DEVELOPING A TEMPLATE FOR
LOGISTICS TEST AND EVALUATION

THESIS

Timothy J. Fennell, Captain, USAF
Jason S. Mogle, Captain, USAF

AFIT/GLM/LAL/94S-17

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GLM/LAL/94S-17

DEVELOPING A TEMPLATE FOR LOGISTICS TEST AND EVALUATION

THESIS

Timothy J. Fennell, Captain, USAF
Jason S. Mogle, Captain, USAF

AFIT/GLM/LAL/94S-17

The views expressed in this thesis are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

AFIT/GLM/LAL/94S-17

DEVELOPING A TEMPLATE FOR LOGISTICS TEST AND EVALUATION

THESIS

Presented to the Faculty of the Graduate School of

Logistics and Acquisition Management of the Air Force

Institute of Technology

Air Education and Training Command

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Logistics Management

Timothy J. Fennell, B.S.

Captain, USAF

Jason S. Mogle, B.A.

Captain, USAF

September 1994

Approved for public release; distribution unlimited

Acknowledgments

We would like to express our appreciation to several individuals for helping us with this thesis. Our sponsor, Major Mary Parker, was an invaluable source of information and assistance throughout the course of our research. Also, we wish to show our appreciation to our thesis advisors, Major Marsha Kwolek and Captain Nathan Jones, whose guidance, expertise, and understanding enhanced our professional development.

We also wish to thank Melissa for her patience and support, without which this thesis could not have been accomplished.

Timothy J. Fennell

Jason S. Mogle

Table of Contents

List of Tables

Table | Page

AFIT/GLM/LAL/94S-17

Abstract

Logistics test and evaluation, performed during the developmental test and evaluation effort, is an integral part of the aircraft acquisition process. However, there has been no standard approach to conduct logistics test and evaluation. This study researched past and present approaches to aircraft logistics test and evaluation to determine the most effective method for future programs.

We conducted this study by using the Delphi method. We solicited the expert opinions of 32 individuals from the logistics test and evaluation field using two rounds of questions. We then statistically analyzed the data to validate the need to test the ten ILS elements, to determine if the logistics test and evaluation process is worth the resources allocated to it, and to determine the most effective approach to aircraft logistics test and evaluation.

Based on this analysis, we concluded that the ten ILS elements are indeed a valid baseline for a logistics test and evaluation template, and that logistics test and evaluation is worth the resources allocated to it. Finally, we concluded that the F-22 logistics test and evaluation approach is the most effective. Collectively, we used this information to develop a template for conducting logistics test and evaluation for future aircraft acquisition programs.

DEVELOPING A TEMPLATE FOR LOGISTICS TEST AND EVALUATION

Chapter I: Introduction

General Issue

This thesis concerns the subject of logistics test and evaluation, which we will refer to as *logistics test*. Logistics test is a test methodology for evaluating and analyzing the ten Integrated Logistic Support (ILS) elements (maintenance planning, supply support, facilities, technical data, support equipment, manpower and personnel, computer resources support, training and training support, packaging/handling/storage/transportation, and design interface) as they apply to the article under test (21:2). Most importantly, this methodology means getting the end users (maintenance personnel) involved in the acquisition of weapon systems--something which unfortunately has not been done in the past. As the technicians perform maintenance on the aircraft during test and evaluation, they evaluate the practicality of the design for supportability (the degree to which the system can be supported) and maintainability (the ease, accuracy, safety, and economy in the performance of maintenance actions) (3:15,20). Decision makers listen to the evaluation and suggestions so the improvements can be incorporated into the design before the weapon system goes into full production.

Logistics test is currently being done on the C-17 aircraft program. The Air Force Flight Test Center at Edwards Air Force Base has conducted logistics tests for the past several years. However, logistics test has not been widely recognized due to a lack of formal structure, documentation, and publicity. The 412th Test Wing at Edwards Air Force Base is sponsoring this thesis as a step toward formalizing and standardizing the logistics test process.

Specific Problem Statement

Because of the lack of a formal structure for logistics test, various acquisition programs have spent large amounts of money, time, and resources just to develop a methodology or procedure to test and evaluate their specific logistics concerns. Essentially, each acquisition program has been forced to invent its own logistics test plans. To alleviate this problem, the purpose of this thesis is to develop a comprehensive template for the logistics test requirements of any aircraft acquisition program. This template will serve as a formal guide for acquisition and test personnel to use for the development of a specific logistics test plan for their particular program. Future acquisition programs will benefit from such a template by having an outline for their logistics test plan already available.

Investigative Questions

In order to properly address the specific problem, the following investigative questions will need to be answered:

1. What are the current regulatory requirements to perform logistics test?

2. What is currently being used as a basis to plan logistics tests?

3. What has been used in the past as a basis to plan logistics tests?

4. What criteria must be considered to appropriately evaluate logistics test items?

After these four questions have been answered, we will be in a position to create a logistics test template.

Scope

This research will apply primarily to aircraft acquisition programs, although the final template may be used by other types of programs if deemed appropriate by the program personnel. The reason for this limitation is that logistical requirements (and the testing of these requirements) can be very different for different types of systems. For example, the logistical requirements of a satellite are much different from the logistical requirements of an aircraft. Therefore, making a template that applies to both types of systems is not practical.

Definitions

To help the reader fully understand this thesis, some terminology needs to be explained. The following terms are found throughout the thesis.

Template: A systematic guide of testing and evaluating the ten ILS elements as they pertain to the article being tested.

Integrated Logistics Support (ILS):

> ILS is a disciplined, unified, iterative approach to the management and technical activities necessary to:
>
> 1. Develop support requirements that are related consistently to readiness objectives, to design, and to each other.
>
> 2. Effectively integrate support considerations into the system and equipment design.
>
> 3. Identify the most cost-effective approach to supporting the system when fielded.
>
> 4. Ensure that the required support structure elements are developed and acquired(9:1-2).

ILS Elements:

1. Maintenance Planning-The process conducted to evolve and establish maintenance concepts and requirements for the lifetime of the system.

2. Manpower and Personnel-The identification and acquisition of military and civilian personnel with the skills and grades required to operate and support the system over its lifetime at peacetime and wartime rates.

3. Supply Support-All management actions, procedures, and techniques used to determine requirements to acquire, catalog, receive, store, transfer, issue, and dispose of secondary items. This includes provisioning for both initial support and replenishment supply support. It includes the acquisition of logistics support for support and test equipment.

4. Support Equipment-All equipment (mobile or fixed) required to support the operation and maintenance of the system. This includes associated multi-use end items, ground handling and maintenance equipment, tools, metrology and calibration equipment, test equipment, and automatic test equipment.

5. Technical Data-Scientific or technical information recorded in any form or medium (such as manuals and drawings). Computer programs and related software are not technical data; documentation of computer programs and related software are. Also excluded are financial data or other information related to contract administration.

6. Training and Training Support-The processes, procedures, techniques, training devices, and equipment used to train civilian and active duty and reserve. This includes individual and crew training (both initial and continuation); new equipment training; initial, formal, and on-the-job training; and logistics support planning for training equipment and training device acquisitions and installations.

7. Computer Resources Support-The facilities, hardware, system software, software development, and support tools, documentation, and people needed to operate and support embedded computer systems.

8. Facilities-The permanent, semipermanent, or temporary real property assets required to support the system, including conducting studies to design facilities or facility improvements, locations, space needs, utilities, environment requirements, real estate requirements, and equipment.

9. Packaging, Handling, Storage, and Transportation-The resources, processes, procedures, design considerations, and methods to ensure that all system, equipment, and support items are preserved, packaged, handled, and transported properly, including environmental considerations, equipment preservation requirements for short and long term storage, and transportability.

10. Design Interface-The relationship of logistics related design parameters to readiness and support resource requirements. These logistics related design parameters are expressed in operational terms

rather than as inherent values and specifically relate to system readiness objectives and support costs of the system. (9:1-2).

Overview of the Thesis

The rest of this thesis will be directed toward answering our investigative questions and creating the logistics test template. In Chapter II, we will review the available literature concerning logistics test to lay a foundation for the thesis. This literature review will provide the necessary knowledge required to appropriately address the subject of logistics test in general. Chapter III will then explain the methodology we will use to answer our four investigative questions and to address our specific problem. Next, in Chapter IV, we will present the data obtained from implementing our methodology and make conclusions after a thorough analysis of the information. These conclusions will result in creating a template for logistics test requirements of aircraft acquisition programs. We will then conclude with our recommendations and suggestions for further research in Chapter V.

Chapter II: Literature Review

Introduction

This literature review traces the background of the need for logistics test and addresses current issues involving how logistics test should be accomplished. However, because a formal logistics test procedure is a relatively new concept, literature concerning the specific subject of logistics test is scarce. We will review DOD instructions, pamphlets, textbooks, and personal correspondence with experts in an attempt to discover why logistics test is necessary, to determine what requirements (if any) govern logistics test, and to learn when logistics test should be accomplished. We will also review past and present logistics test plans and procedures of various aircraft acquisition programs to discover what has been used in the past and what is now being used to conduct logistics test. This review will enable us to establish a baseline for conducting our research. With this knowledge, we should be well positioned to create the template for logistics test and evaluation.

Background

Why does the Air Force need a procedure to perform logistics test? If we look at the life-cycle cost (LCC) of a typical aircraft weapon system, we find that 60% of the system's total LCC is spent on operation and logistical support. This cost is much higher than what is typically spent on system research and development (10%) and production (30%) (17:23). Even though it is accomplished relatively inexpensively as part of research and development, logistics test has a significant influence on the lifetime operation and support costs of the system.

A well-known example is the notorious design of the F-4 radio system. This unreliable radio was located under an ejection seat. Every time this radio failed (which was often), the entire ejection seat had to be removed before the radio was serviced. This required hours of dangerous work (removing and reinstalling a rocket), instead of just a few minutes for the relatively simple replacement of the radio. Other logistical problems of the F-4 contributed to an unimpressive rate of 4.6 sorties between major maintenance (21:11). This unfortunate situation existed because there was not a legitimate procedure to logistically test the weapon system. As a result of logistical inefficiencies in our weapon systems, we now have precedence in the Air Force basic doctrine for emphasizing logistics support:

> The best aerospace systems are worthless if they cannot be refueled, rearmed, and otherwise kept in commission. Aircraft grounded for lack of parts or consumables represent, for the period they are out of service, as much loss to combat capability as aircraft destroyed by enemy action. Aircraft that require excessive maintenance or excessive time to accomplish routine maintenance add to the attrition toll.
> The most obvious solution to the maintenance challenge is to give logistical requirements high priority in designing aerospace weapon systems. Systems designed for easy maintenance and greater reliability decrease the logistics problem. A further benefit is that as the

> general logistics problem decreases, the amount of effort required for logistics decreases, thus easing "tooth-to-tail" ratio problems. (12:255)

Now that we know logistical requirements are critically important to the development of aircraft weapon systems, how do we make sure our plans to test these requirements are efficient, or even legitimate? There has never been a standardized approach to logistics test. In fact, the relatively new B-2 program was the first to employ a systematic method of logistics test, and by all measures, the concept was successful. The C-17 program later refined the logistics test process that the B-2 used. Finally, the F-22 program became the first aircraft acquisition program to incorporate a formal, structured logistics test procedure as early as the demonstration/validation phase of the program (19:15). But each of these programs had to develop its own basic procedure for its logistics test strategy. Consequently, a standardized template for logistics test would enable future aircraft acquisition programs to save money and streamline the logistics test process.

Requirements

In order to begin constructing such a template, we must first discover what requirements exist that might dictate or guide our efforts. General guidance is given as:

> Verifying the accuracy and adequacy of the logistics support identified begins early in the process. Testing, evaluating, and correcting deficiencies in both the design and the support system continue throughout the life cycle. The validity of the analysis results and

> attendant data products must be successfully demonstrated within stated confidence levels. Results of formal test and evaluation programs are analyzed and corrective actions implemented as necessary. (10:4)

But what should we test and evaluate, and how do we accomplish the testing procedure? Before we know exactly what to test, we must be familiar with the user's requirements, know where the system will eventually operate, and know what resources the user will have available when using the system. We must also "manage the contributions to system reliability and maintainability that are achieved by hardware, software, and human elements of the system" (9:6-C-2). Furthermore, we are told that by Milestone 1 (Concept Demonstration Approval) an initial Integrated Logistics Support Plan (ILSP) will be drafted and will include development schedules for each ILS element. Also, by Milestone 2, (Development Approval), the test and evaluation plans should be developed to quantitatively assess achievement of support related thresholds; and the Logistics Support Analysis (LSA) program will be started as the database for ILS documentation (9:7-A-2-1).

The LSA program consists of the set of analyses conducted to examine all elements of the system to determine logistical requirements and to influence the design of the system to ensure that adequate support can be provided at an affordable cost. LSA task 501, Supportability Test, Evaluation and Verification, identifies test resources, procedures, and schedules that will be included in the Test and Evaluation Master Plan (TEMP) (10:25). However, we are not informed on exactly how to conduct the tests that are outlined in the TEMP because the TEMP is an overview document.

Based on the above discussion, the requirements for logistics test seem to be vague and general. This vagueness can be advantageous to the acquisition program, as we have the flexibility to tailor the logistics test approach to suit our specific needs. However, we will not know if we have covered all of the test requirements thoroughly and appropriately. Therefore, it would also be advantageous to have a logistics test template to which we can always refer during the process, saving man-hours, as we now would have a strategy available to use whenever we need it. Because "the ILS effort will encompass the ten ILS elements" and "each of these ten elements must be addressed for both hardware and software" (9:7-A-1-1) during the acquisition process, the cornerstone of such a template will logically be a way of ensuring the ten ILS elements will be evaluated in some way. But when will this evaluation take place?

Considering that the bulk of LCC is committed early in the acquisition process, "it becomes apparent that logistics planning must begin at the front end of an acquisition program" (2:179). Therefore, we should perform logistics test as early as possible during developmental test and evaluation (DT&E). Also, in most programs, private contractors play the key role in producing the required logistical items. So "tests that determine contractual compliance will be conducted independent of the contractor or under program office supervision" and, by Milestone 2, "items will have met...their allocated reliability and maintainability goals..." (9:6-C-1-2).

The "test" of DT&E is a technically driven activity to quantitatively gather data which tells us if the test article meets specified design criteria. The "evaluation" of DT&E is a formal assessment process performed by supporting test organizations, users, and acquisition agencies, possibly with contractor involvement (15:4.4-5). Specifically, DT&E programs shall:

1. Identify potential operational and technological limitations of the alternative concepts and design options being pursued,

2. Support the identification of cost-performance trade-offs. (It should be noted here that "performance" includes both operational characteristics and support characteristics),

3. Support the identification and description of design risks,

4. Substantiate that contractual technical performance and manufacturing process requirements have been achieved, and

5. Support the decision to certify the system is ready for operational test and evaluation. (9:8-4)

This guidance tells us what we need to have as a result of our logistics test efforts during DT&E. However, we are still left with the problem of not having a standardized template for <u>conducting</u> logistics test.

