biases. There is no sure way to avoid this. Appropriate attention must be given to insure that the end value of any data produced will be enhanced by knowledge of those involved, in addition to how these data were collected and evaluated. Several criteria for staff selection suggest themselves:

Assure that the test planning group contains members who are expert in equipment operation and tactics in combat, operational procedures, the particular equipment to be tested, statistical and mathematical test design, methods of simulating combat in test situations, and methods of test data collection and test data analysis.

Assure that no constraints are placed upon the test team which would prevent any members from making the fullest use of their expertise in their respective fields.

Policy statements require that all operational testing be accomplished by the type of personnel at experience levels planned for the materiel items end use. However, blind adherence to this requirement can seriously impact the cost/schedule/performance of the test. This may dictate that some Contractor or Development Agency test and evaluation specialists be added to achieve an effective test team complement. Identification of these added T&E specialists and characterization of their role in planning and execution of the test is offered as a way to remove criticism based on fear of developer bias and concern for truly independent evaluation.

Experts in operation and tactics should come from personnel who have actual combat experience. They should also include officers, non-commissioned officers, and enlisted personnel who have had actual combat operational support experience. For a team such as described, rank and specialized expertise are not necessarily in a straight-line relationship. For example, if in the course of planning, an officer who is expert on tactics supersedes the views of an enlisted maintenance man about the difficulties of repairing the system, or inhibits free expression of views, valuable data may be lost or test problems may intensify.

Items of Equipment. Material for test and test support equipment/facilities (not including brick and mortar) to be used or expended during the test must be identified in the planning phase. Listing and categorization in terms of source, condition, configuration status, schedule, support requirements, modifications, documentation, training operator characteristics, etc., should be addressed.

After all items of equipment to be employed in the conduct of the test have been identified they must be categorized as unique to the test and other. The categorization as "test unique" is necessary because in joint OT&E, funds for unique items are provided by the DD(T&E) from Program Element #65804D.

The Air Force Manual, AFM 173-10, and the equivalent Army and Navy publications, should be used to the maximum extent practical in estimating O&M and other costs. It will undoubtedly be found that AFM 173-10 is more useful in establishing cost in the O&M category. For equipment identified as unique, test item modifications, and equipment not in operational inventory, it will be necessary to seek implementing and supporting command assistance.

Operations and Maintenance. The participating DOD Components in Joint Tests are required to provide the necessary O&M support without reimbursement (from 65804D). This means that the participating services will fund for all other requirements not specifically accepted for funding under the DD(T&E) determination of test unique costs. Foreseeable situations exist wherein the test can involve test ranges or projects which are funded from the RDT&E appropriations category or are industrially funded facilities which require reimbursement by projects using the test facilities. The solution to this problem is clear understanding at the outset. This, in turn, depends on accurate identification and estimation of test needs, schedules and costs. Careful consideration of O&M resource matters including responsibilities for the planning, programming and budgeting (PPBS) for them is essential to good test management. This should also include the identification of an agency or agencies responsible to monitor the PPBS of all participating elements.

The support provided to joint test programs by operating commands must be recognized as secondary to the mission of the Command. Diversion of operational systems and expenditure of related supporting resources can only be viewed as a reduction in the operational readiness of the unit. Therefore, the test organization and plan must include provision for contingencies and flexibility to capitalize on opportunities as they arise. For example, consider a change in DEFCON status.

Military Construction. Situations can arise in the design of joint tests where, in spite of all attempts to eliminate needs for high dollar value real property items, they just cannot be avoided. When this occurs the Joint Test Director will be immediately confronted with schedule incompatibilities and significantly increased potential program slippages. Beyond adequate planning for lead time and MILCON unique budgeting considerations similar to PPBS relationships previously discussed, there are some additional approaches to these problems which can be identified and discussed.

Certain instrumentation and Communications/Electronic items can be specified and justified for procurement on a "turn key" basis. Although this has been an abused practice and has created rigorous standards for approval and use, it remains as a viable approach when it is properly undertaken.

