DOT/FAA/AR-xx/xx

Office of Aviation Research
Washington, D.C. 20591

Final Report: A Comparison of Three Evaluative Techniques for Validating Maintenance Documentation

February 2004

Final Report

This document is available to the U.S. public through the National Technical Information Service (NTIS), Springfield, Virginia 22161.

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents or use thereof. The United States Government does not endorse products or manufacturers. Trade or manufacturer's names appear herein solely because they are considered essential to the objective of this report.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AR-xx/xx		

4. Title and Subtitle	5. Report Date
Final Report: A Comparison of Three Evaluative Techniques for Validating Maintenance Documentation	
	6. Performing Organization Code
7. Author(s) Alex Chaparro, Bonnie Rogers, Chris Hamblin, Barbara Chaparro	8. Performing Organization Report No.
9. Performing Organization Name and Address National Institute for Aviation Research Wichita State University 1845 Fairmount Wichita, KS 67260	10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address U.S. Department of Transportation Federal Aviation Administration Office of Aviation Research Washington, DC 20591	13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code AFS-300

15. Supplementary Notes

The FAA William J. Hughes Technical Center Monitor was Michael Vu.

16. Abstract

Surveys and interviews reveal that general aviation manufacturers rely on user to identify problems in maintenance documentation with corrections typically initiated in response to users-reported problems found in the manual. This investigation compares techniques that manufacturers can use to improve the quality of the maintenance documentation developed by technical writing groups. The techniques, User Performance and Cognitive Walkthrough, were used to identify problems in aircraft maintenance documentation. The techniques high light probable mismatches between the intent of the technical writer and the interpretation of the maintenance procedures by aircraft maintenance technicians (AMTs). The purpose of this investigation was to determine whether these techniques eliciting unique information from participants, to establish whether different groups of evaluators (engineers, technical writers, AMTs) identify different types of problems in the written documentation, and to identify common types of problems in the documentation. We also reviewed the heuristics technical writers use to develop maintenance documentation. The results of this investigation show that User Performance and Cognitive Walkthrough evaluations are complementary techniques for identifying problems with maintenance documentation, that the errors identified by individual participants vary in significant ways according to experience and training (engineers versus AMTs), and that procedure errors (e.g., sequencing, completeness) and language errors (e.g., clarity, meaning) are the most commonly cited errors in the maintenance documentation. Ways in which manufactures can employ the evaluated techniques in their manual development process are discussed.

17. Key Words	18. Distribution Statement
Error, Human Factors, Maintenance Manuals	This document is available to the public through the National Technical Information Service (NTIS), Springfield, Virginia 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	16	

Form DOT F1700.7 (8-72) Reproduction of completed page authorized

ACKNOWLEDGEMENT

We would like to acknowledge the contribution and commitment of the following individuals and their organizations to this effort: Daniel J. McPartland and Rex Rathegeber (Cessna Aircraft Company), Jack Kramer, Gary Bolivier, and Jeff Coyendall (Learjet/Bombardier Aerospace), Terry Clark and Ron Jaerger (Raytheon Aircraft Company). We would also like to thank our FAA technical monitor Michael Vu for his support and assistance in completing the project.

TABLE OF CONTENTS

		Page
EXECUTIVE SUMMARY		9
1.	INTRODUCTION	10
1.1.	Background	11
1.2	A Conceptual Framework for Maintenance	12
1.3	Evaluation of Maintenance Documentation	14
1.4	Purpose	15
2.	METHODOLOGY	15
2.1.	Interviews with Technical Writers	15
2.1.1.	Background of Technical Writers	15
2.1.2.	Manual Development Process and Heuristics	16
2.2	Selection and Description of Evaluation Techniques	16
2.2.1	Materials	17
2.2.2	Participants	17
2.2.3	General Procedures	18
3.	RESULTS	19
3.1	Participant Demographics	19
3.2	Error Taxonomy	19
3.3	Cognitive Walkthrough Method	20
3.3.1	Participants	20
3.4	User Performance Evaluation Methods	21
3.5	Comparison of Evaluation Methods	22
3.5.1	Errors Reported in Evaluation Methods	22
3.5.2	Unique Identified by Different Evaluator Groups	24
3.5.3	Analysis of Procedure and Language Errors	24
3.6	Satisfaction Measures	25
4.	DISCUSSION	27
4.1	Recommended Best Practices	29
4.2	Integrating Evaluation Techniques in the Technical Writing Process	30

4.3 Technical Writers' Training 31

5. FUTURE RESEARCH 31

6. REFERENCES 33

7. APPENDICES 34

A. User Satisfaction Questionnaire 34
B. User Background Questionnaire 36

LIST OF TABLES

Table		Page
1	Participant Demographics	19
2	Number of Errors Reported in CW by Evaluator Group	20
3	Number of Errors Reported in UP Evaluation by Evaluator Group and Method	21
4	Unique Errors Found as a Function of Evaluation Method	24
5	Number of Language Error Subcategories Reported by Evaluation Method	24
6	Number of Procedure Error Subcategories Reported by Evaluation Method	24
7	Results of Satisfaction Measures between CW and UP Evaluation Methods on a Five-Point Scale.	26
8	Mean satisfaction values as a function of user group and evaluation method.	26

LIST OF FIGURES

Figure		Page
1	Norman's (1988) Action Cycle	13
2	Aircraft Maintenance Manual Development Process	16
3	Error Frequencies as a Function of Evaluation Method and Error Type	23
4	Percent of Errors Reported by Evaluator Background and Experience	23
5	Comparison of the Corrective Actions by Evaluation Method	25
6	User-Centered Iterative Design in the Aircraft Manual Development Process	30

EXECUTIVE SUMMARY

Surveys and interviews reveal that general aviation manufacturers rely on user to identify problems in maintenance documentation with corrections typically initiated in response to users-reported problems found in the manual. This investigation compares techniques that manufacturers can use to improve the quality of the maintenance documentation developed by technical writing groups. The techniques, User Performance and Cognitive Walkthrough, were used to identify problems in aircraft maintenance documentation. The techniques high light probable mismatches between the intent of the technical writer and the interpretation of the maintenance procedures by aircraft maintenance technicians (AMTs). The purpose of this investigation was to determine whether these techniques eliciting unique information from participants, to establish whether different groups of evaluators (engineers, technical writers, AMTs) identify different types of problems in the written documentation, and to identify common types of problems in the documentation. We also reviewed the heuristics technical writers use to develop maintenance documentation. The results of this investigation show that User Performance and Cognitive Walkthrough evaluations are complementary techniques for identifying problems with maintenance documentation, that the errors identified by individual participants vary in significant ways according to experience and training (engineers versus AMTs), and that procedure errors (e.g., sequencing, completeness) and language errors (e.g., clarity, meaning) are the most commonly cited errors in the maintenance documentation. Ways in which manufactures can employ the evaluated techniques in their manual development process are discussed.

1. INTRODUCTION.

Since the 1960s commercial aviation has seen a dramatic decline in accident rates. This decline may be attributable to several factors including improved aircraft design, improved component reliability and pilot training initiatives (Line Oriented Flight Training, Crew Resource Management). However, aircraft accident rates have changed little over the last decade attention has turned to the identification of other factors that contribute to aircraft accidents. One factor that has received increased scrutiny is aircraft maintenance. Estimates of the number of major aircraft accidents that can be attributed to maintenance vary considerably. For instance, in 1994 Marx and Graber [1] estimated that 12% of major aircraft accidents were related to maintenance; in contrast a summary of commercial airplane accidents between 1959-1999 estimated this number to be ≈6% [2]. More recently, John Goglia, member of the National Transportation Safety Board has argued that informal analyses of accidents occurring between 1995 and 2000 suggest this number may now be as high as 40% [3]. For comparison, analyses of Naval Aviation Class A mishaps (e.g., those involving loss of an aircraft or fatality) reveals maintenance as a causal factor in 17% of accidents [4].