Past Logistics Test Methods

Logistics testing during previous aircraft acquisition programs relied heavily on the particular contractor (or contractors) responsible for the item being tested. Although the Air Force would typically charter a skeleton crew of personnel to oversee the logistics concerns of the system being tested, the vast majority of logistics testing was performed by and supervised by the contractor, with the government merely serving in an observer capacity. Also, detailed logistics test plans were typically not made by the government, and, to a certain extent, logistics concerns were tested "by default" just by flying the aircraft. If a particular logistics-related problem surfaced as a result of operational testing, then that problem was addressed. If no logistics-related problems surfaced, then the logistical design of the system was considered to be sound. Although this approach was relatively simple in the short run, it often resulted in excessive amounts of time-compliance technical orders (TCTOs) to correct logistics problems on production aircraft that were not adequately tested during the DT&E process (23). The following programs (B-1 and F-16) were among the earliest aircraft programs to at least recognize the need for some plan concerning logistics test and evaluation.

B-1 Program. The B-1 program experienced concurrent testing and production. Nevertheless, a logistics plan was developed to guide the logistics testing efforts. Logistics testing and evaluation consisted of eight major objectives (summarized below) to be accomplished during the testing process (11:B-1, B-2).

1. Reliability, Maintainability, and Availability: Evaluate mission reliability, logistics reliability, and support equipment reliability. Evaluate quantitative and qualitative maintainability parameters to include maintenance man-hours per flying hour (MMH/FH), mean time to repair (MTTR), accessibility, and ease of maintenance.

2. Technical Data: Schedule the aircraft to perform verification and validation work.

3. Training: Evaluate contractor and Air Force developed training courses to ensure they meet objectives. Place special emphasis on areas of new technology.

4. Support Equipment: Functional tests to assess performance and interfaces with the test item or system, and maintainability evaluation demonstrations.

5. Supply Support: Use data (reliability, maintainability, and actual supply data collected at the test site) from the flight test program to verify spares calculations.

6. Maintenance Planning: Evaluate the adequacy and effectiveness of the maintenance planning effort to ensure the logistics elements were adequate to maintain the weapon system and provide the level of system operational effectiveness and availability required.

7. Software: Assess software reliability and maintainability.

8. Integrated Diagnostics: Evaluate the aircraft self-diagnostic capabilities.

F-16 Program. This program also did not use a detailed logistics test plan nor did it use the Logistic Support Analysis process during the original acquisition and testing process. The testing effort was almost completely dependent upon contractor testing and data obtained by isolated tests performed by the contractor (16). The F-16 ILSP contained little useful information on logistics testing, but it provided two objectives. The first objective was to "ensure support equipment compatibility, training adequacy, technical order adequacy and system maintainability" (13:28). The second was to "identify hardware/firmware deficiencies (reliability)" (13:28). Also, logistics testing focused on reliability and maintainability aspects of the system as a whole, and not on each logistics element, and distributed the reliability and maintainability data to all participating agencies for their review. However, sometimes maintainability could not be quantitatively evaluated due to the contractor testing concept impacting the validity of the data and due to the limited amount of data collected by Air Force personnel (4:19).

These programs are representative of the early attempts to legitimize, or formalize, the logistics testing process. Although they outline major objectives to accomplish during the testing process, they still do not sufficiently outline a detailed method to achieve these objectives and they still rely almost completely on contractor furnished data to evaluate the logistical aspects of the system. The next section shows how current programs are taking logistics test to the next level of legitimacy.

Current Logistics Test Methods

The three aircraft acquisition programs that are currently conducting logistics testing as a part of DT&E are the B-2, C-17, and F-22 programs. Logistics testing for these programs is more formalized and structured than the attempts made in the past. Although the approach used by each program is somewhat different, the one underlying common trait among these logistic test programs is that they all seek to involve Air Force personnel as an integral part of the testing process, rather than merely being observers who analyze contractor data. They also outline in detail how logistics test objectives will be met. The following summaries are provided as general overviews of these logistic test methods.

B-2 Program. This program uses an integrated team concept for logistics test. The DT&E personnel at the Air Force Flight Test Center (AFFTC) team with the contractor and Air Force Operational Test and Evaluation Center (AFOTEC) to develop an essential task listing (ETL) of major items or procedures that need to be tested, and these tasks are prioritized according to urgency of need. These ETLs are then broken down into Logistic Test Work Orders (LTWOs) that describe in detail how an item is to be tested. These LTWOs are categorized into one of three categories:

1. Simple tasks that can be conducted anytime.
2. More complicated "middle ground" tests that need a small level of planning.
3. Complicated tasks that require a detailed plan and close scheduling (18).

Once the tasks are categorized, logistics testing is accomplished as opportunities arise during flight testing. A goal of the B-2 logistics test program is to cause minimal disruption to the flight test program. Therefore, aircraft downtime dedicated specifically to logistics test is very limited. As testing progresses, the LTWOs and ETLs are "checked-off" as they are tested and reliability and maintainability data are collected as needed (18).

This approach focuses on the tasks and subsystems that need to be tested and is based largely on technical data and support equipment deliverables. As tests are conducted, they are linked to an ILS element (or several) to ensure that all of the ILS elements are evaluated. Also, an important aspect of this approach is the need for AFOTEC to maintain "independence" by keeping their own databases, performing separate analyses of the data, and generating separate reports. The advantage of including AFOTEC (and other organizations) in the DT&E process is that larger amounts of data can be gathered and shared, resulting in larger sample sizes for a more thorough analyses (18).

C-17 Program. The C-17 logistics test plan is based on a systematic evaluation of each ILS element, and the test concept is derived from AFR 800-8, Acquisition Management. This concept consists of the quantitative and qualitative aspects of the ten ILS elements as they relate to three disciplines. The first discipline is reliability and maintainability (R&M), and it focuses mainly on quantitative analyses. The second discipline, human factors (HF), is a qualitative analysis of how people "receive information through their senses, store this information, and process it in making decisions" (24:7). The third discipline is called logistics test (LT) and it includes both quantitative and qualitative measurements. This discipline focuses on "assessing the adequacy of C-17 aircraft system performance and system supportability through direct man-machine interface" (24:10). These disciplines are evaluated for each ILS element and are broken down into measurable units called logistics test measures (LTMs) within each ILS element, resulting in the following matrix (24:16-20).

Elements

1. Maintenance Planning
   - R&M: LTM 10: Reliability
     - 11: Maintainability
   - HF: None
   - LT: LTM 12: Scope
     - 13: Frequency
     - 14: Task Completion Times

2. Manpower & Personnel
   - R&M: LTM 20: Reliability
     - 21: Maintainability
     - 22: Average Crew Size
     - 23: AFSC
   - HF: LTM 24: Human Performance during All-Weather Activities
   - LT: LTM 25: Crew Size
     - 26: AFSC

3. Support Equipment
   - R&M: None
   - HF: LTM 30: Ease of Use
     - 31: Handling
     - 32: Safety
     - 33: Compatibility
   - LT: LTM 34: Supportability
     - 35: Utilization Rate

4. Supply Support
   - R&M: LTM 40: Reliability
   - HF: None
   - LT: LTM 43: SMR Codes
     - 44: Availability

5. Tech Data
   - R&M: None
   - HF: LTM 50: Safety
     - 51: Adequacy
     - 52: Clarity of Instructions
   - LT: LTM 53: Understandability
     - 54: Ease of Use

6. Training and Training Support
   - R&M: None
   - HF: None
   - LT: LTM 60: Knowledge Training (Contractor Furnished)
     - 61: Knowledge Training (AF Furnished)
     - 62: Proficiency Training
     - 63: Safety

7. Computer Resources Support
   - R&M: LTM 70: CRS Reliability
     - 71: CRS Maintainability

72: BIT System Adequacy
HF: LTM 73: Functional Utility
74: Ease of Use
LT: LTM 75: Functional Utility
76: Ease of Use

8. Facilities
R&M: None
HF: None
LT: LTM 80: Supportability
81: Safety
82: Compatibility

9. PHS&T
R&M: None
HF: None
LT: LTM 90: Suitability
91: Safety

10. Design Interface
R&M: LTM 100: Reliability
101: Maintainability
HF: LTM 102: Interoperability
103: Accessibility
LT: LTM 104: Energy Consumption
105: Accessibility
106: System/Component Preservation
107: Component Standardization

F-22 Program. This approach uses a large team concept consisting of DT&E, AFFTC, AFMC, AFOTEC, AETC, and contractor personnel, with each organization analyzing their own data writing separate reports as needed. The plan defines the objectives of logistics test, identifies the ILS elements to test during DT&E (and what elements not to test), describes the interface between organizations, and outlines the entire process. This program seeks to give logistics testing the same priority as flight testing, and aircraft downtime is scheduled specifically for logistics testing purposes. The process begins by identifying the ILS elements to be evaluated and the major subsystems to be tested (1:1-1,3-8). Then a test information sheet (TIS) is developed for each logistics test task that is required to be tested as a part of the subsystems. Each subsystem can have several, perhaps hundreds, of TISs. These TISs specifically explain how an item is to be tested and evaluated, and how the item relates to the ILS elements. The goal is to integrate reliability and maintainability analyses, technical order verification, support equipment verification, and human factors assessments into an all-encompassing, overlapping testing effort, rather than evaluate each of these aspects in isolation as was done in past logistics test efforts (23). This process generates a periodic three-part report. Part one gives the overall status of logistics test resources by ILS element and potential impacts on the flight test program. The second part gives an assessment of all ILS elements of the system, with information such as how many technical orders have been validated and what pieces of support equipment have been accepted. The third part of the report provides for comments and feedback on the testing process (1:3-10,11).

Comparison of the Methods

Now that we have seen various methods of conducting logistics test, it will be helpful to compare these approaches to see where the similarities and differences are. This comparison is found in Table 1.

Table 1

Comparison of the Logistics Test Methods

FEATURE	TRADITIONAL METHOD	F-16	B-1	C-17	B-2	F-22
Scheduled Log Test	NO	YES	YES	YES	NO	YES
Integrated	NO	NO	NO	YES	YES	YES
Focus on ILS Elements	NO	NO	YES	YES	NO	YES
Addresses All 10 ILS Elements	NO	NO	NO	YES	YES	YES
Test in Isolation	NO	YES	NO	NO	NO	NO
System Emphasis	YES	YES	YES	YES	YES	YES
Subsystem Emphasis	NO	NO	YES	YES	YES	YES
AFOTEC Involvement	NO	NO	NO	NO	YES	YES

As Table 1 suggests, there are several similarities and differences among the different approaches. For example, the Traditional method and the B-2 method are the only two approaches which do not use a specific schedule for logistics test. The next feature, "Integrated," refers to whether the logistics test effort is performed by using a team concept. As we can see in Table 1, only the current programs stress an integrated team approach. The feature of "Focus on ILS Elements" refers to whether the methods use the ILS elements to drive the logistics test procedure. The B-1, C-17, and F-22 programs have this feature. The next feature, "Addresses All 10 ILS Elements," is similar to the previous feature. Since DODI 5000.2 now mandates that all 10 elements be addressed, it is not surprising to find that the current programs comply with this instruction. The B-2 method is unique in that it does not focus its testing process on the ILS elements, but it does ensure that all ten elements are addressed. The feature of "Test in Isolation" refers to the concept of testing one part of the system completely and then moving on to another part of the system. The F-16 method is the only method which tests in isolation. The features of "System Emphasis" and "Subsystem Emphasis" are self explanatory. Although all of the methods use an emphasis on the system, only the B-1, C-17, B-2, and F-22 methods emphasize testing of subsystems. Finally, we see that only the B-2 and F-22 have meaningful AFOTEC involvement in the logistics test process during DT&E. We will use a panel of experts (described in Chapter III) to distinguish between the approaches and to find which features are desirable for use in a standardized logistics test approach.

Conclusion

This literature review serves to provide background information with the intent to help develop a standardized template for the logistics test requirements of any aircraft acquisition program. Proper logistics test is required to reduce total LCC. Even though logistics test is currently being conducted on several weapon systems, the concept of logistics test has only recently been formalized and no standard method of conducting logistics test exists. Furthermore, the requirements for logistics test are vague, which allows us to tailor the logistics test approach to meet our specific needs. It is also apparent that the best way to conduct logistics test is by somehow evaluating the ten ILS elements as they apply to the system being developed. Also, we discovered that logistics test should be conducted during DT&E, in order to have an early and meaningful influence on system design. Furthermore, we reviewed aspects of past and current logistics test methods and compared their features. With all of this information, we have a clear idea of what we need to know as a result of a logistics test effort. The remainder of the thesis will focus on exactly what we need to do to get these results.

Chapter III: Methodology

Introduction

In the previous chapter, literature concerning logistics test was reviewed to obtain background information which will enable us to develop an appropriate research methodology. In this chapter, we will present the methodology we used to answer the four investigative questions:

1. What are the current regulatory requirements to perform logistics test?

2. What is currently being used as a plan to perform logistics test?

3. What has been used in the past as a plan to perform logistics test?

4. What criteria must be considered to appropriately evaluate logistics test items?

The first three questions were answered by analyzing the literature review. This analysis will be conducted in Chapter IV. To answer the fourth question, we used the Delphi method. The remainder of this chapter will address research design issues, discuss advantages and disadvantages of Delphi, and explain why we believe it is an appropriate approach for our research. We will also address the development of our test instrument, define what we consider to be an expert, and explain how we will know when a consensus is reached. Finally, we will explain how the data will be analyzed.

Research Design

The design of our methodology centers around the fourth investigative question, which was answered using the Delphi method. Specifically, we sent a set of questions to our predetermined experts to evaluate criteria used in past and present logistics test plans. Our definition of an expert was a senior noncommissioned officer, commissioned officer, or civilian equivalent who has achieved his/her rank and position because of a thorough understanding of the logistics, test and evaluation, or logistics test and evaluation. The sample population included past and present members of the acquisition logistics field at System Program Offices, the Air Force Flight Test Center, various Major Command Headquarters, the Air Force Operational Test and Evaluation Center, the Directorate of Acquisition Logistics, the Defense Systems Management College (DSMC), and the Air Force Institute of Technology (AFIT). Because the population of experts was thought to be relatively small, the sample consisted of approximately 30 carefully chosen experts.

In order to maximize our response rate, we contacted the experts prior to sending the questionnaires to them to ensure their cooperation. They returned their responses, along with their justifications of their responses, to us and we analyzed the data. Round two began by sending the questionnaires back to the experts, along with the results of the first round. We asked them to review the data and reconsider their first responses based on this data. This process continued until consensus was reached on all of the target questions on the questionnaire. Our requirements for reaching a consensus are discussed later in this chapter. Consensus was reached after two rounds. Using the final consensus data as input, we developed the template we need to create.

To ensure the reliability of our methods, we concentrated on *equivalence*, which "is concerned with variations at one point in time among observers and samples" (15:186). Equivalence among the observers was achieved by having both the investigators analyze the data together. To help minimize bias, equivalence among the experts was achieved by their thorough understanding that we had no interest in achieving any particular result. We also standardized the conditions among the experts to a reasonable extent by sending the same questionnaire to each expert simultaneously, with each expert having the same amount of time to respond.

Background of the Delphi Method

The Delphi method is a technique for gathering expert opinions and forming solutions after reaching consensus. It "replaces direct confrontation and debate by a carefully planned, orderly program of sequential individual interrogations usually conducted by questionnaires" (5:3). Application of the technique begins by selecting a panel of experts. An expert is defined based upon status among peers, by the amount of experience in a particular field, or by some combination of these. The experts are presented with a questionnaire and asked to respond to questions on an ordinal scale. They are also asked to answer why they respond as they do. The results of the questionnaire are statistically analyzed, usually addressing mean scores (6:3). A second round instrument is then developed and the results of the first round are sent to the expert panel. They are asked to review the first round data and to reconsider their first responses. The questionnaires are then sent back for the researchers to analyze again. There may be as many as four rounds of questionnaires needed to reach consensus (5:5-6).