Another approach is to incorporate Combat Engineering, Air Force "Prime Beef" or SEABEE type organization into the plan. Justifications for this can be most readily seen in Joint Operational Test and Evaluations. The point is, without lengthy belaboring, that there are many innovative ways to circumvent MILCON type problems if the services of a professional Civil Engineering Officer are included in the organizational phase. Details of the MILCON program procedures are contained in AFM 85-26.

7. THE TEST PLAN AND TEST REPORT

Recall that the actual planning of the test, test designing, establishing measures of effectiveness, data collection plans, range support requirement documents, test methods, execution of the test, data analysis, evaluation and preparation of the Test Report — all aimed at meeting the originally directed test objectives — are, for all practical purposes, the same in JOT&E as with any OT&E. All of these matters are treated elsewhere in this manual. What we have been discussing in this chapter are those things unique to JOT&E.

One disgression from this is necessary in the form of a short comment about the Test Plan and the Test Report.

The Test Plan and Test Report formats given in this manual are for USAF standard use. They may not be the format that the Joint Test Director will decide to use; however, they should be considered. The JOT&E plan and report will also have lengthy approval chains. They will have to go through each participating Department as well as the Office of the DD (T&E). This approval chain should be established at the outset to avoid later surprises. The approvals should then be coordinated with all in the chain — this step might even cause it to get shortened by some wise and sympathetic senior official.

8. THE POLITICS OF JOT&E

The politics of a JOT&E can become very hazardous. Remember that JOT&E is often pitting the materiel, personnel, and tactics of one military department against another. The "loser" could easily have one of his favorite weapons shown to be obsolete. Recall what Bill Mitchell did to the battleship and what the waving of white bed sheets did to the horse mounted calvary.

Those involved in JOT&E "like Ceasar's wife must be above suspicion." Scrupulous attention to one of the best aspects of all operational testing, test realism, will go a long way to avoiding becoming meshed in debilitating and even career ruining situations.

9. ACHIEVING REALISM IN TESTING

Experience indicates that realism has a positive effect on the validity, credibility and utility of an OT&E program. Major characteristics of realistic tests for establishing the feasibility of test concepts and subsequent detailed test planning are:

Operational and support personnel having relatable combat experience should participate in the selection of test objectives and development of the test design. They should also be participants in data collection and evaluation (White Force) functions.

System operators should not have test data collection responsibilities; this function should be the responsibility of trained White Force personnel.

Real-time casualty assessment and damage assessment should be used.

Instrumentation should be provided which permits the operators and the threat to exercise (by simulation) the kinds of options available to them in combat.

The program should consist of exercises simulating combat or combat support situations in which collection of test data is superimposed. These may be supported by experiments and other measurements not involving exercises.

The exercises should be optimized for simulations of two-sided engagements.

Operators of systems being evaluated should be subjected to combat zone conditions and status during the test.

Crews should stay at action stations for times representative of the military situations being simulated, including bad weather and night.

The system operators and interface system operators should be put under stress as in combat, using isolation, rumors, noise, odors, smoke, pyrotechnics, blank charges, time constraint, foreign languages, and other psychological devices and procedures to a practical degree.

System operators should be typical operational personnel with varying levels of skill (including for system maintenance) and should use typical operational procedures.

Operational constraints and opportunities typical of the situation and locale should be imposed.

Tactical profiles and combat delivery modes should be used.

Testing should be performed in a variety of operationally representative physical environments, including climate and season, weather, terrain, vegetation, visibility and EW characteristics.

Scenarios should provide planned and unplanned conditions in which crews will react to "enemy" actions, and system problems.

Secrecy should be maintained about the "enemy's" moves and strength except for that normally gained from intelligence in the specific situations.

Threats should be simulated according to the best current intelligence about enemy doctrine, deployments, tactics, appearance, signals and actions.

Tactics should be consistent with the selected military situation.