Maintenance procedures have been cited as primary factors contributing to maintenance errors [5-8]. A review of Naval Aviation Maintenance mishaps that occurred between 1990 and 2003 [4] showed that 28% of the accidents involved problems in maintenance procedures including missing procedural steps, incorrect sequence of steps, inadequate procedures for inspection and troubleshooting, incorrect technical information, and incorrect diagrams. Similar data from commercial aviation shows that problems with procedures were cited in 11.4% of critical incidents [5]. Critical incidents were defined as "*events that could have prevented an aircraft from operating normally, or could have put the safety of anyone at risk*" (p. 200). Maintenance procedures were also cited as contributing factors in NTSB reports for two recent aircraft accidents [9, 10]. One should use caution in relying on Class A accidents or critical incidents to draw conclusions about the impact and frequency of documentation related maintenance mishaps. Mishaps are rare events, so they will underestimate the frequency of incidents in which poor documentation resulted in maintenance errors. In addition, mishaps do not account for the other effects of poor documentation including the costs of incorrectly executed or slowed maintenance.

Maintenance documentation has recently begun to receive attention from academic researchers, the Federal Aviation Administration, and manufacturers. For instance, the methods and techniques employed by the aviation industry to develop maintenance documentation were only documented three years ago [11]. This research identified a number of problems with the development of maintenance documentation. This included reactive rather than proactive evaluations of the manual, the limited use of aircraft maintenance technicians' (AMTs') input and procedure validation, the absence of systematic attempts to track error, and the lack of standards for measuring document quality. Some of the participating aircraft manufacturers have modified their manual development practices; however, these efforts have been informal and have not taken advantage of techniques and methods developed in other industries.

1.1 Background.

Chaparro and colleagues [11, 12] reported that the development and revision of technical maintenance documentation is similar across aviation manufacturers. Usually, the writer works from engineering drawings and vendor specifications using a predetermined formatting style or template adopted by a manufacturer. Writers often rely on each other or customer service personnel (customer technical support or field service) to proofread a draft maintenance procedure. Input from the user (AMTs) population is not usually sought.

Corrections to maintenance documentation are made post release through reports of customer problems, called Publication Change Requests (PCRs), or Engineering Change Requests (ECRs). PCRs or "squawks" are submitted by customer technical support personnel based on customer calls from the field. Even then, unless the maintenance procedure is judged to impact safety of flight, it will not be validated. In these extreme cases, the procedure is performed by an AMT on the aircraft, corrections are made to the manual, and the revision is distributed to the customer through a Service Bulletin. This reactive process relies heavily on the end user to find problems in the documentation after publication. The assumption that users will report errors in maintenance procedures may be incorrect. Chaparro, et al. [12] found that 53% of aviation mechanics reported only occasionally, rarely, or never reporting errors found in aircraft documentation.

The "correctness" of the document design is not the only factor that assures its reliability or usability. The accurate and clear communication of information is also critical. In other words, the AMT's interpretation of the procedure must match the intent of the writer. A mismatch has two likely outcomes. First, the AMT may become frustrated and call customer support for assistance in performing a procedure, or second, the AMT may "work-around" the procedure. This entails trying to deduce the writers' intent when a procedure is confusing, or the information is incomplete or inaccurate. This is not an uncommon occurrence.

In their survey, Chaparro, et al. [12] found that 64% of AMTs reported finding their own way of performing a procedure. Nearly 60% of aircraft maintenance personnel also reported continuation of an unfamiliar task despite not being sure if they were performing it correctly [13]. Similarly, McDonald et al. [8] reported that 34% of routine maintenance tasks are performed in ways different than outlined in the maintenance procedure. AMTs are often very good at deriving a plausible interpretation of incomplete information by drawing on their `knowledge and that of other mechanics [14]. This very ability may result in an AMT misinterpreting procedures in such a manner that it is difficult to discover the error in their interpretation and subsequent actions. Although the AMTs' training and experience may allow them to correctly identify the writers intent, this will not always be the case. This uncertainty can be reduced by the proactive approach of assessing documentation quality before publication using tools originally developed to test the usability of computing software programs.

Inaccurate or unreliable documentation can have a broad negative impact. The costs to the manufacturer include technical support personnel's time, writers' time to revise the manual based on publication change requests (PCRs) and costs of printing revisions to the manual. The cost to the operator includes increased downtime, damage to components, and the potential loss of an aircraft and lives. There can be considerable pressure on AMTs to minimize aircraft down time and return it to revenue service. Under such circumstances, mechanics may be more inclined to "work around" (i.e., attempt to identify the writer's intent) unclear procedures. Mechanics report that such violations of recommended practice are partly motivated by unclear or inefficient procedures [8]. However, cases where the AMT "works around" the procedure are the most problematic and may be related to aviation accidents that were attributed to maintenance error (see [10]).

1.2 A Conceptual Framework for Maintenance.

Donald Norman's action cycle [15] provides a conceptual framework for thinking about how maintenance information can facilitate or impede performance of a maintenance task (see Figure 1). The performance of maintenance task includes determining the identification of the task goal, identifying appropriate set of actions given the current state of the system, executing the actions and evaluating the outcome of the actions. The technicians' behavior reflects two classes of action: execution and evaluation shown on the left and right sides of Figure 1.

Consider the case of replacing an aircraft tire. The technician is given paperwork that specifies a goal (i.e., replace tire). Before maintenance can begin, the technician must identify the set of specific actions that must be executed. Some of this these actions are specified in the maintenance manual; however, the manual does not outline the maintenance actions in minute detail. Rather, the manual specifies more intermediate level goals (i.e., remove securing lock pin). The AMT relies on several information sources including diagrams or schematics, their prior experience or what we refer to as Knowledge In (the) Head (referred to hereforth as KIH), and information available in the environment or Knowledge In (the) World (i.e., KIW) to identify the specific set of actions to be executed. Sometimes the sequence of maintenance actions is indicated by the design itself. For instance, by looking at the tire, the maintenance technician recognizes that a safety lock/pin, washers and bearings must be removed first before the tire can be removed. Likewise, during reassembly, the different sized screws, or shape or color coding of electrical connectors can facilitate maintenance by effectively telling the mechanic which ones to connect. These types of *physical constraints* are examples of what Norman called KIW. Putting knowledge in the environment (i.e., world) can facilitate execution of maintenance tasks by serving as memory aids (labels, signal, warnings, etc.), guiding assembly and disassembly (some parts will only fit into certain holes) and reducing the information the maintenance technician must remember.

Once the action has been performed, the AMT evaluates whether it has produced the expected outcome. For example, is the tire seated properly, is it aligned correctly, does it move freely? This part of the cycle involves comparing what they now see in their environment (KIW) to what they think they should see (KIH).

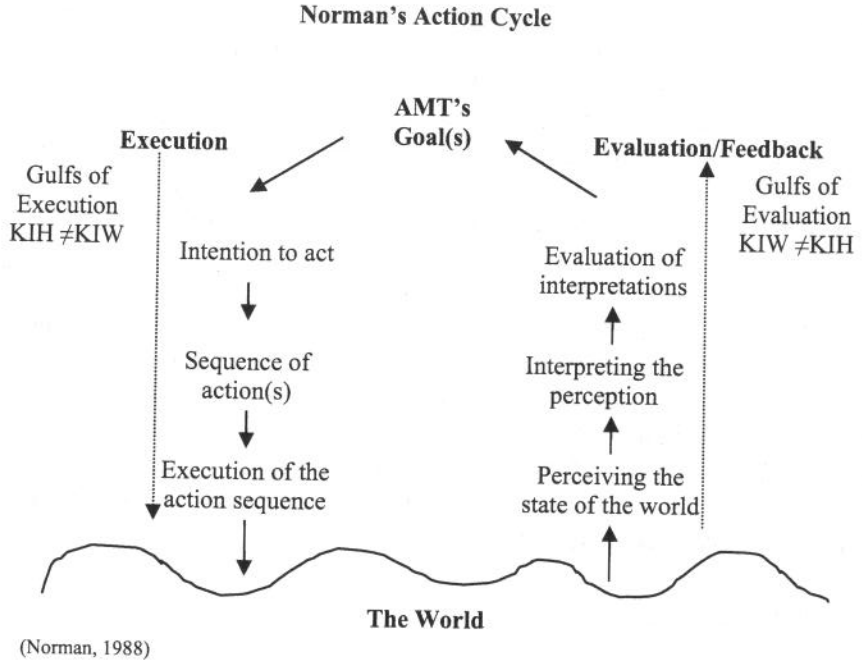

Figure 1. The Action Cycle. Adapted from Norman, 1988.