Advantages of the Delphi Method. The use of the Delphi method has several advantages. First, the negative aspects of group interaction are minimized, yet consensus is still reached. There are three reasons for this advantage: anonymity, controlled feedback, and statistical group response (6:3). Anonymity is the condition where the respondents do not complete the questionnaire together in one location. This anonymity eliminates the influence of a dominant person. Controlled feedback consists of several rounds where the results of the previous round are provided to the respondents. This feedback ensures that all of the experts receive the same information as the others. Statistical response refers to an objective measure of the opinions of the respondents. Also, the consensus is not biased by political pressure toward conformity.

Another advantage of the Delphi method is that the technique is appropriate when very little documented information exists on the subject and the best available information exists as opinion (7:4). This aspect is especially useful when researching a relatively new subject such as logistics test. Yet another advantage is that research has shown that the results of the Delphi method are just as accurate as using committee techniques (8:4).

Disadvantages of the Delphi Method. Despite the many advantages of the Delphi method, there are also some key disadvantages which need to be recognized. One disadvantage is the relatively small sample size of experts typically associated with the Delphi method. Our research included only 32 experts. Also, with Delphi, the selection of experts is neither scientific nor random. One assessment of the Delphi method states "In no sense is Delphi found to be a serious contender in scientific questionnaire development and in the experimentally controlled and replicable application of questionnaires" (25:27). Even with these disadvantages, Delphi has potential uses in several research problems (5:13).

Appropriateness of Delphi for This Research. One reason that the Delphi method was appropriate for this thesis is that very little documented information exists on logistics test. In Chapter I, we stated that one of the reasons why logistics test has not been widely recognized was due to a lack of formal documentation. The vast majority of information on logistics test exists as expert opinion, and Delphi is well suited to analyze that information. Another reason why Delphi is appropriate is because of the geographic separation of the experts and the researchers. It was impractical and too costly to assemble the experts in one location to conduct this research.

Development of the Questionnaire

In order to document the responses of the experts, the instrument we selected was a printed questionnaire. We selected it over other forms, such as a telephone interview, because the printed questionnaire reduces documentation errors and reduces the amount of time required to contact all of the experts (27:277). In this section, we will discuss how we constructed the questionnaire.

Round One. In developing the round one instrument, we organized it from the more general questions to the more specific. Research has shown that respondents feel less threatened to answer questions organized in such a way (27:207). We began by asking each respondent about his/her experience level and organizational affiliation. Next, we asked the experts to express their opinions on the value of logistics test. Following this, the experts were asked to rank order the ten ILS elements and determine what percentage of the logistics effort should be spent on each element. Finally, we asked them to rate the effectiveness of six approaches to logistics test, and provide specific comments about what they liked and disliked about each approach. For these questions, a five-point Likert scale was used and the experts were asked to respond to each statement in terms of five degrees of agreement. The Likert scale is appropriate for this research because it is easy to construct and simple to interpret. Also, the Likert scale is more reliable than other scales and provides a greater volume of data. Another advantage of the Likert scale is that it enhances respondent comfort, ease, and understanding (15:273).

Attached to the eight-page questionnaire was a cover letter explaining our work and the questionnaire, the definitions of the ten ILS elements, and lastly, summaries of the five past and present logistics test approaches described in the last chapter (B-1, F-16, B-2, C-17, F-22). We also added a sixth, fictitious approach. In order to reduce bias, we did not identify the approaches.

This 12-page document was electronically transmitted to 32 experts in various organizations throughout the Air Force. These organizations were: F-22 System Program Office (SPO), F-15 SPO, B-1 SPO, C-17 SPO, B-2 SPO, F-16 SPO, F-117 SPO, Air Force Operational Test and Evaluation Center, the Acquisition Logistics Directorate of Aeronautical Systems Center, Defense Systems Management College, Air Force Institute of Technology, Air Combat Command Headquarters, and Air Mobility Command Headquarters. We provided a one week suspense for the return of the questionnaire. A sample of the Round One questionnaire is included in Appendix A.

Round Two. After we received the results of Round One, we determined if a consensus had been reached during Round One. A consensus was not reached on all questions. Consequently, we proceeded with the Round Two questionnaire development. To develop the Round Two questionnaire, we analyzed the Round One data to determine the mean scores for the questions. We also summarized the "likes and dislikes" comments of the six approaches provided by the experts during Round One.

For Round Two, we provided the respondents with the mean score from Round One for each approach along with the summarized comments. We asked them to reconsider their first round response given this new information and to mark their new response. We also asked them to provide three specific comments for each approach as to why they chose the answer they did. Provided with each questionnaire was a cover letter and the summaries of the six approaches. We again electronically transmitted this package to the original 32 experts who responded in Round One. The experts received a one week suspense for completing Round Two. A sample of the Round Two questionnaire is provided in Appendix B.

Determination of a Consensus

For the questions involving the Likert scale, consensus was determined using a two-rule process. If a question met either of the two rules, a consensus was reached. First, if one specific response (strongly disagree, disagree, agree, or strongly agree) was selected by at least 50% of the experts, then there was a consensus. Second, if one type of response (strongly disagree and disagree, or strongly agree and agree) was selected by 50% of the experts, then there was a consensus (6:9). Thus, for our research, if at least 16 of the 32 experts satisfied one of the above rules, there was a consensus. For the questions involving rank order, consensus was determined by adding the numeric values of each item and then ranking the items according to the least total value being first.

Data Analysis

Round One. In the Round One questionnaire (Appendix A), the first six questions give demographic data about the experts. This data was used to verify that the respondents met our criteria for being an expert. This data also enabled us to group the experts into subgroups. The responses of the subgroups were compared using the Wilcoxon signed rank test to determine if bias existed among the subgroups. The subgroups compared were experts with logistics experience versus experts with test and evaluation experience. This comparison showed if there was a bias based on the experts' area of experience.

Another comparison was a three-way comparison among the places where the experts worked. The three subgroups were SPOs, the AFFTC, and an "other" category consisting of AFOTEC, MAJCOMS, and AFIT. The Wilcoxon signed rank test showed if the probability distribution of the responses were identical, and thus indicated if bias existed within the subgroups. This comparison told us if there was a bias based on where the experts worked.

The next two questions were designed to ascertain the importance of logistics test. The responses to these questions were evaluated only to see if there was consensus among the experts regarding the importance of logistics test. Also, questions 9, 10 and 11 were designed to show if the experts believed the requirement to test the ILS elements is valid. To analyze question 10, we added the ranks given to each element and listed them according to the lowest total rank. This rank order told us the importance of each element relative to the others. The percentages were requested only to allow an expert to show that an element should not be tested at all by assigning a percentage of zero to that element.

Also in the first round, questions 12-17 were analyzed in order to ascertain if a consensus existed among the experts as to which approach is the most effective. We provided the mean ratings for these questions to the experts, along with the likes and dislikes from questions 18-23, in a Round Two questionnaire. For our research, the mean rating provided more information to the experts than the median because several approaches had the exact same median. For example, for a particular approach, if 17 experts rated the approach as "strongly disagree" and 15 experts chose "strongly agree", the median would be "strongly disagree", which would be misleading. Therefore, the experts were asked to again rate the approaches based on this new information about the mean. This process continued until a consensus was reached. Finally, the responses to question 24 were compiled for use in developing the final logistics test template.

Round Two. The data obtained from the Round Two questionnaire (Appendix B) consisted of new ratings for the approaches and the top three reasons why each expert rated each approach the way he/she did. If a consensus was reached, the ratings were compared using the Friedman Test for a randomized block design. This nonparametric test does not require the assumption of normal probability distributions, and it does not require equal variances among the distributions.

The null hypothesis of the Friedman test was that the probability distributions for the ratings of the approaches were identical (meaning the experts think that each approach had the same effectiveness). The alternative hypothesis was that at least two of the probability distributions differed in location (meaning at least two approaches had a different level of effectiveness). If the test told us that two approaches differed, then the Wilcoxon signed rank test for paired difference designs was used to compare the approaches in order to discover the approaches that were statistically different (20:981). This comparison told us which approach (or approaches) were rated significantly higher than the others. The statistically higher approach (or combination of approaches) was used as the baseline for our logistics test template.

Template Development. After the best approach was discovered, we began building our template. The basic methodology of the best approach (or approaches) served as the starting point. Then the comments from question 24 of the round one questionnaire were integrated into this basic methodology. This question asked the experts to provide any additional information about conducting logistics test that was not addressed in our questionnaire. Finally, the responses to our open-ended questions in the round two questionnaire were integrated into the template. These questions asked the experts to list the top three reasons why they liked or disliked each approach. This development provided a standardized template which included inputs from all of the experts.

Summary

This chapter outlined the methodology we used to answer the investigative questions of this thesis. Questions one, two, and three were answered by thoroughly analyzing the literature reviewed in Chapter Two. After considering the advantages and disadvantages of the Delphi method, we chose this method to answer investigative question four. We also discussed the development of the questionnaire that was sent to our predetermined experts and how consensus among the experts was achieved. Finally, we discussed how the data generated by the questionnaires was analyzed and how this data was used to develop the template. In the next chapter we analyze the data obtained from this methodology.

Chapter IV: Analysis

Introduction

The analysis of the data was performed in two phases. The first phase of analysis occurred after the experts returned the round one questionnaire. We first verified that the respondents met our definition of being experts and then we grouped the experts by type of experience and by the organization in which they worked. This allowed us to perform tests for bias among these groups. We also calculated the means and determined if consensus was reached for the effectiveness of the six logistics test approaches. The findings of this first phase analysis were used to construct the round two questionnaire. The second phase of analysis occurred when the experts returned the round two questionnaire. We again calculated the means and determined if consensus was achieved during this round. We then conducted statistical tests to determine the most effective logistics test approach. This chapter provides a detailed discussion of how we analyzed the data during each phase and how the data influenced the development of the logistics test template.

Round One

Demographics. We sent 31 questionnaires to 31 experts and received 30 questionnaires back, for a 96.8 percent response rate. We expected such a high response rate because we had contacted potential respondents in advance to ensure their cooperation. The first data we analyzed were the first six questions of the round one questionnaire (see Appendix A). These questions ask for demographic information about the experts. The first task we performed verified that each respondent met our definition of an expert. Question 1 asked "What is your current rank or grade?" All of the experts were grouped into categories as shown in Table 2.

Table 2

Breakdown of Experts

	Number	Percent
Civilian	10	33%
Commissioned Officer	13	43%
Senior Noncommissioned Officer	7	24%
Total	30	100%

As Table 2 shows, there were 10 civilians, 13 officers, and 7 noncommissioned officers. Also, the lowest civilian grade was GS-12, the lowest officer grade was captain, and the lowest senior noncommissioned officer was master sergeant.

Questions two, three, and four asked the respondents about their years of experience in test and evaluation, logistics, and logistics test and evaluation, respectively. Table 3 shows mean and median years of experience in each category.

Table 3

Years of Experience of Experts

	Logistics	Test and Evaluation	Logistics T&E
Mean	12.36	5.73	2.7
Median	11	4	2

As Table 3 illustrates, as a group, the respondents had the most experience in logistics, with a mean of 12.36 years and a median of 11 years. Our experts' experience in test and evaluation was a mean of 5.73 years and a median of 4 years. As expected, because formal logistics testing is relatively new, the experts had the least amount of experience in the field of logistics test and evaluation, with a mean of 2.7 years and a median of 2 years. Thus, our respondents met our definition of an expert, which was defined as a senior noncommissioned officer, a commissioned officer, or civilian equivalent who has achieved his/her rank and position because of a thorough understanding of logistics, test and evaluation, or logistics test and evaluation.

The next question asked the experts to identify where they currently worked. This question was important to later determine if bias existed among the experts dependent upon where they worked. This question also ensured we obtained a thorough representation of experts in the Air Force. A breakdown of where the experts worked is depicted in Table 4.

Table 4

Breakdown of Experts by Organization

	AFFTC	SPOs	OTHER	TOTAL
NUMBER	9	9	12	30
PERCENT	30%	30%	40%	100%

We obtained a fairly even distribution across these previously selected categories, with the Air Force Flight Test Center (AFFTC) and the System Program Offices (SPOs) each containing 30 percent of the experts. The "Other" category contained 40 percent of the experts. A breakdown of the "Other" category is outlined in Table 5.

Table 5

Breakdown of "Other" Organizations

	HQ/AMC	DSMC	AFIT	HQ/ACC	ASC/AL	AFOTEC
Number of Experts	1	2	2	2	3	2
Percent of "Other" Organizations	.08%	.17%	.17%	.17%	.25%	.17%
Percent of Total Experts	.03	.07	.07	.07	.10	.07

As Table 5 shows, the experts from the "Other" category worked in the major command headquarters of Air Combat Command

(HQ/ACC) and Air Mobility Command (HQ/AMC), Defense Systems Management College (DSMC), Air Force Institute of Technology (AFIT), Air Force Operational Test and Evaluation Center (AFOTEC), and the Aquisition Logistics Directorate of the Aeronautical Systems Center (ASC/AL).

Importance of Logistics Test. The next two questions in the round one questionnaire were designed to ascertain the importance of logistics test. Question 7 asked if a logistics test template would benefit future aircraft acquisiti. programs. Table 6 shows the frequency of response for each value rated by the experts.

Table 6

Frequency of Response to Determine Consensus for Question 7

	Value	Frequency	Percent
	1	0	0.0%
	2	0	0.0%
	3	1	3.3%
	4	13	43.3%
	5	16	53.4%
Total		30	100.0%

The results show that we achieved consensus after the first round because 96.7 percent of the experts either agreed or strongly agreed that a logistics test template would benefit future aircraft acquisition programs. Also, question 8 asked if the results of aircraft logistics test are worth the resources allocated to the test effort. Table 7 illustrates this data.

Table 7

Frequency of Response to Determine Consensus for Question 8

	Value	Frequency	Percent
	1	0	0.0%
	2	0	0.0%
	3	1	3.3%
	4	8	26.7%
	5	21	70.0%
Total		30	100.0%

Again, consensus was reached after round one because 96.7 percent of the experts either agreed or strongly agreed, with a full 70 percent strongly agreeing that results of logistics test are worth the resources allocated to the test effort. These two questions provide strong evidence that the time is right for a formal logistics test template.

ILS elements. After determining that a template is needed and, as discussed in Chapter II, because any logistics test template should logically include a way of addressing the ten ILS elements, questions 9-11 asked the experts to validate the ten ILS elements outlined in DODI 5000.2. Question 9 asked if the ten ILS elements cover what needs to be logistically tested for an aircraft program. Table 8 provides the frequency of response data for this issue.

Table 8

Frequency of Response to Determine Consensus for Question 9

	Value	Frequency	Percent
	1	1	3.3%
	2	1	3.3%
	3	0	0.0%
	4	22	73.4%
	5	6	20.0%
Total		30	100.0%

Over ninety-three percent of the experts either agreed or strongly agreed with this statement. Thus, consensus on this question was achieved. Furthermore, question 10 asked the experts to rank order the importance of the ten ILS elements and to provide a percentage of the total logistics test effort that should be devoted to each element. Also, question 11 gave the experts an opportunity to provide any other ILS element that was not included in the original ten elements but should have been included in addition to the original ten. Our purpose in asking questions 9-11 was to validate the ten ILS elements by determining if the experts reached a consensus on whether some elements should be deleted or new elements should be added. Table 9 shows how the experts ranked the importance of each ILS element.