10. SUMMARY

In summary it can be seen that JOT&E differs very little from OT&E as a single military department would approach it. The major differences are at the front end (where the incentive comes from); in the organization (two or more departments, each with its peculiarities, are involved); source of funds (test unique costs paid by OSD); politics of planning (each participating Department will want to look as good as possible); and at the back end (more people/organizations to satisfy). These are formidable differences. Significant dangers await those assigned to JOT&E roles who do not recognize and work hard to cope with these differences.

CHECKLIST FOR JOINT OPERATIONAL TEST & EVALUATION

1. Review the report of the Joint Test Feasibility Working Group
   a. Are the objectives of the test program clearly stated?
   b. Are intelligence estimates current and compatible with those of operating command?
   c. Are the environmental factors realistic?
   d. What test design has been accomplished?
2. Have all participants been identified?
3. What financial planning, programming, and budgeting actions have been initiated at what levels?
4. What manpower will be required? What skills and backgrounds?
5. What provisions have been made for test items, and support equipment?
6. What range will be used? Have Statements of Capability been issued?
7. Are schedules compatible with availability of test items?
8. Are sample sizes compatible with the availability of hardware, range schedule, and time available?
9. Is full use being made of simulations to reduce physical testing?
10. Is an environmental impact statement required?
11. Are "Unique" items of equipment defined for funding by DOD?
12. Has the approval chain been defined and reduced as much as possible?
13. Has the "charter" been drawn up and approved by DD (T&E)?
14. Are memorandum agreements or Letters of Instructions with participating DOD components concluded?
15. Have the formats for the Test Plan and Test Report been defined?

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER RADC-TR-74-270, Volume I (of two)	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) STANDARD PROCEDURES FOR AIR FORCE OPERATIONAL TEST AND EVALUATION		5. TYPE OF REPORT & PERIOD COVERED Final Report
		6. PERFORMING ORG. REPORT NUMBER None
7. AUTHOR(s) D.E. Simon (RCA) et al		8. CONTRACT OR GRANT NUMBER(s) F30602-73-C-0375
9. PERFORMING ORGANIZATION NAME AND ADDRESS Braddock, Dunn & McDonald, Incorporated* 5301 Central Avenue Albuquerque, New Mexico 87108		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS PE 65804D JO 51050102
11. CONTROLLING OFFICE NAME AND ADDRESS Rome Air Development Center (IRAA) Griffiss Air Force Base, New York 13441		12. REPORT DATE October 1974
		13. NUMBER OF PAGES 450
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Same		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

Distribution limited to US Gov't agencies only; test and evaluation; October 74. Other requests for this document must be referred to RADC(IRAA),GAFB,NY 13441.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

*with subcontractors RCA Government and Commercial Systems Missile and Surface Radar Division and the Xerox Corporation Data System Division.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

OT&E
Initial OT&E
Weapon System Life Cycle
Measures of Effectiveness OT Planning
Joint OT&E Simulations
OT&E Data Handling
AF Testing
OT&E Mathematics and Statistics
OT&E Statistical Design
Constant Improvement Program
Test Report Formats

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes the overall structure of Air Force Operational Test and Evaluation process from the appearance of the test directive to the production of the final report. It describes the major documentation requirements applicable to major and minor weapon system acquisitions and provides guidance to the Air Force OT&E community in the areas of formulation of test objectives, selection of test concepts, determination of test planning criteria, OT&E data collection and analysis requirements, formulation of OT&E conclusions and

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (continued)

recommendations, and test reporting. It further describes procedures for the development of statistical design of an operational test. This report culminates an effort to standardize the management and analytical procedures applicable to Air Force Operational Test and Evaluation. Toward this end, standardized data elements of measures of effectiveness are developed.

MISSION
of
Rome Air Development Center

RADC is the principal AFSC organization charged with planning and executing the USAF exploratory and advanced development programs for electromagnetic intelligence techniques, reliability and compatibility techniques for electronic systems, electromagnetic transmission and reception, ground based surveillance, ground communications, information displays and information processing. This Center provides technical or management assistance in support of studies, analyses, development planning activities, acquisition, test, evaluation, modification, and operation of aerospace systems and related equipment.

Source AFSCR 23-50, 11 May 70