In any one of the six steps of execution and evaluation, the AMT may encounter difficulties in moving toward the goal, which are termed 'gulfs'. Gulfs of execution and evaluation are said to exist when the AMT has difficulty translating goals into action due to a) difficulty identifying which actions will allow them to achieve their goal (Gulf of execution) and b) difficulty identifying the system state (Gulf of Evaluation) due to poor or absent feedback. The execution of procedures can also be hindered by using unfamiliar part references. In this case the AMT must rely on their prior knowledge (i.e., KIH), technical drawings, and expertise of others to identify the part thus bridging the gulf of execution. Failure of this process can result in the wrong part being removed perhaps causing damage to other parts and/or systems.

Maintenance technicians rely on a variety of cues to identify the current state of the system. When replacing the tire they may use formal cues including hydraulic pressure readings, indicator lights, etc. to determine whether the system is safe for maintenance. Other cues may be provided by "notes" and "warnings" or maintenance "hints" included in some manuals. They also depend on other less formal cues provided by touch and vision to judge the alignment of parts or torque values. The systems state may be difficult to judge in the absence of such cues or feedback. The correct execution of a long maintenance procedure on a complex system, like an engine, may only be revealed through the feedback provided by functional tests. The functional test results (or KIW) supplement KIH narrowing the gulf of evaluation allowing the user to determine if further actions are necessary.

The maintenance manual is a necessary aid to bridge the gulfs of execution and evaluation in performing lengthy, new or rarely performed maintenance tasks. Gulfs in execution and evaluation may arise due to differences in the way writers and engineers communicate the procedure in the documentation. Writers and engineers must rely on their writing experience, terminology, knowledge of the users, and knowledge of the aircraft systems in deciding how best to describe a procedure. Unfortunately, the experiences, terminologies, and training of the writer, engineer, and the AMT are all different. These differences may explain why mismatches occur between the information provided by the writer and the interpretation by the AMT.

Narrowing the gulfs or the "mismatches" is the principle reason for evaluating the usability of the document throughout the development process. During the writing process decisions are made regarding what information to include, the sequence of steps to perform, and the level of detail to include. These decisions are based on information and assumptions regarding the skills and knowledge of the typical AMT. Evaluation techniques allow the technical writer to determine if their decisions were correct. The optimal result is a manual that the AMT perceives as a useful resource.

1.3 Evaluation of Maintenance Documentation

Several of the aircraft manufacturers who participated in the original survey [11] have initiated limited efforts at testing some maintenance procedures. The selection of maintenance procedures for evaluation is based on their potential impact on aircraft safety. These validations consist of informal evaluations of written procedures using technical or customer support engineers, design engineers, and technical writers. Members of the user population (i.e., AMTs) are not usually included in the evaluations.

Evaluations of maintenance documentation are rarely performed despite the fact that writers recognize the value of validation. Seven of the ten writers interviewed in this study said the best way to find errors is through a validation process. As one writer responded when asked, "What suggestions do you have toward improving your job process to make technical documentation more effective?"

> *"Validation! Having held the position as a support engineer, I know what it is like to be 'out in the field' with a customer's A/C (aircraft). The sheer frustration of using documentation which calls up incorrect parts in kit lists or incorrect accomplishment instructions is not only embarrassing but extremely costly – especially when the A/C becomes grounded for extended periods as a result. Validation of technical documentation would alleviate much of the headache."*

Although the benefits of validation are recognized, there has been no systematic attempt to develop the techniques in a manner that they could be adopted by other technical writers. However, a range of evaluative techniques (e.g. cognitive walkthrough, user performance evaluation, etc.) developed by the software industry may be adapted for use in this domain. These techniques represent a set of standardized tools that can be employed by technical writers

The evaluative techniques are part of an iterative, user-centered approach to design and development. The goal of a user-centered design process is to optimize the user's interaction with the tool, system or product so that the user's goals are supported. This proactive approach requires a better understanding of the user's needs, expectations, and how they perform their tasks to ensure the quality of the deliverable. By employing methods borrowed from the software industry aviation technical writers may be better able to identify the ways in which the manual either supports or impedes successful task completion. Similarly, interventions to improve the quality of maintenance procedures can be guided by understanding the types of errors commonly found in maintenance documentation. This information can also be used to gauge the effectiveness of interventions by tracking error types and rates across time. Additional benefits that are expected to follow from the use of these techniques include a reduction in the number of Publication Change Requests (PCR's) submitted by customers, reductions in maintenance errors and completion times, and fewer calls to technical support.

1.4 Purpose.

The purpose of this research was to investigate the applicability of three software *Usability* methods in evaluating aviation documentation and to document the types of errors found in maintenance documentation. A diverse set of participants were recruited to participate in the evaluations in order to document how experience and training affect error detection.

2. METHODOLOGY.

Prior to selecting the usability methods to use in our experiments we first reviewed how technical writers approached their task and which heuristics they employed. The following text describes, 1) the interviews conducted with technical writers, 2) the methods selected for comparison, 3) participants, materials and procedures employed in the evaluations.

2.1. Interviews with Technical Writers.

Structured interviews with aviation technical writers were conducted at two local aircraft manufacturers in Wichita, Kansas. A total of ten writers—five from each manufacturer—were selected to reflect different levels of experience and background. The writers were recommended by the supervisors at the participating Technical Publications departments.

2.1.1 Background of Technical Writers.

All writers reported receiving training on the use of templates, guidelines, and style guides from their employers. The writers averaged 8.5 years of technical writing experience, with a range of 2 to 29 years. Educational backgrounds were diverse, including a master's degree in aeronautical engineering, undergraduate degrees in English (n = 2), Business (n = 3), Engineering (n = 1), and Associates degrees (n = 3). Five of the writers had previous experience as maintenance technicians– four in the USAF and one in the automotive industry. Specific technical writing responsibilities varied from writing new procedures,

making revisions to existing procedures to researching PCRs to identify which changes to the manual were justified.

2.1.2 Manual Development Process and Heuristics.

Regardless of employer, all writers reported using similar processes to write and test maintenance procedures (See Figure 1). Technical writers described several heuristics they used in their writing tasks. These included reviewing safety considerations, writing similar installation procedures together to improve the consistency between them, writing procedures in order of logical steps and location on aircraft, writing procedures as a guide rather than giving a lot of detail, and reviewing the complete maintenance process as a whole before beginning to write a procedure.

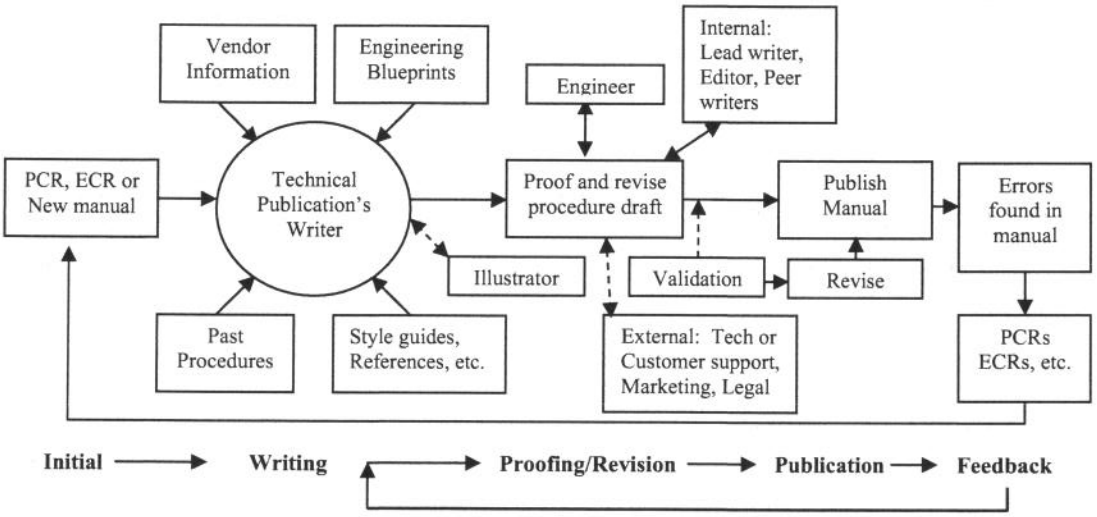

Figure 2. Aircraft Maintenance Manual Development Process.