Table 9

Expert Ranking of ILS Elements

Rank	Element
1.	Technical Data
2.	Design Interface
3.	Support Equipment
4.	Maintenance Planning
5.	Training and Training Support
6.	Manpower and Personnel
7.	Supply Support
8.	Computer Resources Support
9.	Facilities
10.	Packaging, Handling, Storage, and Transportation

This ranking in itself is not as important as the consensus reached where the experts overwhelmingly agreed that there should be no deletions to the list of ten elements. Eighty percent of the experts agreed that all of the elements need to be tested to some extent. Furthermore, 87 percent of the experts agreed that no additional ILS elements are needed. Thus, we concluded that the ten ILS elements outlined in DODI 5000.2 are valid and should serve as the basis for a logistics test template.

Logistics Test Approaches. Questions 12-17 asked the experts to rate the effectiveness of each of the six approaches discussed previously. Our purpose for these questions was to determine if a consensus existed as to the effectiveness of each approach and to obtain mean ratings to compare the effectiveness of each approach. Question 12 asked the experts about approach one, which represents the approach used by the

F-16 program. Table 10 shows the frequency of response data for this question. The mean rating for this question was 2.39.

Table 10

Frequency of Response to Determine Consensus for Approach 1

	Value	Frequency	Percent
	1	3	10.0%
	2	15	50.0%
	3	10	33.3%
	4	2	6.7%
	5	0	0.0%
Total		30	100.0%

As Table 10 shows, 60 percent of the experts rated this approach as either very ineffective or ineffective. Thus, consensus was reached for this question.

Question 13 asked about the effectiveness of approach two, which is the method used by the C-17 program. Table 11 provides the data for this question. The mean rating for this approach was 3.55.

Table 11

Frequency of Response to Determine Consensus for Approach 2

	Value	Frequency	Percent
	1	0	0.0%
	2	3	10.0%
	3	8	26.7%
	4	19	63.3%
	5	0	0.0%
Total		30	100.0%

Because 63.3 percent of the experts rated this approach as effective, we had a consensus for this question.

In the next question, the experts were asked to rate approach three, which represents the B-1 program. The data for this question follows. The mean rating for this approach was 2.90.

Table 12

Frequency of Response to Determine Consensus for Approach 3

	Value	Frequency	Percent
	1	0	0.0%
	2	11	36.7%
	3	10	33.3%
	4	8	26.7%
	5	1	3.3%
Total		30	100.0%

We did not achieve consensus for this question in this round. Only 36.7 percent of the experts rated this approach as either "1" or "2" and only 30 percent rated this approach as a "4" or "5". Also, no individual rating was greater than 50 percent. Therefore, a round two questionnaire was necessary.

Question 15 asked about approach 4, which the B-2 program uses. Table 13 provides the data for this question. The mean rating for this approach was 3.71.

Table 13

Frequency of Response to Determine Consensus for Approach 4

	Value	Frequency	Percent
	1	0	0.0%
	2	2	6.7%
	3	10	33.3%
	4	12	40.0%
	5	6	20.0%
Total		30	100.0%

The table shows that 60 percent of the experts rated this approach as either effective or very effective. Thus, we had a consensus for this question.

The next question pertains to approach five, which is the method used by the F-22 program. The following table outlines the frequency of response data for this question. The mean rating for this approach was 4.10.

Table 14

Frequency of Response to Determine Consensus for Approach 5

	Value	Frequency	Percent
	1	0	0.0%
	2	3	10.0%
	3	6	20.0%
	4	7	23.3%
	5	14	46.7%
Total		30	100.0%

Because 70 percent of the experts rated this approach as either effective or very effective, we achieved consensus for this question.

Question 17 asked about approach six, which is the unstructured, traditional approach. Table 15 shows the data for this question. The mean rating for this approach was 1.96.

Table 15

Frequency of Response to Determine Consensus for Approach 6

	Value	Frequency	Percent
	1	11	36.7%
	2	11	36.7%
	3	6	20.0%
	4	2	6.6%
	5	0	0.0%
Total		30	100.0%

The table shows that 73.4 percent of the experts rated this approach as either very ineffective or ineffective. Thus, we had a consensus for this question.

Analyzing questions 12-17 provided information on whether a consensus was reached for all of the approaches. Table 16 summarizes these results and shows the mean scores for each approach. This data indicated that we do not have a total consensus among our experts because they could not agree on the effectiveness of approach three. Also, the mean scores were used to provide the experts additional information to consider for the next round questionnaire.

Table 16

Summary of Questions 12-17

Approach	Consensus	Mean Score
F-16 (1)	YES	2.39
C-17 (2)	YES	3.55
B-1 (3)	NO	2.90
B-2 (4)	YES	3.71
F-22 (5)	YES	4.10
Traditional (6)	YES	1.94

After the experts rated the approaches, we then asked them to list the characteristics in questions 18-23 that they liked and disliked about each approach. The purpose of these questions was to provide the experts with additional information for the round two questionnaire. We condensed these comments by listing each comment only once, even if it was repeated among experts. This format allowed the experts to consider the views of other experts along with the mean scores for each approach. These comments are located in the round two questionnaire in Appendix B.

The final question in the round one questionnaire was question 24. The experts were asked if there were any other approaches or aspects of logistics test that were not included in the questionnaire but that they felt needed to be addressed.

These comments, which are contained in Appendix C, were used during the template development.

Round Two

For the round two questionnaire (see Appendix B), we provided the experts with a list of likes and dislikes for each approach and the round one mean rating for each approach. We sent out 30 questionnaires to the 30 respondents from round one, and we received a total of 25 round two questionnaires, for an 83.3 percent response rate. Five respondents could not respond due to extended absences, such as temporary duty assignments, vacations, or illnesses. We asked the experts to consider the new information regarding the means and likes and dislikes, and to rate the approaches again. Our purpose for asking these questions was to establish a consensus among the experts pertaining to the effectiveness of each approach. Also, to assist us in template development, we asked the experts to list the top three reasons (likes, dislikes, or a combination of both) for their rating for each approach.

Even though we had consensus on five of the six approaches after round one, we asked the experts to again rate these approaches so that we could get their top three reasons for their ratings and so they could consider the new information and change their rating if desired. Also, instead of isolating the one approach on which we did not have consensus, we chose to have the experts rate all of the approaches as a group because of the dependence that one rating may have on another.

Approaches. The first question in round two asked the experts to rate the effectiveness of approach one, the F-16 approach, at accomplishing logistics test. Table 17 provides the frequency of response data for this question. The mean rating for this approach was 1.92, which represented a moderate decrease from the round one mean rating of 2.39.

Table 17

Frequency of Response Data, Round Two, Approach 1

	Value	Frequency	Percent
	1	2	8.0%
	2	23	92.0%
	3	0	0.0%
	4	0	0.0%
	5	0	0.0%
Total		25	100.0%

A full 100 percent of the experts rated this approach as very ineffective or ineffective. This, of course, is a consensus.

The next approach to be considered was approach two, which is the C-17 approach. Table 19 shows the data for this question. The mean rating for this approach was 3.48, which was a slight decrease from the round one mean rating of 3.55. As Table 18 illustrates, 52 percent of the experts rated approach 2 as either effective or very effective. Thus, we have a consensus.

Table 18

Frequency of Response Data, Round Two, Approach 2

Value	Frequency	Percent
1	0	0.0%
2	2	8.0%
3	10	40.0%
4	12	48.0%
5	1	4.0%
Total	25	100.0%

Approach three, the B-1 approach, was the next approach that the experts considered. Table 19 provides the data for this question. The mean rating for this approach was 2.68, which was a moderate decrease from the round one mean rating of 2.90.

Table 19

Frequency of Response Data, Round Two, Approach 3

Value	Frequency	Percent
1	0	0.0%
2	10	40.0%
3	13	52.0%
4	2	8.0%
5	0	0.0%
Total	25	100.0%

Because 52 percent of the experts rated this approach as neutral, we have consensus for this question.

The next question asked the expe to rate approach four, which is the B-2 approach. Table 20 shows the data for this question. The mean rating for approach four was 3.72, or just 0.01 higher than the round one mean rating of 3.71.

Table 20

Frequency of Response Data, Round Two, Approach 4

Value	Frequency	Percent
1	0	0.0%
2	1	4.0%
3	9	36.0%
4	11	44.0%
5	4	16.0%
Total	25	100.0%

A consensus of 60 percent of the experts rated this approach as either effective or very effective.

Question nine asked about the effectiveness of approach five, which is the F-22 approach. Table 21 provides the frequency of response data for this question. The mean rating for this approach was 4.28, compared to the round one mean rating of 4.10.

Table 21

Frequency of Response Data, Round Two, Approach 5

Value	Frequency	Percent
1	0	0.0%
2	0	0.0%
3	2	8.0%
4	14	56.0%
5	9	36.0%
Total	25	100.0%

Ninety-two percent of the experts rated this approach as either effective or very effective, which constitutes a consensus for this question.

The last approach was approach six, which represents the traditional approach. Table 22 outlines the data for this question. The mean rating for this approach was 1.48, which was a substantial decrease from the round one mean rating of 1.9.

Table 22

Frequency of Response Data, Round Two, Approach 6

	Value	Frequency	Percent
	1	14	56.0%
	2	10	40.0%
	3	1	4.0%
	4	0	0.0%
	5	0	0.0%
Total		25	100.0%

Because 96 percent of the experts rated this approach as either very ineffective or ineffective, we have a consensus on this question.

Summary of Round Two Ratings. The previous six tables showed the frequency of response data for the six approaches from the round two questionnaire. The next table, Table 23, provides a summary of this data as it pertains to determining consensus and the mean ratings for each approach. This information shows that consensus was reached on all of the questions. Therefore, a round three questionnaire was not necessary.

Table 23

Summary of Approach Ratings, Round Two

Approach	Consensus	Mean Score
F-16 (1)	YES	1.92
C-17 (2)	YES	3.48
B-1 (3)	YES	2.68
B-2 (4)	YES	3.72
F-22 (5)	YES	4.28
Traditional (6)	YES	1.48

Top Three Reasons for Experts' Ratings. The round two questionnaire also asked the experts to address the top three reasons for rating each approach the way they did. These reasons are summarized in Appendix D and were used in the development of the final logistics test template. The use of these reasons is further discussed in Chapter V.

Statistical Tests for Approach Effectiveness. Once we achieved consensus on all of the questions, we were prepared to statistically analyze the data to discover if one approach (or approaches) was rated by the experts to be significantly more effective than the other approaches. The Friedman Test for a randomized block design was used to compare the ratings of the approaches. Table 24 shows the data used to conduct the Friedman Test.

Table 24

Probability Distributions of the Approach Ratings
(Friedman Test)

Exp	Ap1	Rk	Ap2	Rk	Ap3	Rk	Ap4	Rk	Ap5	Rk	Ap6	Rk
1	2	1.5	3	3.5	3	3.5	4	5.5	4	5.5	2	1.5
2	1	1.5	3	4	2	3	4	5	5	6	1	1.5
3	2	2.5	3	4.5	2	2.5	3	4.5	5	6	1	1
4	2	1.5	3	3.5	3	3.5	5	6	4	5	2	1.5
5	2	2	5	5.5	3	3	5	5.5	4	4	1	1
6	2	1.5	4	5	3	3	4	5	4	5	2	1.5
7	2	1	4	4.5	3	2.5	4	4.5	5	6	3	2.5
8	2	2	3	4.5	2	2	3	4.5	4	6	2	2
9	2	2.5	4	5	2	2.5	3	4	5	6	1	1
10	2	1.5	4	4.5	3	3	4	4.5	5	6	2	1.5
11	2	1.5	3	3	4	4.5	4	4.5	5	6	2	1.5
12	2	2.5	3	4	2	2.5	4	5.5	4	5.5	1	1
13	2	1.5	3	3.5	4	5.5	3	3.5	4	5.5	2	1.5
14	2	2.5	4	4.5	2	2.5	4	4.5	5	6	1	1
15	2	1.5	4	5	3	3	4	5	4	5	2	1.5
16	2	2.5	4	5.5	3	4	2	2.5	4	5.5	1	1
17	2	1.5	3	4	3	4	3	4	4	6	2	1.5
18	2	2	4	4.5	3	3	5	6	4	4.5	1	1
19	2	3	2	3	2	3	3	5	5	6	1	1
20	2	2.5	4	5.5	2	2.5	3	4	4	5.5	1	1
21	2	2	3	3.5	3	3.5	4	5	5	6	1	1
22	2	2	4	5.5	2	2	3	4	4	5.5	2	2
23	1	1.5	4	5.5	2	3	3	4	4	5.5	1	1.5
24	2	2	3	4	3	4	4	6	3	4	1	1
25	2	2.5	2	2.5	3	4.5	5	6	3	4.5	1	1
Total		48.5		108		80		118.5		136.5		33.5

KEY: Exp = Expert Ap1-6 = Approaches 1-6 Rk = Rank

The 25 experts served as the "blocks" for this comparison, and the six approaches served as the treatments. Thus, for each expert, we ranked how he or she rated each approach. For example, for each expert, the lowest rated approach was given a

"1" ranking, the second lowest approach received a "2" ranking, and the third lowest approach received a "3" ranking. This process continued for each expert, giving the top rated approach for each expert a "6" ranking. In case of ties, the tied ratings were given the average of the rankings they would have received if they were unequal but occurred in successive order. Then the rankings for each approach were added and a total for each approach was determined. For example, approach one had a total of 48.5.

The null hypothesis for this test was that the probability distributions of the ratings for each approach were identical. The alternative hypothesis was that at least two of the six distributions differed. The test statistic for this test was calculated using equation (1).

$$F_r = \frac{12}{bp\,(p+1)} \sum_{j=1}^{p} R_j^2 - 3b(p+1) \qquad (1)$$

where b = number of blocks (25), p = number of treatments (6), and R_j = rank sum of the j^{th} treatment. Substituting the values into equation (1) gives F_r = 94.577. The rejection region is the Chi-square value with (p-1) = 5 degrees of freedom and a significance level of 0.1. This Chi-square value is 9.236 (20:1187). Because the test statistic of 94.577 is greater than the Chi-square statistic of 9.236, we rejected the null hypothesis (20:980). Thus, we concluded that at least two of the approaches were rated by the experts at a different level of effectiveness. We now needed to determine which approaches were rated differently, and which approach (or approaches) was rated as significantly more effective than the others.

In order to determine which approaches were rated significantly different from the others, we chose the Wilcoxon signed rank test as discussed in Chapter III. This test allowed us to compare the approaches in pairs. The first pair to be compared were the two approaches with the highest mean ratings, approaches four (B-2, 3.72) and five (F-22, 4.28). We conducted this Wilcoxon signed rank test manually to demonstrate the steps involved in the process of how the comparisons were made. Table 25 shows the organization of the data for this test.