2.2 Selection and Description of Evaluation Techniques.

Based on our interviews with technical writers, two techniques (described below) were chosen for the evaluation: Cognitive Walkthrough (CW) and User Performance (UP). Two experiments were performed to evaluate each of these evaluative methods. The rationale for selecting these techniques included the following: 1) CW is similar to what manufacturers are currently using to evaluate documentation; however, this process is very informal and rarely involves representatives from the user population (AMTs), and 2) UP evaluation was selected because of an earlier recommendation by Chaparro and Groff [16] and evidence demonstrating the effectiveness of the technique.

Cognitive Walkthrough (CW) is a review technique in which experienced evaluators, familiar with the design and development of the task, review or "walk through" each step of a written maintenance procedure. Participants are instructed to visualize performance of each step as if they were doing the task. The purpose of this method is to identify incorrect technical and

factual information, poor wording choices, and inadequate information regarding what actions to take if the desired outcome is not obtained.

User Performance Evaluation (UP): Unlike CW, a UP evaluation involves a participant physically performing a maintenance task using the written documentation. Participants are chosen who are not familiar with the procedure or its development to ensure that they are representative of the user population (i.e., AMTs), and the written communication can be evaluated without the potential biases arising from knowledge of the writer's intent, or familiarity with the system's design. Two forms of the UP were compared: 1) a single user performs the evaluation and 2) a co-discovery user evaluation where two participants perform the task together. One of the proposed advantages of this method is that more issues may be identified when two evaluators work jointly.

2.2.1 Materials.

A general aviation aircraft manufacturer provided an unpublished maintenance procedure describing step-by-step instructions for rigging a cabin door to an aircraft fuselage. This procedure was chosen because 1) it was unfamiliar to the pool of AMTs and their prior experience did not transfer readily to the new door design, and 2) a computer simulation and physical prototype were available for use in testing. The prototype door was an exact replica of the aircraft fuselage and door assembly. The interior door handle and handrail installation on the new door design were the same as on other cabin doors. However, the hinges and latching mechanisms were new designs.

The procedure consisted of four pages of text and one page of illustrations. There were six major tasks with a total of 105 subtasks. Prior to the experiments, the maintenance procedure was evaluated by production line mechanics and design engineers familiar with the door-rigging task to estimate the number and types of errors within the document. The procedure was not modified as it was judged to have a sufficient number and types of errors. The same maintenance procedure was used for all three evaluations.

2.2.2 Participants.

In this investigation we conducted two experiments using three different evaluation techniques. Significant differences in experience and background existed among the participants in each of the evaluations. Subjects were not assigned randomly to the evaluations; instead, they were selected intentionally based on their level of familiarity with the door design, and technical background to investigate the role experience and training play in error detection.

Cognitive Walkthrough (CW). Nineteen participants, 17 male and 2 female, completed the CW evaluation. The participants were assigned to one of four groups based upon their familiarity with the new door design (expert vs. naïve) and technical background (engineers vs. AMTs). A total of three expert engineers, 5 expert AMTs, 6 naïve engineers, and 5 naïve AMTs participated in the evaluation.

The expert AMTs had performed maintenance on the new cabin door during the aircraft development and flight testing. The expert engineers used in the study were responsible for the design of the cabin door and the creation of documentation including system description, functional tests, engineering drawings containing details regarding component assembly and disassembly. The naïve AMTs were experienced with cabin door rigging procedures on the manufacturer's other aircraft but had no experience with the new cabin door. The naïve engineers were experienced with design, troubleshooting and rigging cabin doors on other aircraft but had no experience with the new cabin door. With the exception of the expert engineers none of the participants had previously seen the maintenance procedure.

User Performance Evaluations (UP). A total of ten naïve AMTs and five naïve engineers (all unfamiliar with the new cabin door rigging procedures) from the manufacturer's service facility participated in the UP Evaluations. Five of the AMTs were assigned to the single-user (SU) evaluation and five were assigned to the Co-discovery (CD) evaluation. The five naïve engineers were teamed with the five naïve AMTs in the CD evaluations. All of the participants in this evaluation were male.

2.2.3 General Procedures.

Prior to the experiment all participants were informed of the purpose of the experiment and were asked to read and sign a consent form and privacy statement. Following the experiment they were asked to complete short background and satisfaction questionnaires (see Appendixes A and B).

Cognitive Walkthrough (CW). Naïve mechanics and engineers watched a short animated video of the door that illustrated the key parts of the cabin door's design and provided an overview of the door's operation. All participants read a paper copy of the maintenance procedure and were asked to note any errors they found including typos, missing or incorrect information and any instructions that were out of sequence, confusing, or did not make sense. Any materials typically referenced (i.e., engineering drawings) while proofing the maintenance procedure were available to the participants while they reviewed the written procedure. The time required to complete the cognitive walkthrough was recorded upon completion (M = 40 minutes, range 26-70 minutes).

User Performance Evaluation (UP). AMTs and engineers were instructed to perform the procedure as written in the maintenance procedure and to verbally describe what they were doing at each step and why they were doing it. They were asked to inform the researcher of any instruction (or part of an instruction) that was incorrect, missing, out of sequence, confusing, or simply did not make sense. The time required to complete the cognitive walkthrough was recorded upon completion (M = 142 minutes, range 105-210 minutes).

3. RESULTS.

The purpose of this project was to document the qualitative differences between the techniques in terms of the types and frequency of errors reported by participants. Statistical analyses were restricted to mean differences between methods of the satisfaction questionnaire results.

3.1 Participant Demographics.

Table 1 shows a breakdown of the participant demographics by experience, background and FAA licensing. Service center AMTs who completed the UP evaluations had the most experience. All participants had extensive experience with the manufacturer's aircraft.

Evaluator type	n	Mean years experience	A & P Certification
CW Naïve AMT	5	5.4	2
CW Naïve Engineer	6	4.0	N/A
CW Expert AMTs	5	6.6	2
CW Expert Engineer	3	7.3	N/A
SU Naïve AMTs	5	14.6	5
CD Naïve AMTs	5	13.0	5
CD Naïve Engineer	5	7.0	N/A

Table 1. Participant Demographics

3.2 Error Taxonomy.

According to Reason and Hobbs (p. 39, 2003) *"An error is the failure of planned actions to achieve their desired goal, where this occurs without some unforeseeable or chance intervention."* Since the problem areas we asked participants to identify in the maintenance procedure had not resulted in an actual failure of planned actions, by this definition, these are technically not errors. However, within the context of this report, errors are defined as those items identified by participants as potential problems areas in the documentation.

To facilitate analysis and interpretation, a taxonomy was developed to categorize the errors identified by the participants. Four error type categories were identified in the evaluations:

- technical (tools, values, tolerances)
- language (clarity of wording/terminology, grammar, typos)
- graphics (dimensions, parts, clarity of illustration)
- procedural (sequencing, separating, combining)

The associated corrective actions suggested by the participants were recorded for later analyses.

3.3 Cognitive Walkthrough (CW) Method.

Although the CW evaluation in the software industry is usually performed by a subject matter expert we tested this method using both naïve and expert groups of AMTs and engineers. The purpose of this experiment is to identify how the number and types of errors vary according to participant background (AMT vs. engineer) and experience (expert vs. naïve).

3.3.1 Participant Responses.

Table 2 shows a summary of the types of errors identified in the CW evaluation for each evaluator group. The values in the table represent the sum of all the errors reported by the participants in each group. The results show that experts (AMTs and engineers) identified more errors (154 vs. 126) than their naïve counterparts and this is true despite the fact that there fewer expert participants (n=8 vs. 11); finally, engineers reported more errors than either group of AMTs (i.e., expert and naïve). However, it should be noted that the data for the naïve engineers was skewed by one participant who reported 55 of the 98 errors. For all groups, the two most common types of errors were language and procedural. These results will be discussed later in more detail.

EVALUATOR	n	TECH	LANG	GRAP	PROC	TOTALS
CW Naïve AMT	5	3	13	3	9	**28**
CW Naïve Engineer	6	4	63	8	23	**98**
CW Expert AMT	5	6	17	4	17	**44**
CW Expert Engineer	3	9	46	25	30	**110**
TOTALS	**19**	**22**	**104**	**38**	**66**	**226**

Table 2. Number of errors reported in the CW method by evaluator group.

A review of the comments made by each user group revealed several differences. Comments by naïve participants were associated with query's regarding the meaning or interpretation of the text. The experts reported more errors that were factual in nature. This result is not surprising since only individuals familiar (i.e., experts) with the door design and its operation can readily identify whether descriptive or factual information is incorrect.

Expert Engineers. The engineers familiar with door design often cited errors that were technically or procedurally incorrect. Examples of comments included "Metric torque values are incorrect.", "The upper side (add "does not") touch the latch roller.", and "(substep) d should be after e."

Expert AMTs. AMTs familiar with the door procedure frequently found errors that would impede task completion. For example, one instruction stated, "If necessary, remove the gears and replace the key …" and the comment by an expert AMT during CW was, "Gears are sealed on – is there actually access to these?"

A comparison of the comments made by the two groups of expert participants showed that of the 154 errors found by both groups (110 by engineers and 44 by AMTs) only 11 comments pertained to the same steps in the procedure.

Naïve Engineers. The naïve engineers frequently pointed out issues related to language (typos, grammar, and wording) and procedural steps (sequencing, separating, and combining). For instance they cited the absence of detail indicating what action to perform next if a stated action did not meet the specified requirement (e.g., What if it doesn't measure 0.10?). The naïve engineers were more likely to find grammatical or typographical language errors which, when changed, would make the written presentation better, but would most likely have not prevented the AMT from completing the task.

Several issues identified by the naïve engineers in the CW were later reported as problems in the UP evaluation. Three naïve engineers reported that the wording "Adjust …until the force needed to close the cabin entry door and the handle forces are the best between them." needed clarification. In the UP evaluations, this step was cited as unclear by three of the naïve AMTs in Single-User evaluations, two naïve engineers and two naïve AMTs in the Co-Discovery evaluations.

Naïve AMTs. The naïve AMTs reported the fewest errors, and their comments usually were in the form of queries to clarify the tasks. For instance, "How much fuel is in the aircraft?", "Can all of the rollers be removed and work them one at a time? "How can we take such an exact measurement?" Although the naïve AMTs made significantly fewer comments than the naïve engineers, their comments were also similar to those subsequently obtained in the UP evaluations.

The results illustrate the unique contributions made by different evaluators. Because of their familiarity with the procedure, system experts were better able to identify errors in technical information and system descriptions. However, due to their familiarity with the system they were less likely to identify vague, unclear, and imprecise procedural descriptions reported by the naïve participants.

3.4 User Performance (UP) Evaluation Methods.

Unlike CW, UP evaluations were performed by members of the user population and involve actually conducting the procedures on a physical article. The purpose of this experiment was to identify the relative benefits of Single User (SU) vs. Co-Discovery (CD) methods and to investigate how the identified error types varied by method.

EVALUATOR	N	TECH.	LANG.	GRAPHIC	PROC.	TOTAL
SU Naïve AMT	**5**	**14**	**47**	**34**	**67**	**162**
CD total		**34**	**115**	**40**	**142**	**331**
CD Naïve AMT	5	20	89	29	107	245
CD Naïve Engineer	5	14	26	11	35	86
TOTALS	**15**	**48**	**162**	**74**	**209**	**493**

Table 3. Number of errors reported in UP Evaluation by evaluator group and method. The contributions of the naïve AMTs and engineers in the CD are indicated.

The values in Table 3 show that CD evaluation method was relatively more effective in identifying errors than the SU method. Roughly twice as many issues were reported by participants using the CD vs. the SU method. A comparison of the contributions made by AMT and engineers in the CD method show that AMTs identified many more errors (roughly three-fold greater) associated with procedural, language and graphics than did the engineers.

Like the results from the CW, procedure and language errors were again the most frequently cited problems. The most common types of procedural errors were missing information (n = 89), followed by problems with wording (n = 159). Instances of missing information were usually associated with the absence of instructions regarding what actions to perform if a stated value or condition was not met, steps for disassembling or reassembling components, and steps to open or close the door. Problems with language clarity included the use of unfamiliar part names, lack of consistency in the procedure, and subjective language, such as "...seal *can* be removed (AMT comment, "Does it need to be removed or not?" or "make sure ... operates *correctly*" (AMT comment, "What is correctly? Correct gap or correct position?). When unfamiliar part names were referenced the AMTs would often rely on their experience to identify the relevant part. This was not always sufficient as several of the AMTs volunteered that they would have taken apart or adjusted the wrong component.

These results illustrate the ways in which a written statement can be misinterpreted. Ambiguities are more salient to the user when they have to convert written statements into action. In addition, physical obstructions that make the procedure difficult or impossible to perform become obvious. The results also demonstrate the benefits derived from having evaluators work as a team.

3.5 Comparison of Evaluation Methods.

The analyses below summarize the relative frequencies of different error types as a function of evaluative technique, the percentage of errors reported by different evaluator groups, the number of unique errors identified using each method, and a breakdown into subcategories of the two most frequently cited error types.

3.5.1 Errors Reported in Evaluation Methods.

Figure 3 illustrates the average number of the four major error types (language, graphic, procedural and technical) reported by participants using the two evaluation methods (CW and UP). Unlike the earlier analysis, the values for each bar represent the total number of errors divided by the number of participants in each group. Thus these data show the number of errors that an individual participant may contribute in an evaluation. These results demonstrate the benefits of performing the maintenance procedure on an aircraft. On average a participant in the UP Evaluations identified twice as many errors (means of ~33 vs. ~15) as participants who performed a CW. Note that the CW was relatively more effective at detecting language errors while the UP evaluations resulted in more procedure errors.

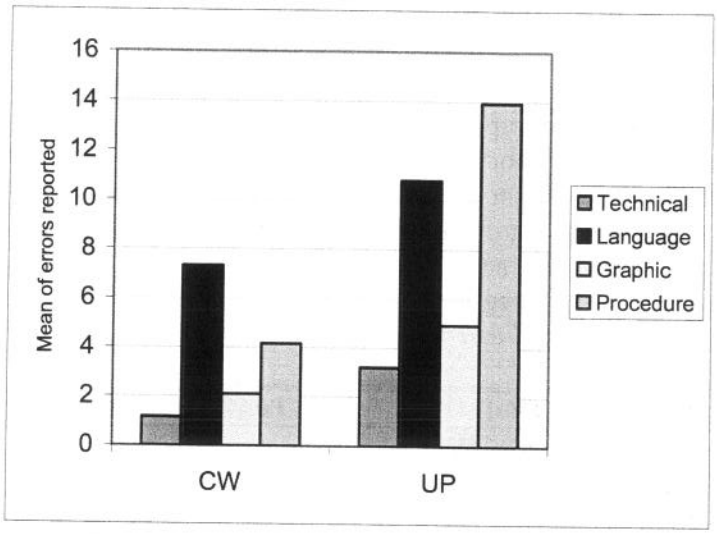

Figure 3. Error frequencies as a function of evaluation method and error type.

The number of errors identified by all participants was summed and Figure 4 shows the percentage of total as a function of participant background (Engineer vs. AMT and Expert vs. Naïve) and evaluation method (CW, CD, SU). No attempt was made to eliminate errors that were identified by more than one participant. In the CW a comparison of AMTs and Engineers (both naïve and expert) shows that the majority of errors, 27% vs. 9%, were identified by engineers. In contrast, the results for the UP Evaluations show a very different pattern: AMTs using the SU and CD methods reported 53% of the errors, while engineers in the CD method contributed roughly 11%.

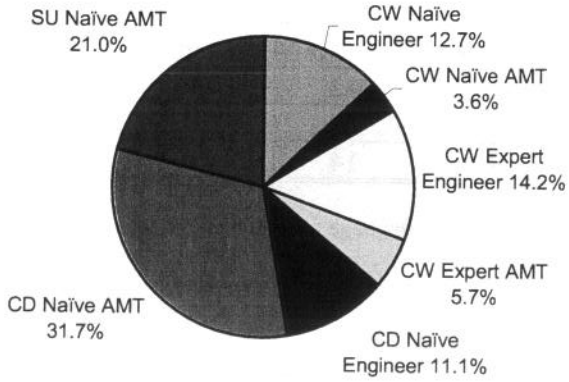

Figure 4. Percent of errors reported by evaluator background and experience.