The table shows the ratings given to approaches four and five by each expert, the difference between the ratings, the absolute value of the difference, and the rank of the difference. In case of ties, the tied ratings received the average of the ranks they would have received if they were unequal but occurred in successive order. After ranking the differences between the approaches, the positive ranks were added for a total of 53. Similarly, the negative ranks were added for a total of 178. The test statistic for this comparison is the smaller of these two totals, which is 53. The null hypothesis for this test is that the two approaches have identical probability distributions. The alternative hypothesis is that the probability distribution of approach

Table 25

Determination of a Difference Between Approaches 4 and 5
(Wilcoxon Signed Rank Test)

Expert	Ratings For: Ap4	Ap5	Dif	Abs Val of Dif	Rank
1	4	4	0	0	2.5
2	4	5	-1	1	13
3	3	5	-2	2	23.5
4	5	4	1	1	13
5	5	4	1	1	13
6	4	4	0	0	2.5
7	4	5	-1	1	13
8	3	4	-1	1	13
9	3	5	-2	1	13
10	4	5	-1	1	13
11	4	5	-1	1	13
12	4	4	0	0	2.5
13	3	4	-1	1	13
14	4	5	-1	1	13
15	4	4	0	0	2.5
16	2	4	-2	2	23.5
17	3	4	-1	1	13
18	5	4	1	1	13
19	3	5	-2	2	23.5
20	3	4	-1	1	13
21	4	5	-1	1	13
22	3	4	-1	1	13
23	3	4	-1	1	13
24	4	3	1	1	13
25	5	3	2	2	23.5

T(+)=53
T(-)=178

four is shifted to the right or to the left of that for approach five. The rejection region for this test using a significance level of 0.1 is 101 (20:1185). Because the test statistic of 53 was less than the rejection region of 101, we rejected the null hypothesis of identical probability distributions between approaches four and five (20:965). Thus, we concluded that approach five, the F-22 approach, was rated as significantly more effective than approach four, the B-2 approach. Furthermore, because approach five had the highest mean of all the approaches, and approach four had the next highest mean, it follows that approach five is also significantly more effective than all of the other approaches. Using a statistical computer program (26:111), we confirmed our reasoning by comparing all possible combinations of approaches. In fact, every combination of the effectiveness of the approaches proved to be significantly different from the others with one exception. Approach two (C-17) and approach four (B-2) were the only two approaches with identical probability distributions. These results are contained in Appendix E.

Statistical Tests for Bias. Having determined which approach was rated most effective by the experts, we were interested in determining if the ratings were influenced by the respondents' field of expertise (logistics or test and evaluation) or by where they worked (SPO, AFFTC, other). To be consistent with the previous tests, we used a significance level of 0.1 for all tests for bias. The first test for bias is shown in Table 26.

The purpose of this test was to determine if bias existed based on the experts' experience. Experts were grouped as having either logistics experience or test and evaluation experience. Experts with experience in both areas were grouped according to the area in which they had the most experience. The mean ratings for each approach within these groups are

Table 26

Determination of Bias Based on Experience
(Wilcoxon Signed Rank Test)

	Mean Rating From:			Abs Val of	
Approach	Log	Test	Dif	Dif	Rank
1	1.90	2.00	-0.10	0.10	2
2	3.40	3.80	-0.40	0.40	6
3	2.70	2.60	+0.10	0.10	2
4	3.65	4.00	-0.35	0.35	5
5	4.30	4.20	+0.10	0.10	2
6	1.45	1.60	-0.15	0.15	4
					T(+)=4
					T(-)=17

shown in the table, along with the differences and absolute values of the differences of the mean ratings. The null hypothesis is that the probability distributions of the mean ratings for each group are identical. The test statistic was four, and the rejection region was two. Because the test statistic was greater than the rejection region, we accepted the null hypothesis and concluded that no bias existed based on the respondents' experience.

The next test for bias was actually a series of tests to determine if bias existed based upon where the experts worked. The first test in this series was between experts from the SPOs versus experts from the "other" category (see Table 6 for a breakdown of this category). Table 27 shows the data for this test.

Table 27

Determination of Bias Based on Organization, SPO vs. Other (Wilcoxon Signed Rank Test)

Approach	Mean Rating From: SPOs	Other	Dif	Abs Val of Dif	Rank
1	1.86	2.00	-0.14	0.14	3
2	3.71	3.56	+0.16	0.16	4
3	2.71	2.89	-0.17	0.17	5
4	4.29	3.33	+0.96	0.96	6
5	4.43	4.33	+0.10	0.10	2
6	1.71	1.67	+0.05	0.05	1
					T(+)=13
					T(-)=8

The null hypothesis for this test was that the probability distributions of the mean ratings for the approaches are identical between experts from SPOs and experts from the "other" category. The test statistic was eight and the rejection region was two. Because the test statistic was greater than the rejection region, we accepted the null hypothesis and concluded that no bias existed between these organizations.

The next test for bias was between experts from the AFFTC and the "other" category. Table 28 shows the data for this test. The null hypothesis for this test was that the probability distributions of the mean ratings for the approaches were identical between the AFFTC and the "other" category. Because the test statistic of 3.5 was greater than 2, we accepted the null hypothesis and concluded that no bias existed between experts from the AFFTC and experts from the "other" category.

Table 28

Determination of Bias Based on Organization, AFFTC vs. Other (Wilcoxon Signed Rank Test)

Approach	Mean Rating From: AFFTC	Other	Dif	Abs Val of Dif	Rank
1	1.89	2.00	-0.11	0.11	1
2	3.22	3.56	-0.34	0.34	3.5
3	2.44	2.89	-0.45	0.45	5
4	3.67	3.33	+0.34	0.34	3.5
5	4.11	4.33	-0.22	0.22	2
6	1.11	1.67	-0.56	0.56	6
					T(+)=3.5
					T(-)=17.5

The last comparison was made between experts who worked in the SPOs versus those who worked at the AFFTC. Table 29 provides the information for this test.

Table 29

Determination of Bias Based on Organization, SPO vs. AFFTC (Wilcoxon Signed Rank Test)

Approach	Mean Rating From: SPOs	AFFTC	Dif	Abs Val of Dif	Rank
1	1.86	1.89	-0.03	0.03	1
2	3.71	3.22	+0.49	0.49	4
3	2.71	2.44	+0.27	0.27	2
4	4.29	3.67	+0.62	0.62	6
5	4.43	4.11	+0.32	0.32	3
6	1.71	1.11	+0.60	0.60	5
					T(+)=20
					T(-)=1

The null hypothesis for this test was that the probability distributions of the mean ratings for the approaches from the SPOs and the AFFTC were identical. Because the test statistic of one was less than the rejection region of two, we rejected the null hypothesis in favor of the alternative hypothesis, which was that the probability distributions of the approach mean ratings were not identical. Therefore, we concluded that bias existed between experts in SPOs and in the AFFTC. It appears that the experts from SPOs rated the approaches higher than experts from the AFFTC. We could not determine any cause for this bias and concluded that this bias would not affect the development of the template.

Conclusion

In this chapter, we analyzed the data collected from both round one and round two questionnaires. In round one, we verified that our respondents met our definition of being experts and grouped the experts by type of experience and by organization. We also gained consensus that a logistics test template would benefit future aircraft acquisition programs and that logistics test is worth the resources allocated to it. Also in round one, we verified that the experts had no additions or deletions to the ten ILS elements outlined in DODI 5000.2. This is important to template development, as any template would logically be based on the ten ILS elements. Finally, in round one we analyzed the data for the effectiveness of the six approaches and determined that total consensus had not been reached, thus necessitating a round two questionnaire.

In analyzing the round two data, we finally gained consensus on the effectiveness of all six approaches. We then performed a Friedman test for a randomized block design to determine if there were any differences in the effectiveness of the approaches. This test revealed that at least two approaches differed. We then performed a Wilcoxon signed rank test between the approaches with the two highest mean ratings, which were the B-2 and the F-22 approaches. This test revealed that the effectiveness of the F-22 approach was significantly higher than the effectiveness of the B-2 approach, and therefore higher than all other approaches. Thus, the F-22 approach served as the baseline for our template. This approach was modified using data from question 24 of round one (see Appendix C) and the top three likes and dislikes from round two (see Appendix D).

Also in round two, we tested for bias among the experts according to their experience and their organization. We concluded that the only bias was between experts from SPOs and experts from the AFFTC. This bias, however, should not affect the development of the logistics test template.

Based on the information gained from this analysis, the template was developed and is presented in the next chapter. Chapter V also includes conclusions, recommendations, and suggestions for further research.

Chapter V: Conclusion

Introduction

In the first four chapters, we identified the need for a logistics test template and researched the available literature to discover the requirements necessary for a logistics test and evaluation program. Using the Delphi method, we then assembled a panel of experts to discover the most effective approach to logistics test and evaluation. After statistically analyzing the data and discovering that the F-22 approach is most effective, we were then able to create a template by incorporating the comments made by the experts into the F-22 logistics test plan.

Using the F-22 logistics test and evaluation plan as the framework, we ensured that all of the comments from question #24 (round one), which asked for other aspects of approaches not included in the original six approaches, were included in the template. Additionally, the top three likes and dislikes of the six approaches identified in round two were also incorporated into the template. The following section is the result of this effort.

Logistics Test and Evaluation Template

Purpose. The purpose of aircraft logistics test and evaluation is to assess the efforts in design and decision making which seek to improve reliability, maintainability and supportability, to reduce life-cycle costs and to minimize risks associated with transitioning from production to operation of the aircraft being tested. The logistics test and evaluation process begins early in the DT&E process, and it measures, evaluates and documents information on all of the ILS elements outlined in DODI 5000.2. The findings of the test process are reported so decision makers can take corrective actions or make trade-off decisions in designs or concepts (1:C.1-1).

The template we developed provides an approach rated by a panel of experts as the most effective in conducting logistics test and evaluation. As shown in Chapter IV, approach five was rated to be the most effective. Because approach five represented the approach used by the F-22 program, it served as the framework for this template. The template should be used as a guide only and would need to be modified and/or outlined in more detail according to the requirements of the aircraft being tested. The template's purpose is to provide logistics test and evaluation personnel with a scientifically evaluated approach to logistics test and evaluation out of the many conceived and executed in the past. This template will save time and money because the development of a logistics test and evaluation approach from scratch is unnecessary.

Organization. Broadly explained, the organizational elements necessary for this approach include members from the aircraft contractor, the Air Force Flight Test Center (AFFTC), the Air Force Operational Test and Evaluation Center (AFOTEC), the using command, the implementing command, which is usually

Air Force Materiel Command (AFMC), and Air Education and Training Command (AETC).

How the members from these agencies are organized to perform logistics test and evaluation will depend heavily upon the inputs from all of these agencies. However the effort is organized, the key point to this approach is that members from all of the previously mentioned agencies must work as one team and that adversarial relationships are a thing of the past. Even though an agency such as AFOTEC must remain independent to properly perform its mission, such cooperation allows larger sample sizes to be collected with little or no duplication of tests.

Planning. This section describes in general terms the steps to take to plan this approach.

1. Identify which of the ILS elements are to be tested as part of a specific logistics test and evaluation plan. While it is true that DODI 5000.2 states that all ILS elements must be addressed, some of the elements may be included in separate test efforts. For example, the F-22 program has a separate Training System Test Plan which is not included as part of the DT&E Logistics Test and Evaluation Integration Plan. Also, facilities are not being tested as part of this plan because the facilities at the AFFTC are not representative of the facilities of the using command. The other eight ILS elements were selected to be tested in one plan because they can be tested as part of the flight test program.

2. Identify the major subsystems to be tested. The number and identity of subsystems to be tested is a group decision to be made by the contractor, AFFTC and AFOTEC.

3. Create Test Information Sheets (TISs) for each subsystem. A TIS provides information such as the objectives of the test, test description, test conditions, test requirements, and test schedule. There may be any number of TISs for each subsystem, depending on its complexity. Each TIS should identify what ILS elements apply to ensure that all ILS elements are addressed as part of the test effort. Additionally, the TISs should be developed such that the system is tested in an integrated manner rather than in isolation.

4. Create a table of all of the ILS elements. Break down the ILS elements into subelements and use specific, measurable criteria to evaluate these ILS subelements. See Appendix F for an example of how this table may be accomplished.

Implementation.

1. Integrate logistics test scheduling with flight test scheduling. Logistics test and evaluation efforts should be given the same priority as flight test activities.

2. Conduct tests according to TIS priority and the schedule previously coordinated with the flight test effort.

3. Conduct a post-test debriefing to review the data collected from the test and to ensure that all requirements of the test have been met.

Reporting. AFOTEC, AFFTC, and the contractor should compile independent reports. The reports should be in three parts. Part 1 should be an overall status of the test effort by ILS element. This part will detail the progress being made to test all ILS elements. Part 2 should be an assessment of all logistics elements of the system. This part addresses the test effort by subsystem and documents the progress being made to test the subsystems. Part 3 should provide an opportunity for feedback.

Conclusions

Our research has enabled us to reach several conclusions about the aircraft logistics test and evaluation process and the development of a template.

1. The ten ILS elements are indeed a valid baseline for a logistics test and evaluation template. We reached consensus among the experts that there should be no additions nor deletions to the ten ILS elements listed in DODI 5000.2, and the approach rated most effective by the experts utilized the ten ILS elements to conduct logistics test and evaluation.

2. The aircraft logistics test and evaluation process has continuously evolved. Even today, improvements are still being made to the process in current aircraft programs. For example, F-22 program decision makers learned from the B-2 program that individual logistics tests should be

scheduled events rather than unscheduled events to be accomplished as opportunities arise.

3. Logistics test and evaluation is worth the resources allocated to the test effort. This conclusion suggests that a greater amount of resources spent on the logistics test process during DT&E would yield greater dividends throughout the life cycle of the system. For example, during DT&E of the C-17 program, evaluators discovered that the main landing gear pins had excessive play. The contractor was able to redesign the pins before production, thus preventing costly changes in the future (24:17B).

4. The F-22 program represents the most effective approach to aircraft logistics test and evaluation. It incorporates all of the aspects of logistics test which were important to the experts.

Recommendations and Suggestions for Future Research

Based on the above conclusions, we recommend that aircraft acquisition programs use our template (which emulates the F-22 approach) to conduct logistics test and evaluation. To obtain a copy of the F-22 Logistics Test and Evaluation Integration Plan, Contract F33657-91-C-006, Contract Data Requirements List (CDRL) A191, contact the F-22 System Program Office, Wright-Patterson AFB, Ohio. This document was not included in this thesis because it was incomplete and still in draft form as of the publishing date of this thesis. Because this approach is still in development, future research could focus on the following areas:

1. Conduct a case study to track the F-22 logistics test and evaluation program to determine its effectiveness.

2. Perform a cost/benefit analysis to determine the optimum proportion of resources to be allocated to the logistics test and evaluation effort.

3. Computerize the template and make it more detailed based on the actual effectiveness of the F-22 logistics test and evaluation program. Only then can this template be finalized and incorporated into an official publication to be used by aircraft acquisition programs.

4. Determine if this template is applicable to future programs such as the Joint Primary Aircraft Training System (JPATS).

Logistics test and evaluation during DT&E is only now being recognized as an important part of the aircraft acquisition process. The continuous evolution of logistics test has resulted in a comprehensive, measureable, and useful program developed by the F-22 SPO. The Air Force should continue its efforts to legitimize and formalize the logistics test and evaluation process. This would ensure that the most cost effective, maintainable, and best performing aircraft systems are developed for our national defense needs.