3.5.2 Unique Errors Identified by Different Evaluator Groups.

In many instances, the same error was reported by more than one participant in the experiment; these redundant reports were eliminated and the sum of these single instance or "*unique*" errors for each method are shown in the right most column of Table 4. The results for UP are broken down to show the number of unique errors for both SU and CD. This analysis shows again that significantly more errors were reported using the UP technique than CW and that of the UP methods more errors were identified using CD than SU. It is important to note that the three techniques were *not* redundant as the CW method had 21 errors in common with the SU and 45 errors in common with CD.

EVALUATION METHOD	TOTAL ERRORS	UNIQUE ERRORS
Cognitive Walkthrough (CW)	226	**152**
User Performance (UP) Evaluations:	493	**219**
Single User (SU)	162	**78**
Co-Discovery (CD)	331	**141**
TOTALS	719	**371**

Table 4. Unique errors found as a function of evaluation method used.

3.5.3 Analysis of Procedure and Language Errors.

The two most cited error types were language and procedure. Language and procedure errors were classified into subcategories shown in Tables 5 and 6, respectively. The analysis reveals that the UP evaluations were effective in spotting language errors related almost exclusively to clarity (n = 159 of 162 total) whereas the CW technique identified a more diverse set of language errors.

EVAL. METHOD	LANGUAGE ERRORS					
	CLARITY	INCORRECT	TYPO	GRAMMAR	MISSING	UNNESS
CW	101	12	11	9	5	1
UP	159	2	0	1	0	0
TOTAL	**260**	**14**	**11**	**10**	**5**	**1**

Table 5. Number of language error subcategories reported by evaluation method.

The four most commonly cited procedure errors were similar for both methods and included missing procedures, sequencing of steps, unnecessary information, and incorrect information. With one exception, "unnecess" category, more errors were identified by subjects in UP.

EVAL. METHOD	PROCEDURE ERRORS					
	MISSING	SEQUENCE	UNNECES.	INCORRECT	COMBINE	CLARITY
CW	20	12	28	13	4	2
UP	89	44	19	34	13	10
TOTAL	**109**	**56**	**47**	**47**	**17**	**12**

Table 6. Number of procedure error subcategories reported by evaluation method.

A corrective action was usually implied when errors were reported. Figure 5 shows a summary of this data. The majority of these comments for both User Performance (SU and CD) and Cognitive Walkthrough (CW) techniques requested either changing or adding more information to the procedure. Note that twice as many comments requesting that information be included in the procedures were obtained through UP (n = 280) than CW (n = 138).

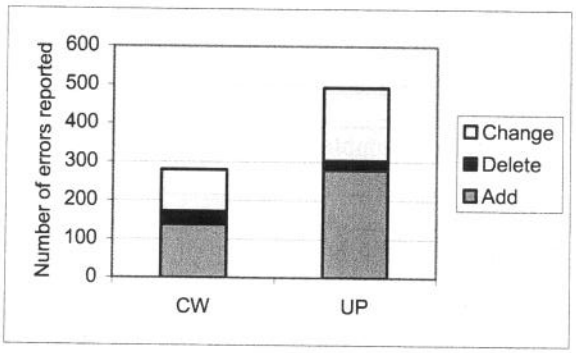

Figure 5. Comparison of the corrective actions by evaluation method.

3.6 Satisfaction Measures.

A scale was developed to assess the participants' satisfaction with the written procedure. The scale had ten individual statements of satisfaction measured on a 5-point agreement scale; Strongly Disagree to Strongly Agree (See Appendix A). Using statistical software (SPSS), a Cronbach's Alpha of .92 was calculated revealing good scale reliability in measuring satisfaction. Three additional statements asked for a judgment of the procedure's complexity relative to other procedures, whether additional instructions would be needed to complete the procedure and an open-ended query of what would improve the procedure.

Results (see Table 7) of the satisfaction measures were analyzed by method, i.e., CW and UP evaluations (Single-User (SU) & Co-Discovery (CD), and by user group, (expert engineer, expert AMT, naïve engineer, and naïve AMT). Generally participants in the CW method were more satisfied with the written procedure, giving it a mean rating of 3 or higher (i.e., greater satisfaction) on the ten satisfaction statements and the overall satisfaction query; whereas, those who participated in UP evaluations rated the procedure <3, (less satisfaction). Table 8 shows the mean satisfaction score by user group and evaluation method. A comparison of Table 8 and Figure 4 reveals that generally satisfaction scores seem to be related to the number of errors found. For example, naïve AMTs and expert engineers reported the fewest and most errors in the CW evaluation, respectively. The data in Table 8 show that the naive AMTs and expert engineers reported the highest and lowest satisfaction scores, respectively. Both groups indicated that the procedure needed more instructions and were neutral that this procedure was "more complex than most."

SATISFACTION STATEMENT	COGNITIVE WALKTHROUGH *n* = 18 *M(SD)*	USER PERFORMANCE *n* = 15 *M(SD)*
I am satisfied with the number of steps included.*	3.74 (1.10)	2.33 (.90)
The procedure was clearly written.*	3.47 (.96)	2.40 (.83)
The illustration was helpful.*	3.68 (1.64)	2.80 (1.27)
The amount of information included was useful.*	3.78 (1.00)	3.20 (.56)
Total Satisfaction Score*	68.35 (15.32)	54.00 (14.38)
I can accomplish the task quickly using this procedure.	3.16 (1.07)	2.53 (1.19)
It is easy to understand what is needed.	3.16 (1.21)	2.53 (.92)
It uses the fewest steps possible to accomplish the task.	3.21 (1.03)	2.60 (.99)
It is easy to learn the procedure.	3.16 (.96)	2.87 (.83)
It was easy to remember the steps.	3.11 (.88)	2.87 (1.19)
The terminology used in the procedure is easy to understand.	3.33 (1.09)	2.87 (.83)
Overall, how would you rate this procedure?	3.39 (1.09)	2.87 (.64)
I would need additional instructions to complete the procedure.	3.63 (1.38)	3.40 (1.45)
This procedure is more complex than most.	2.94 (1.21)	2.93 (.80)

*Significant differences between CW and UP at $p < .05$ level

Table 7. Results of satisfaction measures between CW and UP evaluation methods on a 5-point scale (Strongly disagree … Strongly agree).

USER GROUP	*n*	*M(SD)*
CW Naive AMT	5	80.00 (9.05)
CW Naive Engineer	6	63.00 (6.66)
CW Expert AMT	5	68.80 (9.65)
CW Expert Engineer	3	41.33 (14.46)
UT AMT	5	58.00 (15.23)
CD AMT	5	48.00 (8.48)
CD Engineer	5	54.40 (20.51)

Table 8. Mean satisfaction values as a function of user group and evaluation method.

Listed below are the comments made in response to the open-ended question, "What would you do to improve this procedure?"

Cognitive Walkthrough:

- Make it more to the point. Need to help mechanic not hold his hand. Clean some verbiage up. Too many filler words.
- More steps and illustrations.
- More illustrations with views and parts labeled. Reference to specific views on engineering drawings where applicable.
- Not enough background on the subject to adequately answer.
- If you could publish the video it would be extremely helpful.
- More pictures, have door in front of me.

- Add more visuals such as figures.
- Add a specific chart for torque and clearance of key components. The CATIA model was really helpful. Some parts could be faster and others slower.
- Add more illustrations; identify parts and locations on illustrations.
- Find out if you should have a specific load of fuel on. I would also have to put it to practical use to see if I had any problems.
- Details on troubleshooting – cabin door.
- Add more illustration. There should be a good, clear illustration for each procedure that shows all the parts involved.

User Performance Evaluations:

- Reword nomenclatures. Add illustrations. Label tooling.
- More illustrations on door latch measurements.
- Needs more illustrations.
- Rearrange some steps and include better descriptions to accomplish procedures in proper sequence or rigging door and provide better access to accomplish spade door hatching rig and check.
- More illustrations. Steps in correct order, more clear directions.
- Better definition of nomenclature (i.e. use figure to define part, locations)
- Illustrations
- Unify/simplify terminology and illustrations to define locations and parts
- Reorder steps to ease operations.
- After the procedure is finalized go back through it one last time.
- Improve on terminology of parts and remove some steps.
- More illustrations for clarity
- Separate some tasks.
- More pictures with items pointed out.

Nine of the fourteen participants in SU and CD and six of the twelve CW participants who wrote additional comments mentioned that addition of illustrations would improve understanding of the documentation.

4. DISCUSSION.

The results of this investigation show that 1) User Performance and Cognitive Walkthrough evaluations are complementary techniques for evaluating maintenance documentation, 2) the errors identified by individual participants varied in significant ways according to experience (expert vs. naive) and training (engineers vs. AMTs), and 3) that procedure and language errors are the most commonly cited errors in the maintenance documentation.

The User Performance (UP) evaluation was found to be very effective in identifying potential problems in maintenance documentation. This advantage derives from the fact that execution of a procedure will reveal how the users' interpretation differs from the intent of the writer. A proof reader cannot know if his/her interpretation of the procedure is in error unless they are queried about their understanding. Also, ambiguous language becomes more salient when a user is confronted with the task of translating written statements into specific

actions. For the naïve AMT, the meaning or intent of the text may not be obvious given that they don't share the writer's or engineer's familiarity with the system, its design or functioning. Finally, performance of a procedure is more likely to reveal physical constraints (i.e., location of access panels, space limitations, presence of physical obstructions) that make performance of the procedure difficult if not impossible. Of the two UP techniques, the CD method proved to be superior in terms of the total number of problems that the participants reported. This advantage seems to derive from the interaction of the participants. By working as a team they appear to more readily identify areas where their interpretations and understanding of written procedures differ.

The results for the CW and UP methods suggest that the two techniques should be viewed as complementary techniques. The errors identified using CW and UP overlap only partially, indicating that each may be effective in identifying different kinds of user problems (see Table 4). This also reflects in part an influence of participant background. The expert engineers and AMTs may more readily detect incorrect technical information or erroneous system and functional descriptions. Application of CW by experts early in the development of a procedure may be an effective way to identify and eliminate technical or factual errors that can create problems during execution of the procedure.

The CW method was selected for evaluation because it was similar to the process employed by manufacturers to proof documentation. However, our methods differed in several respects from those typically employed by manufacturers including 1) use of expert AMTs to review and make comments during the development of the written procedure; 2) use of multiple reviewers during the evaluation; 3) expert participants were asked to evaluate the procedure in a controlled setting, without distraction and 4) an exit interview followed evaluation to ensure that the evaluators comments ware interpreted correctly. We believe that these methods maximize the effectiveness of the evaluative techniques.

In this investigation, we evaluated the contribution made by naïve engineers and AMTs using the CW method of evaluation. Even though these participants were unfamiliar with the design or function of system described in the procedure, they identified a number of problems that were not reported by the expert group. Comments made by naïve AMTs, although fewer in number, were similar to those made by the AMTs in the UP evaluations. Thus, CW evaluations performed by naïve engineers and AMTs can complement the evaluations performed by experts and represents a useful alternative in cases where a UP evaluation cannot be performed.

At present, members of the user population (AMTs) are not usually involved in evaluating maintenance documentation. This is true despite the fact that they may be the best judges of how well the maintenance documentation meets their needs, or more importantly, how it fails to meet those needs. The participants in the UP evaluations that involved execution of the procedures by an AMT reported less satisfaction with the procedure than participants in the CW evaluations. The UP evaluations identified more procedural and language errors than the CW evaluations. In most cases the evaluators requested additional information be included in the manual and the editing existing text for clarity but rarely did they request the deletion of information.

The current system of document evaluation relies on the customer to identify these errors. The cost of these errors to an operator may be reflected in maintenance costs given that the majority of the language and procedure errors (see Tables 5 and 6) cited by the participants relate to *clarity* and *missing* procedures. The AMT participants were selected to be representative of the population that utilizes maintenance documentation. However, this participant group may be more familiar with the manufacturer's aircraft than most AMTs and customer support engineers in Fixed Base Operations (FBO) facilities. Therefore, our results represent a conservative estimate of what might be obtained using AMTs less familiar with this manufacturer's aircraft.

The technical literature on usability evaluation methods for testing technology applications shows that the majority of errors are found using three to five evaluators [17, 18]. Using multiple evaluators ensures that differences in the user population are represented. However, it is usually recommend that three to five evaluators be used as there are diminishing benefits associated with more evaluators.

4.1 Recommended Best Practices.

Observations made during testing and analysis of the participant comments revealed a number of frequently cited problems with the presentation of information in the written procedures.

- Steps which are rechecks should include necessary numeric information so the AMT does not have to go back nor rely on memory for the values.
- Identify changes in location on aircraft, e.g., perform inside or outside aircraft.
- Illustrations should identify visual orientation of part to aircraft.
- Procedures should be broken into separate logical subtasks. These subtasks should be clearly separated (e.g., spaces between).
- Include necessary information within the step rather than referring to another part of the manual.
- Installation and re-installation processes should contain more information than disassemblies. Cues or checks should be included within the procedure to ensure correct assembly.
- After initial assembly of the aircraft, most maintenance practices will be for trouble shooting customer complaints and routine maintenance practices. Make sure the manual addresses the critical information needed for each of these AMT needs.
- Consider the environment in which the AMT is performing the procedures. For example, space on the aircraft is often cramped and light is scarce, so some of the measurements may be obstructed and difficult, perhaps impossible, to obtain.
- Make sure the nomenclature is clear. AMTs need to know the naming of components and have an overview of how the individual components work within the system.

- When there are similar aircraft, make sure that differences are emphasized to prevent transfer of the AMT's previous knowledge when it is not applicable.
- Every disassembly process should be followed by a reassembly process.
- Human factors principles in presentation of documentation should be incorporated into technical writers' guidelines. For example, sentences should not be in all caps, the number of sub-steps should be limited (there were 39 in task 4), consistent terminology and symbols should be used, e.g., (4) *as substep*, (4) *as the number of parts*, four (4) *as number of parts*; and *bolts*, *screws*, and *fasteners* were all used interchangeably.

4.2 Integrating Evaluation Techniques in the Technical Writing Process.

As illustrated in Figure 6, there are three main phases to the development of aircraft maintenance manuals: 1) Manual Concept, 2) Manual Design, and 3) Manual Development.

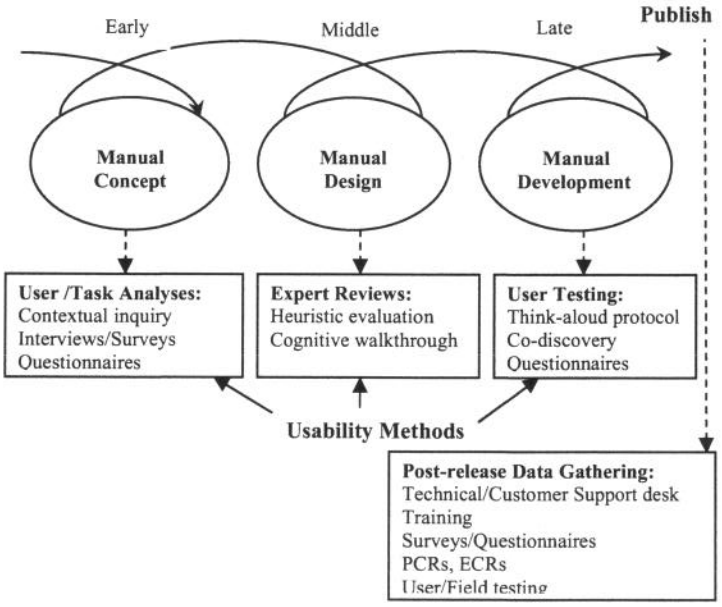

Figure 6. User-centered iterative design in the aircraft manual development process.

Different techniques can be used to collect user information during each phase (i.e., User/Task Analyses, Expert Reviews, and User Testing). For instance, CW and UP evaluations may be performed during the manual design and manual development phases, respectively.