Appendix A: Round One Questionnaire

As you should already know, we are students at the Air Force Institute of Technology. We are conducting research with the intent to develop a template for accomplishing logistics test and evaluation during the DT&E of aircraft acquisition programs. This research is sponsored by the 412 Test Wing at Edwards AFB.

We are using the Delphi method to discover the most appropriate way to conduct logistics test and evaluation. The Delphi method is an analysis technique which uses the opinions of carefully chosen experts in order to evaluate the pros and cons of a particular subject. You are one of 30 experts chosen to participate in this research.

Enclosed you will find a set of six different approaches to logistics test and evaluation that have been used in the past or are currently being used. Please read them carefully and understand their differences and similarities. Each one is unique and if some seem more detailed than others it is because they were originally intended to be so by their creators.

You will also find a four-page questionnaire. Based on your experience and the six approaches, please answer the questions as thoroughly as possible. A list of the ten ILS elements as they are defined in DOD Instruction 5000.2 is included for your reference. We estimate a total time of 30 minutes required to complete the questionnaire.

The goal of the Delphi method is to reach a consensus among the experts. This may require several rounds of questions. This questionnaire is considered as Round One. After Round One, we will compile the answers that you provided and return them to you so that you can review all of the answers and possibly modify your responses based on this new information. This will be Round Two. After Round Two, the answers will be statistically evaluated in order to ascertain the existence of a consensus.

This questionnaire is being sent to experienced logisticians and test and evaluation personnel (who will remain anonymous). The Delphi method requires that each expert be given the same amount of time to respond. Accordingly, we request that you FAX the questionnaire only (we do not need the rest of the material) to us at DSN 986-7988 or commercial (513)476-7988, NOT LATER THAN 15 April 1994. If you have any questions about this questionnaire, please contact us immediately at DSN 785-7777, ext. 2179.

Thank you for taking the time to answer our questionnaire. We feel the results of this research will be well worth the effort.

ILS Elements:

1. Maintenance Planning-The process conducted to evolve and establish maintenance concepts and requirements for the lifetime of the system.

2. Manpower and Personnel-The identification and acquisition of military and civilian personnel with the skills and grades required to operate and support the system over its lifetime at peacetime and wartime rates.

3. Supply Support-All management actions, procedures, and techniques used to determine requirements to acquire, catalog, receive, store, transfer, issue, and dispose of secondary items. This includes provisioning for both initial support and replenishment supply support. It includes the acquisition of logistics support for support and test equipment.

4. Support Equipment-All equipment (mobile or fixed) required to support the operation and maintenance of the system. This includes associated multi-use end items, ground handling and maintenance equipment, tools, metrology and calibration equipment, test equipment, and automatic test equipment.

5. Technical Data-Scientific or technical information recorded in any form or medium (such as manuals and drawings). Computer programs and related software are not technical data; documentation of computer programs and related software are. Also excluded are financial data or other information related to contract administration.

6. Training and Training Support-The processes, procedures, techniques, training devices, and equipment used to train civilian and active duty and reserve This includes individual and crew training (both initial and continuation); new equipment training; initial, formal, and on-the-job training; and logistics support planning for training equipment and training device acquisitions and installations.

7. Computer Resources Support-The facilities, hardware, system software, software development, and support tools, documentation, and people needed to operate and support embedded computer systems.

8. Facilities-The permanent, semipermanent, or temporary real property assets required to support the system, including conducting studies to design facilities or facility improvements, locations, space needs, utilities, environment requirements, real estate requirements, and equipment.

9. Packaging, Handling, Storage, and Transportation-The resources, processes, procedures, design considerations, and

methods to ensure that all system, equipment, and support items are preserved, packaged, handled, and transported properly, including environmental considerations, equipment preservation requirements for short and long term storage, and transportability.

10. Design Interface-The relationship of logistics related design parameters to readiness and support resource requirements. These logistics related design parameters are expressed in operational terms rather than as inherent values and specifically relate to system readiness objectives and support costs of the system.

APPROACH #1

This approach does not use a detailed logistics test plan. It begins by identifying the objectives of the logistics test effort, which include determining support equipment compatibility and training adequacy, technical order verification and validation, measuring system maintainability, and identifying hardware and software deficiencies (reliability). These objectives are tested in isolation, which means that one objective is tested first, then it is closed out, and then another objective is tested, and so on. The reliability and maintainability (R&M) testing focuses on R&M aspects of the entire system, rather than on subsystems. Any other deficiency or needed improvements will be recognized and identified as a result of flight testing; therefore, no additional planning efforts are required.

APPROACH #2

This approach is a systematic evaluation of the ten ILS elements as they pertain to three disciplines: Reliability and Maintainability (quantitative measures), Human Factors (qualitative), and Logistics Test (quantitative and qualitative).
These disciplines are broken down into measurable units called Test Measurements (TM) for each element. For example, for the ILS element of Design Interface, the three disciplines are broken down as follows:

Design Interface
1. Reliability and Maintainability:
   TM: Reliability (this might include a list of top logistic problems that need to be addressed)
   TM: Maintainability (this might include elapsed repair times for maintenance tasks)
2. Human Factors:

TM: Accessibility (qualitative assessment of how easily the system is serviced and maintained)

3. Logistics Test:
   TM: Energy Consumption (fuel, oil requirements)
   TM: Component Standardization

Each ILS element is analyzed in a similar manner, using the same three disciplines, but with different TMs that are specific to the particular ILS element.

This approach is framed around the ten ILS elements, and once each ILS element is analyzed, then Logistics Testing can be considered complete.

APPROACH #3

This approach focuses on eight logistics test objectives aimed primarily at confirming data supplied by the contractor. They include: 1) *Reliability, Maintainability, and Availability*. Reliability testing will generate data from flight testing to develop reliability parameters including: Mean Time Between Maintenance Actions, logistics reliability, and support equipment reliability. Maintainability testing will consist of qualitative and quantitative parameters such as Maintenance Man Hours per Flying Hour, Mean Time To Repair, accessibility, and ease of maintenance. Availability will be developed from test data and the R&M data. 2) *Technical Data*. Schedule the aircraft to perform verification and validation work. 3) *Training*. Evaluate contractor and Air Force developed training courses to ensure they meet their objectives. Determine if training is compatible with maintenance planning and T.O. content. Place emphasis on new technology areas. 4) *Support Equipment*. Perform functional tests to assess performance and to verify interfaces with the test item and system. Perform climatic testing and nuclear certification; and perform maintainability evaluations. 5) *Supply Support*. Use data from flight testing to verify spares calculations. 6) *Maintenance Planning*. Evaluate the adequacy and effectiveness of the maintenance planning effort to ensure the logistics elements are adequate to maintain the system and provide the level of system operational effectiveness and availability required. 7) *Software*. Assess software reliability and maintainability. 8) *Integrated Diagnostics*. Evaluate the aircraft self-diagnostic capabilities.

No detailed plan is made for these objectives. Logistics testing and evaluation of these objectives are performed using an "ad-hoc" approach.

APPROACH #4

This approach stresses an integrated team concept of Logistics Test. The DT&E testing organization teams with AFOTEC and the contractor to develop an Essential Task Listing (ETL) of items to be tested (such as engine change, towing operations, etc.), and these tasks are then prioritized according to urgency of need. These tasks are then broken down into Test Measurement Instructions (TMIs) that describe how the item is to be tested. These TMIs are categorized in three ways:

1. Simple tests that can be conducted anytime.
2. Complicated tests that require close scheduling (i.e., fuel tank tasks).
3. "Middle-ground" tests that need minimal planning.

Once the tasks are categorized, logistic testing is accomplished as opportunities arise during flight testing. In other words, the goal is not to disrupt flight testing, with the hope that logistics testing can be completely accomplished as a part of flight testing. No specific downtime is dedicated for logistics test, if possible. As testing progresses, the TMIs and ETLs are "checked-off" as they are tested and reliability and maintainability data are collected as needed.

This approach focuses on the tasks and subsystems that need to be tested, (based largely on Tech Data and Support Equipment deliverables) rather than the ILS elements. As tasks are tested, they are "linked" to an ILS element (or several) to ensure that the ILS elements are addressed. Also, since AFOTEC is involved in the testing at this point, they need to maintain "independence" by keeping their own databases and generating separate reports.

APPROACH #5

This approach uses a large team concept consisting of DT&E, AFFTC, AFMC, AFOTEC, AETC, and contractor personnel, with each organization writing their own reports and conclusions as needed. It begins by identifying what ILS elements are to be tested (and what elements are not to be tested) during DT&E. Then the major subsystems to be tested are identified. Next, Test Measurement Instructions (TMIs) are developed which specifically explain how each subsystem of the Logistics Test effort is to be evaluated. Each subsystem has several (perhaps hundreds) of TMIs, depending on the complexity of the subsystem. Included in each TMI is a matrix of how the ILS elements apply to the particular TMI. This approach gives Logistics T&E the same scheduling priority as flight testing. Therefore, aircraft downtime will be scheduled specifically for Logistic T&E purposes, and each item identified in the TMIs will have a scheduled time to be tested. As items are tested, data is collected and the items are closed out. A periodic report is generated in three parts. Part 1 gives the overall status of test resources by ILS element. Part 2 gives an assessment of all logistic elements of the system with inputs from Reliability and Maintainability, Human Factors, and items to watch for suspected deficiencies. The third part allows for feedback.

This approach stresses the need for an integrated analysis of the ILS elements and the subsystems, rather than testing a particular element or subsystem in isolation and then systematically moving on to the next element or subsystem.

APPROACH #6

This approach is the least detailed of all the approaches. It simply relies on the Test and Evaluation Master Plan and the Integrated Logistics Support Plan to outline logistics test objectives. No formal logistics test plan is developed. Logistical deficiencies, problems, and areas for improvement are noted as the flight test program evolves. If no deficiencies in a particular element or area are noted, it is assumed that the element or area is logistically acceptable.

QUESTIONNAIRE

1. What is your current rank or grade? ________________

2. How many years experience do you have in aircraft Test and Evaluation? (conducting or managing T&E programs, even at a SPO, Majcom, etc.) __________

3. How many years experience do you have in aircraft logistics? (maintenance, supply, transportation, logistics planning) __________

4. How many years experience do you have specifically with aircraft Logistics Test and Evaluation? __________

5. Where do you currently work? SPO_____ AFOTEC_____ AFFTC_____ MAJCOM_____ AFIT_______ OTHER______

6. Are you: MILITARY____ FEDERAL CIVILIAN____ CONTRACTOR____ OTHER______?

For questions 7 through 9, please use the following scale.

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree

7. A logistics test template would benefit future aircraft acquisition programs.

1 2 3 4 5

8. The results of aircraft logistics test are worth the resources allocated to the test effort.

1 2 3 4 5

9. The ten ILS elements cover what needs to be logistically tested for an aircraft program.

1 2 3 4 5

10. Rank order, with 1 being the most important, the 10 ILS elements according to the contribution each should be to the aircraft logistics test effort. Also place the percentage of total aircraft logistics test resources which should be

allocated to each element. You may use zero percent if you think an element should not be included in the test effort.

RANK	ELEMENT	PERCENTAGE
____	Maintenance Planning	________%
____	Supply Support	________%
____	Training	________%
____	Technical Data	________%
____	Manpower and Personnel	________%
____	Support Equipment	________%
____	Computer Resources	________%
____	Facilities	________%
____	Packaging, Handling, Storage, Transportation	________%
____	Design Interface	________%

11. Are there any other ILS elements, *other than* the 10 defined in DODI 5000.2 and presented in this questionnaire, that you believe should be considered in an aircraft logistics test effort?

For questions 12 through 17, please use the following scale.

1	2	3	4	5
Very Ineffective	Ineffective	Neutral	Effective	Very Effective

Please keep in mind cost, schedule, and performance constraints on current and future aircraft acquisition programs. We are interested in realistic answers, not ideal.

12. How effective is approach 1 at accomplishing aircraft logistics test?

1 2 3 4 5

13. How effective is approach 2 at accomplishing aircraft logistics test?

1 2 3 4 5

14. How effective is approach 3 at accomplishing aircraft logistics test?

1 2 3 4 5

15. How effective is approach 4 at accomplishing aircraft logistics test.

1 2 3 4 5

16. How effective is approach 5 at accomplishing aircraft logistics test?

1 2 3 4 5

17. How effective is approach 6 at accomplishing aircraft logistics test?

1 2 3 4 5

18. What are the specific components of approach 1 you liked?

Disliked?

19. What are the specific components of approach 2 you liked?

Disliked?

20. What are the specific components of approach 3 you liked?

Disliked?

21. What are the specific components of approach 4 you liked?

Disliked?

22. What are the specific components of approach 5 you liked?

Disliked?

23. What are the specific components of approach 6 you liked?

Disliked?

24. Are there any other approaches or aspects of approaches you believe were not included as part of this questionnaire but would benefit aircraft logistics test efforts?

Appendix B: Round Two Questionnaire

You should recall that we are students at AFIT and we are conducting research with the intent to develop a template for accomplishing logistics test and evaluation during the DT&E of aircraft acquisition programs. This research is sponsored by the 412 Test Wing at Edwards AFB.

You should also recall that we are using the Delphi method to find the most appropriate way to conduct logistics test and evaluation. The Delphi method uses the opinions of experts in order to evaluate the pros and cons of a particular subject. You are one of 31 experts chosen to participate in this research.

The goal of the Delphi method is to reach consensus among the experts. This may require several rounds of questions. You have already completed Round 1 in which you were asked to rate six approaches and to provide comments on what you liked and disliked about each approach.

This questionnaire represents Round 2 of our research. You will find the exact same approaches and you will be asked to rate them again. However, you will now be provided the average score for each approach (as determined by the panel of experts) and paraphrased comments (also given by the experts) on each approach. Please consider this new information when rating the approaches for this round. Simply circle your choice on each scale, and list the top three reasons (good or bad or combination) why you selected your rating. Your responses will be statistically evaluated in order to ascertain the existence of a consensus among the experts.

Please note that you may find the same "likes" and "dislikes" for a particular approach. For example, for Approach 5, some experts "liked" the idea of giving logistics testing the same priority as flight testing; but other experts "disliked" this idea. Therefore, it will be listed as both a "like" and a "dislike."

The Delphi method requires that each expert be given the same amount of time to respond. Accordingly, we request that you FAX the questionnaire only (we do not need the rest of the material) to us at DSN 986-7988 or commercial (513)476-7988, NOT LATER THAN 11 May 1994. If you have any questions about this questionnaire, please contact us immediately at DSN 785-7777, ext. 2179.

Thank you for taking the time to answer our questionnaire. We feel the results of this research will be well worth the effort.

APPROACH #1

This approach does not use a detailed logistics test plan. It begins by identifying the objectives of the logistics test effort, which include determining support equipment compatibility and training adequacy, technical order verification and validation, measuring system maintainability, and identifying hardware and software deficiencies (reliability). These objectives are tested in isolation, which means that one objective is tested first, then it is closed out, and then another objective is tested, and so on. The reliability and maintainability (R&M) testing focuses on R&M aspects of the entire system, rather than on subsystems. Any other deficiency or needed improvements will be recognized and identified as a result of flight testing; therefore, no additional planning efforts are required.

APPROACH #2

This approach is a systematic evaluation of the ten ILS elements as they pertain to three disciplines: Reliability and Maintainability (quantitative measures), Human Factors (qualitative), and Logistics Test (quantitative and qualitative). These disciplines are broken down into measurable units called Test Measurements (TM) for each element. For example, for the ILS element of Design Interface, the three disciplines are broken down as follows:

Design Interface

1. Reliability and Maintainability:
   - TM: Reliability (this might include a list of top logistic problems that need to be addressed)
   - TM: Maintainability (this might include elapsed repair times for maintenance tasks)
2. Human Factors:
   - TM: Accessibility (qualitative assessment of how easily the system is serviced and maintained)
3. Logistics Test:
   - TM: Energy Consumption (fuel, oil requirements)
   - TM: Component Standardization

Each ILS element is analyzed in a similar manner, using the same three disciplines, but with different TMs that are specific to the particular ILS element.

This approach is framed around the ten ILS elements, and once each ILS element is analyzed, then Logistics Testing can be considered complete.

APPROACH #3

This approach focuses on eight logistics test objectives aimed primarily at confirming data supplied by the contractor. They include: 1) *Reliability, Maintainability, and Availability.* Reliability testing will generate data from flight testing to develop reliability parameters including: Mean Time Between Maintenance Actions, logistics reliability, and support equipment reliability. Maintainability testing will consist of qualitative and quantitative parameters such as Maintenance Man Hours per Flying Hour, Mean Time To Repair, accessibility, and ease of maintenance. Availability will be developed from test data and the R&M data. 2) *Technical Data.* Schedule the aircraft to perform verification and validation work. 3) *Training.* Evaluate contractor and Air Force developed training courses to ensure they meet their objectives. Determine if training is compatible with maintenance planning and T.O. content. Place emphasis on new technology areas. 4) *Support Equipment.* Perform functional tests to assess performance and to verify interfaces with the test item and system. Perform climatic testing and nuclear certification; and perform maintainability evaluations. 5) *Supply Support.* Use data from flight testing to verify spares calculations. 6) *Maintenance Planning.* Evaluate the adequacy and effectiveness of the maintenance planning effort to ensure the logistics elements are adequate to maintain the system and provide the level of system operational effectiveness and availability required. 7) *Software.* Assess software reliability and maintainability. 8) *Integrated Diagnostics.* Evaluate the aircraft self-diagnostic capabilities.

No detailed plan is made for these objectives. Logistics testing and evaluation of these objectives are performed using an "ad-hoc" approach.

APPROACH #4

This approach stresses an integrated team concept of Logistics Test. The DT&E testing organization teams with AFOTEC and the contractor to develop an Essential Task Listing (ETL) of items to be tested (such as engine change, towing operations, etc.), and these tasks are then prioritized according to urgency of need. These tasks are then broken down into Test Measurement Instructions (TMIs) that describe how the item is to be tested. These TMIs are categorized in three ways:

1. Simple tests that can be conducted anytime.
2. Complicated tests that require close scheduling (i.e., fuel tank tasks).
3. "Middle-ground" tests that need minimal planning.

Once the tasks are categorized, logistic testing is accomplished as opportunities arise during flight testing. In other words, the goal is not to disrupt flight testing, with the hope that logistics testing can be completely accomplished as a part of flight testing. No specific downtime is dedicated for logistics test, if possible. As testing progresses, the TMIs and ETLs are "checked-off" as they are tested and reliability and maintainability data are collected as needed.

This approach focuses on the tasks and subsystems that need to be tested, (based largely on Tech Data and Support Equipment deliverables) rather than the ILS elements. As tasks are tested, they are "linked" to an ILS element (or several) to ensure that the ILS elements are addressed. Also, since AFOTEC is involved in the testing at this point, they need to maintain "independence" by keeping their own databases and generating separate reports.

APPROACH #5

This approach uses a large team concept consisting of DT&E, AFFTC, AFMC, AFOTEC, AETC, and contractor personnel, with each organization writing their own reports and conclusions as needed. It begins by identifying what ILS elements are to be tested (and what elements are not to be tested) during DT&E. Then the major subsystems to be tested are identified. Next, Test Measurement Instructions (TMIs) are developed which specifically explain how each subsystem of the Logistics Test effort is to be evaluated. Each subsystem has several (perhaps hundreds) of TMIs, depending on the complexity of the subsystem. Included in each TMI is a matrix of how the ILS elements apply to the particular TMI. This approach gives Logistics T&E the same scheduling priority as flight testing. Therefore, aircraft downtime will be scheduled specifically for Logistic T&E purposes, and each item identified in the TMIs will have a scheduled time to be tested. As items are tested, data is collected and the items are closed out. A periodic report is generated in three parts. Part 1 gives the overall status of test resources by ILS element. Part 2 gives an assessment of all logistic elements of the system with inputs from Reliability and Maintainability, Human Factors, and items to watch for suspected deficiencies. The third part allows for feedback.

This approach stresses the need for an <u>integrated</u> analysis of the ILS elements and the subsystems, rather than testing a particular element or subsystem in isolation and then systematically moving on to the next element or subsystem.

APPROACH #6

This approach is the least detailed of all the approaches. It simply relies on the Test and Evaluation Master Plan and the Integrated Logistics Support Plan to outline logistics test objectives. No formal logistics test plan is developed. Logistical deficiencies, problems, and areas for improvement are noted as the flight test program evolves. If no deficiencies in a particular element or area are noted, it is assumed that the element or area is logistically acceptable.

QUESTIONNAIRE

APPROACH 1

GENERAL AND PARAPHRASED LIKES OF APPROACH 1 FROM ROUND 1:

- Objectives are identified up front
- Test in isolation
- System-level approach
- Test the high points; cheap
- Emphasis placed on TOs and support equipment
- Addresses support equipment, TOs, training, and R&M
- Structured approach; Up-front planning

GENERAL AND PARAPHRASED DISLIKES OF APPROACH 1 FROM ROUND 1:

- Only goes half-way
- Identifies objectives but no plan
- Tests in isolation; need integration
- Not enough documentation
- Doesn't integrate logistics and test and evaluation
- Leaves too much to find during flight test
- Does not inter-relate test-revealed deficiencies
- Only addresses limited ILS elements
- No detailed plan; too vague
- Does not go to subsystem level
- Assumes flight test will catch all problems
- Having no plan results in delays
- Testing in series is suspect
- Focus on entire system, unrealistic. We buy spare parts, not spare planes
- Needs a planning effort for each objective to eliminate duplication of effort
- Needs to address requirements in the ORD
- Random and incomplete
- Needs to be synergistic
- No resources dedicated
- No teaming
- No definition of retest
- "Catch as catch-can"

ROUND 1 MEAN SCORE: 2.3871

1. How effective is approach 1 at accomplishing aircraft logistics test?

1	2	3	4	5
Very Ineffective	Ineffective	Neutral	Effective	Very Effective

2. State the top three reasons why you selected this rating.

APPROACH 2

GENERAL AND PARAPHRASED LIKES OF APPROACH 2 FROM ROUND 1:

- The way each element is broken down
- The way logistics test is developed around 10 ILS elements
- The test measurements process
- Getting engineering, maintenance, and human factors measures together to form a consensus on what needs to be tested
- Insures each of the 10 ILS elements is treated equally
- Systematic and measurable evaluation
- The three disciplines chosen were good choices
- Insures all areas are addressed
- Depth of analysis
- Appears to have metrics
- Simple and straightforward
- Integrates the three disciplines
- Disciplined, Unified, Iterative

GENERAL AND PARAPHRASED DISLIKES OF APPROACH 2 FROM ROUND 1:

- Could be costly
- Perhaps too rigid
- Cumbersome for large, complex, weapon systems
- Tests ILS elements, not the "real" requirements
- Analysis not sufficient to evaluate the ILS elements
- By element, not integrated
- Does not address supportability
- Too complicated
- Adds test requirements to an already full plate
- Too time consuming
- Don't know what discipline #3, logistics test, means
- Are three disciplines enough?
- May gather data of little value
- Qualitative results difficult to act upon
- Assumes testing can be done at any time
- ILS elements not analyzed as a system

ROUND 1 MEAN SCORE: 3.5484

3. How effective is approach 2 at accomplishing aircraft logistics test?

1	2	3	4	5
Very Ineffective	Ineffective	Neutral	Effective	Very Effective

4. State the top three reasons why you selected this rating.

APPROACH 3

GENERAL AND PARAPHRASED LIKES OF APPROACH 3 FROM ROUND 1:

- Attempts to define scope of testing by defining objectives
- Focuses on elements that maintenance can easily evaluate
- Planning
- Addresses practical maintenance arenas
- Training, support equipment, and maintenance planning objectives
- Confirmation of data supplied by contractor
- Comprehensive
- System-level approach
- Focus on ILS elements
- Eight areas are important to overall weapon system
- Includes everything I expect from logistics
- Depth of evaluation/ no isolation
- Uses data from other tests
- Strong in RM&A
- Uses data from flight test rather than separate logistics tests
- Structured approach
- Correct parameters to evaluate during flight test; other logistics elements can't and shouldn't be evaluated in flight test

GENERAL AND PARAPHRASED DISLIKES OF APPROACH 3 FROM ROUND 1:

- Need a plan. Ad-Hoc means it won't be done
- Too fragmented
- No details. Things can fall through the cracks
- Too expensive and too late
- Hard to determine spares requirements
- Does not address all ILS elements
- Too oriented to confirming contractor work
- Doesn't look at weapon system supportability
- Not integrated, no synergism
- Reliance on contractors' data
- Not applicable to large programs
- Data should be developed by Air Force, not contractor
- Random, incomplete, disjointed testing

ROUND 1 MEAN SCORE: 2.9032

5. How effective is approach 3 at accomplishing aircraft logistics test?

1	2	3	4	5
Very Ineffective	Ineffective	Neutral	Effective	Very Effective

6. State the top three reasons why you selected this rating.

APPROACH 4

GENERAL AND PARAPHRASED LIKES OF APPROACH 4 FROM ROUND 1:

- Integrated team
- Task listing
- Prioritization
- Linking tasks to ILS elements
- Realistic
- Useful for narrow scope, short-duration programs
- Tracking system
- May avoid duplication between DT&E and OT&E
- Provides considerable flexibility
- Considers OT&E
- AFOTEC involvement
- Seamless and invisible
- Done without interrupting flight test
- Measurable
- Focus is on tasks, not ILS elements
- Test measurement instructions could translate into actual TO procedures

GENERAL AND PARAPHRASED DISLIKES OF APPROACH 4 FROM ROUND 1:

- No scheduled events
- All ILS elements not addressed
- Separate reports and databases
- Too much room to leave things out
- Does not focus on ILS elements
- No dedicated logistics test time
- Functional managers need to be involved in development of items tested
- Not comprehensive
- Haphazard linking of Essential Task Elements to ILS elements
- Non-interference emphasis implies lack of priority
- AFOTEC involvement may compromise test integrity
- Focuses on tasks and systems instead of missions and systems
- Will it really get done?

ROUND 1 MEAN SCORE: 3.7097

7. How effective is approach 4 at accomplishing aircraft logistics test?

1	**2**	**3**	**4**	**5**
Very Ineffective	Ineffective	Neutral	Effective	Very Effective

8. State the top three reasons why you selected this rating.

APPROACH 5

GENERAL AND PARAPHRASED LIKES OF APPROACH 5 FROM ROUND 1:

- Integrated and large team approach
- Top-down determination of TMIs
- Uses requirements verification matrix
- Equal priority with flight testing
- System approach
- Separate reports
- Integrated analysis of ILS elements
- Systematic and detailed approach
- Dedicated downtime for logistics testing
- Dedicated resources
- Comprehensive and thorough
- Better detailed plan
- Well-structured matrix
- Measurable
- Progress of testing well-documented

GENERAL AND PARAPHRASED DISLIKES OF APPROACH 5 FROM ROUND 1:

- Duplication of effort among AFOTEC and DT&E reporting
- Cost/schedule prohibitive
- Appears to somewhat ungainly
- Building a consensus with so many people involved is next to impossible
- Large test teams are not likely to be approved in a period of a draw down
- Same priority as flight test
- Focuses on subsystems rather than system level
- Dedicated downtime not always necessary
- Requires too much manpower

ROUND 1 MEAN SCORE: 4.0968

9. How effective is approach 5 at accomplishing aircraft logistics test?

1	2	3	4	5
Very Ineffective	Ineffective	Neutral	Effective	Very Effective

10. State the top three reasons why you selected this rating.

APPROACH 6

GENERAL AND PARAPHRASED LIKES OF APPROACH 6 FROM ROUND 1:

- Uses the TEMP
- Inexpensive
- Simple
- Realistic
- Lots of freedom
- Best bet for small, unstructured programs

GENERAL AND PARAPHRASED DISLIKES OF APPROACH 6 FROM ROUND 1:

- No firm, structured program
- Lacks detail
- Does not help further develop TEMP
- It is a bad assumption that "no noted deficiencies" means logistically acceptable
- Does not ensure system is fully tested
- Cavalier attitude
- Too open and too loose
- Can't get there from here
- No schedule
- Not at all effective
- No methodology
- Won't work

ROUND 1 MEAN SCORE: 1.9355

11. How effective is approach 6 at accomplishing aircraft logistics test?

1	**2**	**3**	**4**	**5**
Very Ineffective	Ineffective	Neutral	Effective	Very Effective

12. State the top three reasons why you selected this rating.

Appendix C: Summarized Comments from Question 24, Round One

- Need to look at the ORD
- Need to address a consistent and comprehensive database
- Defined/documented objectives
- Specific, measurable criteria
- Well defined team approach
- Clear/detailed documentation
- System level view with T&E, starting at component/subsystem level
- Combination of approaches 4 & 5
- All organizations must sign up early to the chosen approach
- More emphasis on early AFOTEC involvement
- Contractor should understand and agree on level of testing
- During logistics test planning, test aircraft configurations should be evaluated and tasks prioritized
- Include participation of the user
- Break down the ILS elements into their subelements and determine which can be tested
- Early identification of design problems
- Maintainability demonstrations are scheduled events
- Maintenance and Engineering must work as a team
- Be realistic about dedicated ground testing
- Attempt to define scope of testing and freeze new inputs as much as possible
- Must be approached as an embedded part of the acquisition process
- Trade-offs must be made in order to come up with an affordable plan customized to the scale of the complexity of the system/modification being tested

Appendix D: Top Three Reasons for Experts' Ratings, Round Two

APPROACH 1

Comments	Frequency
- No detailed plan.	13
- Tests in isolation.	12
- Subsystems need to be tested.	9
- No integration.	6
- Needs to address requirements in ORD.	5
- Finding logistics problems is left to chance.	5
- Does not interlock logistics and T&E.	3
- Testing in series.	3
- Does not look at all logistics elements.	3
- Does not inter-relate test-revealed deficiencies.	3
- Objectives are identified up front.	2
- Needs team approach.	2
- Random and incomplete.	2
- Events not scheduled or budgeted.	1
- No criteria mentioned.	1
- Not schedule driven.	1
- Catch as catch can.	1

APPROACH 2

Comments	Frequency
- Addresses all ILS elements.	14
- Systematic and measurable evaluation.	9
- Does not address supportability.	6
- Cumbersome for large, complex weapon systems.	5
- Disciplined, unified, iterative.	4
- Analysis not sufficient to evaluate ILS elements.	4
- Too complicated.	3
- Depth of analysis.	3
- Integrates teams.	2
- Simple and straightforward	2
- Not integrated.	2
- Too time consuming.	1
- Could be costly.	1
- Schedule with scope defined.	1
- Has metrics.	1
- No specific interface testing.	1