Although our investigation did not test methods used in the Conceptual phase, there are evaluation methods that writers may consider using to identify constraints the AMTs have while using the maintenance documentation. As part of this research project, we have developed a technical writer's "toolbox" that outlines evaluative methods which have been adapted for aviation technical documentation. The toolbox consists of descriptions of each evaluation technique, guidelines for using the methods, and various supporting documents

(questionnaires, data collection forms, etc.) that can be used during the evaluations. (The toolbox is available at http://156.26.14.231:90/newsurl/faa_toolbox3/).

The selection of a technique depends on a number of factors including the phase of the manual, criticality of the procedure, availability of personnel, access to the aircraft and time constraints. Some methods can be performed quickly at less cost in terms of personnel and time; however, the scope and depth of the evaluation is similarly reduced. More highly structured techniques may be reserved for testing maintenance procedures that are cost or safety critical.

4.3 Technical Writers' Training.

The absence of standard methods for evaluating technical documentation is a concern for the Technical Publication's departments of aircraft manufacturers. Most of the writers we interviewed saw validation as the preferred method for evaluating maintenance procedures. However, there has not been a set of tools available to writers to perform this type of evaluation systematically and consistently. The evaluation methods tested in this research provide a foundation for the development of new methods to evaluate maintenance documentation that suit the unique demands of the aircraft industry.

The technical writing community will need to be trained in the use of these techniques. We anticipate that the techniques will be incorporated in an existing aircraft technical writing course that grew out of an earlier FAA funded research project [16]. The Usability Toolbox was created to serve as a resource for those in the technical writing community looking to employ these methods. The results of the evaluations can also serve an important educational tool when used to provide feedback to individual writers.

5. FUTURE RESEARCH.

Although aircraft manufacturers are aware that there are costs incurred from inadequate maintenance documentation, to date there have been few quantitative analyses of the costs. These include the significant personnel and customer costs (loss of revenue due to aircraft down time) as well as potentially significant liability costs, associated with deficient maintenance manual documentation. Promotion of validation efforts face serious obstacles given the current economic climate and that that they represent an added cost that most manufacturers do not currently incur. Obviously, the preference is for manufacturers to voluntarily adopt validation practices rather than forcing compliance via regulation. Nevertheless, voluntary compliance may not be an alternative in the future as the NTSB has made the following recommendation to the FAA as part of the accident report pertaining to Air Midwest Flight 5481 [10]:

> "Require 14 *Code of Federal Regulations* Part 121 air carriers to implement a program in which carriers and aircraft manufacturers review all work cards and maintenance manual instructions for critical flight systems and ensure the accuracy and usability of these instructions so that they are appropriate to the level of training of the mechanic performing the work."

The evaluation techniques used in this study represent the basis for formal and systematic attempts to validating maintenance documentation during its development. We believe that these techniques will benefit manufacturers, operators and the flying public.

6. REFERENCES.

1. Marx, D.A. and R.C. Graeber, *Human error in aircraft maintenance*. Aviation Psychology in Practice. 1994, Aldershot: Avebury Technical: Ashgate.
2. Boeing Airplane Group, B.C.A., *Statistical Summary of Comercial Jet Aircraft accidents, World Wide operations 1959-1995*. 1996, Boeing Airplane Group: Seattle, WA.
3. *Maintenance errors appear to contribute to more accidents*, in *Air Safety Week*. 2000, Phillips Publishing International, Inc.
4. Ricci, K. *Human factors issues in maintenance publications design*. in *DOD Maintenance Symposium & Exhibition*. 2003. King of Prussia, PA: SAE International.
5. Hobbs, A. and A. Williamson, *Associations between errors and contributing factors in aircraft maintenance.* Human Factors, 2003. **45**(2): p. 186-201.
6. Reason, J. and A. Hobbs, *Managing Maintenance Error*. 2003, Aldershot, U.K.: Ashgate.
7. Dekker, S., *Field Guide to Human Error Investigations*. 2003, Aldershot, U.K.: Ashgate.
8. McDonald, N., et al., *Safety management systems and safety culture in aircraft maintenance organisations.* Safety Science, 2000. **34**: p. 151-176.
9. NTSB, *Loss of control and impact with Pacific Ocean Alaska Airlines Flight 261 McDonnell Douglas MD-83, N963AS about 2.7 miles north of Anacapa Island, California January 31, 2000*, in *National Transportation Safety Board*. 2002, National Transportation Safety Board: Washington D.C.
10. NTSB, *Report of aviation accident loss of pitch control during takeoff Air Midwest Flight 5481, Raytheon (Beechcraft) 1900D, N233YV Charlotte, North Carolina, January 8, 2003*, in *National Transportation Safety Board*. 2004, National Transportation Safety Board: Washington D.C.
11. Chaparro, A. and L.S. Groff, *Human Factors Survey of Aviation Technical Manuals Phase 1 Report*. 2001, Federal Aviation Administration: Washington D.C.
12. Chaparro, A., et al., *Survey of aviation technical manuals Phase 2 report: User evaluations of maintenance documentation*. 2002, Federal Aviation Administration: Washington D.C.
13. Hobbs, A. and A. Williamson, *Aircraft maintenance safety survey-results*. 2000, Department of Transportation and Regional Services, Australian Safety Bureau.
14. Hobbs, A. and A. Williamson, *Skills, rules and knowledge in aircraft maintenance: errors in context.* Ergonomics, 2002. **45**(4): p. 290-308.
15. Norman, D.A., *The Psychology of Everyday Things*. 1988, New York: Basic Books.
16. Chaparro, A. and L.S. Groff, *Reliability Survey of Aviation Maintenance Manuals Phase 3 Report: Final Report and Recommendations*. 2002, Federal Aviation Administration: Washington D.C.
17. Nielson, J., *Heuristic evaluation*, in *Usability Inspection Methods*, J. Nielson and R.I. Mack, Editors. 1994, John Wiley & Sons: New York.
18. Virzi, R.A., *Refining the test phase of usability evaluation: How many subjects is enough?* Human Factors, 1992. **34**(4): p. 457-468.

7. APPENDICES.

A. User Satisfaction Questionnaire.

1. I can accomplish the task quickly using this procedure.

Strongly Disagree 1 2 3 4 5 Strongly Agree

2. I am satisfied with the number of steps included.

Strongly Disagree 1 2 3 4 5 Strongly Agree

3. It is easy to understand what is needed.

Strongly Disagree 1 2 3 4 5 Strongly Agree

4. It uses the fewest steps possible to accomplish the task.

Strongly Disagree 1 2 3 4 5 Strongly Agree

5. The procedure was clearly written.

Strongly Disagree 1 2 3 4 5 Strongly Agree

7. It is easy to learn the procedure.

Strongly Disagree 1 2 3 4 5 Strongly Agree

8. The illustration was helpful.

Strongly Disagree 1 2 3 4 5 Strongly Agree

9. It was easy to remember the steps.

Strongly Disagree 1 2 3 4 5 Strongly Agree

10. The amount of information included was useful.

Strongly Disagree 1 2 3 4 5 Strongly Agree

11. The terminology used in the procedure is easy to understand.

Strongly Disagree 1 2 3 4 5 Strongly Agree

12. I would need additional instructions to complete the procedure.

Strongly Disagree 1 2 3 4 5 Strongly Agree

13. This procedure is more complex than most.

Strongly Disagree 1 2 3 4 5 Strongly Agree

14. Overall, how would you rate this procedure?

Very Poor 1 2 3 4 5 Very Good

15. What would you do to improve this procedure?

B. User Background Questionnaire.

Please list the aircraft you currently work on and indicate the length of time you have worked on each:

Aircraft Name	Specialty Area (wiring, avionics, engines, ALL)	Less than 3 months	3 – 6 months	6 -12 months	1 - 3 years	More than 3 yrs

Education Level:

- ☐ High school graduate
- ☐ Some college
- ☐ Bachelor's degree
- ☐ Graduate degree
- ☐ Technical degree or training
- ☐ Other (specify)

_____

Type of Maintenance Typically Performed:

- ☐ Line maintenance
- ☐ Base maintenance

Certification

- ☐ None
- ☐ Yes (specify)

_____

Sex

- ☐ Male
- ☐ Female