APPROACH 3

Comments	Frequency
- No detailed logistics test plan	13
- Does not address sustained supportability	7
- Does not address all ILS elements	6
- Ad Hoc	5

- Uses flight data to verify/confirm	5
- Lack of integration	5
- Focus on ILS elements	4
- Focuses too much on verifying contractor data	3
- Eight areas are important to overall weapon system	2
- Structured approach	2
- Random, incomplete, disjointed testing	2
- Data should be developed by A.F., not contractor	2
- Not structured enough	1
- Not sure it is an integrated approach to logistics	1
- Focus on meaningful factors (MTBF,MTTR) to determine system performance	1
- No structured re-testing	1
- Has clear objectives	1
- No team approach	1
- Too fragmented	1
- Lacks a tracking system	1
- Good RM&A approach	1
- Depth of evaluation	1
- Needs to apply to all programs	1
- No early involvement to influence design	1
- No operational input	1
- Not applicable to large programs	1
- System level approach	1
- Comprehensive	1

APPROACH 4

Comments	Frequency
- Integrated teaming	11
- Prioritization	7
- No scheduled events	6
- Provides considerable flexibility	5
- Non-interference emphasis implies lack of priority	4
- No dedicated logistics test time	4
- Linking of tasks to ILS elements	4
- All ILS elements not addressed	4
- Measurable	2
- Plan is developed	2
- TMIs could translate into actual TO procedures	2
- Haphazard linking of ETEs to ILS elements	2
- Will it really get done?	1
- Tracking system	1
- Integrated with flight testing	1
- Treats to weapon system the way the user will	1
- Does not interrupt flight testing	1
- Incomplete testing	1
- Not a good schedule of events	1
- Not comprehensive	1
- Balance between AFOTEC involvement and logistics test	1
- Useful for narrow-scope, short-duration programs	1
- No specific integration testing	1

- Logistics testing is accomplished as opportunities arise- won't work	1
- This would be a good method if logistics were tested as approach 3	1
- Whole system and subsystem need testing, not just what the deliverables require	1
- Focus on tasks and systems, instead of missions and systems	1
- No specifics in tasks	1
- Leaves too much to chance	1
- No detailed approach to looking at integrated logistics	1

APPROACH 5

Comments	Frequency
- Equal priority as flight testing	11
- Integrated team approach	9
- Comprehensive test of each ILS element	8
- Systematic approach	6
- User requirement verification matrix	5
- Dedicated resources	3
- A comprehensive plan	3
- Dedicated downtime for logistics testing	3
- Data collection/periodic reports	3
- Specifically tests integration of system	3
- May be somewhat unrealistic in today's "reduced budget and manpower" environment	3
- Has objective and metrics	2
- Building a consensus with so many people is next to impossible	2
- Top-down determination of TMIs	2
- Cross-functional approach	1
- Need to downsize team for this method to be "very effective"	1
- Dedicated downtime not always needed	1
- Dependent upon size of program	1
- May be too rigid for large programs	1

APPROACH 6

Comments	Frequency
- It is a bad assumption that "no noted deficiencies" means logistically acceptable	12
- No firm, structured program	12
- No schedule/too loose	9
- Lack detail	8
- Does not ensure system is fully tested	8
- No methodology	4
- Won't work	3
- Uses the TEMP	2
- Not at all effective	2

- May be realistic for a small program	1
- needs more detail	1
- Uses the flight test data	1
- Very costly retrofit program will be required	1
- Not possible to complete tech data verification and validation and support equipment testing using this approach	1
- Makes too many assumptions	1
- OK for equipment acquisition if "off-the-shelf"	1
- Not effective for major acquisition	1
- Cavalier attitude	1

Appendix E: Summary of Wilcoxon Signed Rank Test for Differences between Approaches Using Statistix 4.0

Approaches	P-Value	α	Reject Ho?
F-16 vs. C-17	0.00	0.1	Yes
F-16 vs. B-1	0.00	0.1	Yes
F-16 vs. B-2	0.00	0.1	Yes
F-16 vs. F-22	0.00	0.1	Yes
F-16 vs. Traditional	0.00	0.1	Yes
C-17 vs. B-1	0.00	0.1	Yes
C-17 vs. B-2	0.35	0.1	No
C-17 vs. F-22	0.00	0.1	Yes
C-17 vs. Traditional	0.00	0.1	Yes
B-1 vs. B-2	0.00	0.1	Yes
B-1 vs. F-22	0.00	0.1	Yes
B-1 vs. Traditional	0.00	0.1	Yes
B-2 vs. F-22	0.03	0.1	Yes
B-2 vs. Traditional	0.00	0.1	Yes
F-22 vs. Traditional	0.00	0.1	Yes

Approaches: Comparisons of all of the possible combinations of approaches to determine if differences existed between their respective effectiveness probability distributions.

P-Value: The P-Value generated by Statistix 4.0 after we input effectiveness data from round 2 questionnaires and performed the Wilcoxon Signed Rank Test (26:111)

α: Significance level selected for this test.

Null Hypothesis (Ho): The probability distributions of the effectiveness of the two approaches being tested were identical.

Reject Ho?: If the P-value was less than the significance level selected (0.1), we rejected the null hypothesis in favor of the alternative which stated that the probability distributions of the effectiveness of the two approaches being tested were not identical.

Conclusion: The only two approaches which had identical effectiveness probability distributions were the C-17 and the B-2. The experts believed there was no difference in the effectiveness of these two approaches.

Appendix F: Sample of a Systematic and Measurable Method to Evaluate the ILS Elements

The following table contains the ten ILS elements and three subelements (R&M, Human Factors, and Logistics Test) for each element. Within the subelements are measurable units used to evaluate the particular element. Measurements will be taken using surveys to record both qualitative and quantitative information (1:3-11).

Maintenance Planning

R&M: Mean Time Between Maintenance, corrective (MTBMc)
Mean Time Between Maintenance, inherent (MTBMi)
Maintenance Man-Hour Per Flying Hour (MMH/FH)
Mean Man-Hours to Repair (MMTR)

HF: None

LT: Scope of inspection/task
Frequency of inspection/task
Time to perform a given task

Manpower & Personnel

R&M: MMH/FH
Average crew size to perform each task
Total maintenance man-hours for each maintenance action
Air Force Specialty Codes (AFSCs) required for each maintenance action

HF: Human performance during all weather activities

LT: Actual crew size to perform each maintenance action

Support Equipment

R&M: None

HF: Ease of use
Handling
Safety
Compatibility of support equipment and aircraft

LT: Supportability
Utilization rate of support equipment

Supply Support

R&M: Mean Time Between Removal (MTBR) of any repairable or non-repairable equipment
MTBMi

HF: None

LT: Assignment of Source, Maintenance, and Recoverability (SMR) codes
Availability of parts, consumables, and support equipment

Technical Data

R&M: None

HF: Safety
Adequacy
Clarity of instructions

LT: Understandability
Ease of use

Training and Training Support

R&M: None

HF: None

LT: Knowledge training required
Proficiency training required
Safety

Computer Resources

R&M: Hardware and software reliability
Hardware and software maintainability
Built-In-Test (BIT) system adequacy

HF: Ease of use

LT: Quantity of personnel required to operate the computer resources
Functional utility of the computer resources

Facilities

R&M: None

HF: None

LT: Supportability of facilities
Compatibility of facilities with the aircraft and support equipment

Packaging, Handling, Storage, and Transportation

R&M: None

HF: None

LT: Suitability
Safety

Design Interface

R&M: Reliability
Maintainability

HF: Interoperability
Accessibility

LT: Energy consumption
Accessibility
Component preservation
Component standardization

Bibliography

1. Aeronautical Systems Center, Air Force Materiel Command. Contract F33657-91-C-0006 (draft) with Lockheed and Boeing Corporation. Wright-Patterson AFB OH, 15 February 1994.

2. Andrews, Richard A. "The Basics of Integrated Logistics Support," Essay in Course Reader, SMGT 643, System Acquisition Management. School of Systems and Logistics, Air Force Institute of Technology, Wright-Patterson AFB OH, July 1993.

3. Blanchard, Benjamin S. Logistics Engineering and Management. Englewood NJ: Prentice Hall Inc., 1992.

4. Bower, James N. Reliability and Maintainability Evaluation of the F-16C/D Aircraft. Air Force Flight Test Center, Edwards AFB CA, July 1985.

5. Brown, Bernice B. "Delphi Process: A Methodology Used for the Elicitation of Opinions of Experts." Rand Report No. P-3925, The Rand Corporation, Santa Monica CA, September 1968.

6. Dalkey, Norman C. "Delphi." Rand Report No. P-3704, The Rand Corporation, Santa Monica CA, October 1967.

7. -----. "Experiments in Group Prediction." Rand Report No. P-3820, The Rand Corporation, Santa Monica CA, March 1968.

8. -----. "The Delphi Method: An Experimental Study of Group Opinion." Rand Report No. RM-5888-PR, The Rand Corporation, Santa Monica CA, March 1968.

9. Department of Defense. Defense Acquisition Management Policies and Procedures. DOD Instruction 5000.2. Washington: GPO, 23 February 1991.

10. Department of the Air Force. Air Force Logistics Support Analysis Primer. AFLCP 800-17. Washington: HQ USAF, 20 May 1988.

11. Department of the Air Force. B-1 Test and Evaluation Master Plan. Air Force Systems Command. Wright-Patterson AFB OH, 26 May 1982.

12. Department of the Air Force. Basic Aerospace Doctrine of the United States Air Force. AFM 1-1. Washington: HQ USAF, March 1992.

13. Department of the Air Force. F-16 Integrated Logistics Support Plan. Air Force Systems Command. Wright-Patterson AFB OH, February 1991.

14. Downie, N.M. Fundamentals of Measurement: Techniques and Practices. New York: Oxford University Press, 1967.

15. Emory, C. William and Cooper, Donald R. Business Research Methods. Homewood IL: Irwin, Inc., 1991.

16. Gullet, David. F-16 System Program Office, Wright-Patterson AFB OH, Personal Interview, 26 January 1994.

17. Lis, James. "Test and Evaluation," in Military Project Management Handbook. Ed. David Cleland and others. New York: The McGraw-Hill Book Company, 1993.

18. McBrayer, C. Air Force Flight Test Center. Edwards AFB CA. Telephone interview, 4 February 1994.

19. McCarthy, Herbert W. "Logistics," in Military Project Management Handbook. Ed. David Cleland and others. New York: The McGraw-Hill Book Company, 1993.

20. McClave, James T. and Benson, George P. Statistics for Business and Economics. San Francisco: Macmillan Publishing Company, 1991.

21. Parker, Mary H. Chief, Logistics Test Flight, 412 Test Wing, Edwards AFB CA. Personal Correspondence. 24 August 1993.

22. -----. Edwards AFB CA. Personal Correspondence. 5 October 1993.

23. Regan, D. Maintenance/Logistics Test Officer, 412 Test Wing, Edwards AFB CA. Telephone interview, 24 March 1994.

24. Rugama, Jose A. and others. "C-17 Logistics Test and Evaluation Data Analysis Plan." Developed for the Air Force Flight Test Center Combined Test Force, Edwards AFB CA. Not Dated.

25. Sackman, H. Delphi Assessment: "Expert Opinion, Forecasting, and Group Process." Rand Report No. R-1283-PR, The Rand Corporation, Santa Monica CA, April 1974.

26. Siegel, Joan. Statistix 4.0. Tallahassee: Analytical Software, 1992.

27. Sudman, Semour and Bradburn, Norman. Asking Questions. San Francisco: Jossey-Bass Inc., 1983.

Vita

Captain Timothy J. Fennell was born 8 December 1964 at Quantico Marine Corps Base, Virginia. He graduated from Havelock High School, North Carolina, in 1983. Upon graduation, he was awarded a four-year Air Force ROTC scholarship to attend North Carolina State University. He graduated in 1987 with a Bachelor of Science degree in Aerospace Engineering and was commissioned as a second lieutenant in the United States Air Force. Captain Fennell's first assignment was at Chanute Air Force Base, Illinois, where he attended the Aircraft Maintenance Officer Course. He then was assigned to the 4950th Test Wing at Wright-Patterson AFB, Ohio, where he served as the Propulsion Branch Chief, the Flightline Branch Chief, and finally as the Training/Standardization Branch Chief. He has maintenance experience on C-135, C-141, C-18, T-39, and T-37 aircraft. He has attended SOS in residence and was selected to attend the Air Force Institute of Technology in 1993 to obtain a Master of Science degree in Maintenance Management. Captain Fennell has one daughter, Angela Lauren, who is two years old.

Permanent Address: 306 Beechwood Drive
Havelock, NC 28532

Vita

Captain Jason S. Mogle was born on 9 September 1963 in Tiffin, Ohio. He graduated from Columbian High School in 1981. Following high school, he attended Heidelberg College where he graduated in 1985 with a Bachelor of Arts degree in English. He entered Officer Training School shortly after graduation and was commissioned on 13 September 1985. His maintenance career began at the Aircraft Maintenance Officer Course at Chanute AFB, Illinois after which he was assigned to the 82nd Flying Training Wing, Williams AFB, Arizona. There he served in the Field and Organizational Maintenance Squadrons, as well as on the DCM staff maintaining T-37 and T-38 aircraft. In December of 1989, he was assigned to the 4950th Test Wing, Wright-Patterson AFB, Ohio where he served as the Maintenance Supervisor of the Field Maintenance Squadron and subsequently as the Officer-In-Charge of the 4952nd Aircraft Maintenance Unit, maintaining C-135 and C-18 aircraft. In 1993, he was selected to attend the Air Force Institute of Technology to study Maintenance Management. Captain Mogle is married to the former Melissa Massie of Springfield, Ohio. They have one daughter, Danielle, who is two years old.

Permanent Address: 519 East State Route 18
Tiffin, Ohio 44883

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
September 1994

3. REPORT TYPE AND DATES COVERED
Master's Thesis

4. TITLE AND SUBTITLE
DEVELOPING A TEMPLATE FOR LOGISTICS TEST AND EVALUATION

5. FUNDING NUMBERS

6. AUTHOR(S)
Timothy J. Fennell, Captain, USAF
Jason S. Mogle, Captain, USAF

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology,
WPAFB, OH 45433-6583

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/GLM/LAL/94S-17

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
412 TW/LGLL
300 East Mojave Blvd.
Edwards AFB, CA 93524-6325

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
Logistics test and evaluation, performed during the developmental test and evaluation effort, is an integral part of the aircraft acquisition process. However, there has been no standard approach to conducting logistics test and evaluation. This study researched past and present approaches to aircraft logistics test and evaluation to determine the most effective method for future programs. We conducted this study using the Delphi method. We solicited the expert opinions of 32 individuals from the logistics test and evaluation field using two rounds of questions. We then statistically analyzed the data to ultimately develop a logistics test and evaluation template.
We concluded that the ten ILS elements are indeed a valid baseline for a logistics test and evaluation template, that logistics test and evaluation is worth the resources allocated to it, and that the F-22 logistics test and evaluation approach is the most effective.

14. SUBJECT TERMS
Logistics test and evaluation, Integrated Logistics Support (ILS) elements, Developmental Test and Evaluation

15. NUMBER OF PAGES
118

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL