



February 2005, Vol. 7 Issue 1

| [Volume 7 Issue 1](#) | [Past Issues](#) | [A-Z List](#) |

Usability News is a free web newsletter that is produced by the Software Usability Research Laboratory (SURL) at Wichita State University. The SURL team specializes in software/website user interface design, usability testing, and research in human-computer interaction.
[Barbara S. Chaparro](#), Editor

A Comparison of Two Evaluation Techniques for Technical Documentation

by [Bonnie Rogers](#), [Chris Hamblin](#), & [Alex Chaparro](#)

National Institute for Aviation Research & Wichita State University

Summary: This study compared two evaluation techniques, Usability Testing and Cognitive Walkthrough, in their ability to identify errors in aviation maintenance documentation. The techniques were evaluated to see how much unique information they each produced as well as the type of errors identified. Results showed that the techniques were complementary in their findings and both are recommended in the development of technical documentation.

INTRODUCTION

The quality of documented maintenance procedures have been cited as a primary factor contributing to maintenance errors [1-4]. A review of Naval Aviation Maintenance mishaps that occurred between 1990 and 2003 [5] 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, and incorrect technical information and diagrams. However, because mishaps are rare events, they underestimate the frequency of incidents in which poor documentation resulted in maintenance errors. Also, mishaps do not account for the other effects of poor documentation including the costs of incorrectly executed or slowed maintenance.

Surveys reveal that aviation manufacturers rely on aircraft maintenance technicians (AMTs) to identify problems in maintenance documentation. Most corrections to the maintenance documentation are post-release through reports of problems by AMTs, called Publication Change Requests (PCRs). However, assuming that AMTs will report errors in maintenance procedures may be incorrect. Chaparro et al. [6] found that 53% of AMTs reported only occasionally, rarely, or never reporting errors found in the maintenance documentation.

Chaparro et al. also found that 64% of AMTs reported finding their own way of performing a procedure [6]. Nearly 60% of AMTs reported continuation of an unfamiliar task despite not being sure if they were performing it correctly [7]. Similarly, McDonald et al. [4] reported that 34% of routine maintenance tasks are performed in ways different than outlined in the maintenance

documentation. The quality of maintenance documentation may be improved through a systematic evaluation before publication by adapting software usability techniques to maintenance documentation. Proofreading to correct the procedures in the maintenance manual prepublication typically does not solicit user (AMT) input.

METHOD

This investigation compared techniques that manufacturers can use to improve the quality of maintenance documentation developed by technical writing groups. The techniques, Usability and Cognitive Walkthrough evaluations, were used to identify errors in the maintenance documentation. The techniques highlight probable mismatches between the intent of the writer and the interpretation of the maintenance documentation by AMTs. The purpose of this investigation was (i) to determine whether these techniques elicit unique (non-redundant) information from participants, (ii) to establish whether different evaluators (engineers, AMTs) identify different types of errors, and (iii) to identify common types of errors in the maintenance documentation.

Procedure

Participants evaluated a draft of a maintenance procedure for rigging an aircraft door using the two techniques. Participants were asked to either perform the rigging task as described in the procedure on an actual door (i.e., Usability Testing) or to proofread the draft maintenance documentation (i.e., Cognitive Walkthrough). They were instructed to identify any errors (typos, missing or incorrect information, sequencing of steps, etc.) they found in the procedure. Participants varied in their type of training (i.e., AMT vs. Engineer) and level of familiarity (i.e., naïve/unfamiliar vs. expert/familiar) with the aircraft door.

Cognitive Walkthrough (CW). Naïve mechanics and engineers watched a short animated video 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).

Usability Testing (UT). Naïve 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. Usability testing was conducted with both single users (SU) and with two users - co-discovery (CD). The time required to complete the cognitive walkthrough was recorded upon completion (M = 142 minutes, range 105-210 minutes).

RESULTS

Table 1 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 were 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.

Table 1. Number of errors reported in the CW method by evaluator group.

Evaluator	N	Tech	Lang	Graphic	Proc	Total
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	139	40	79	280

Unlike CW, UT evaluations were performed by members of the user population and involved 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.

Table 2. Number of errors reported in Usability Testing by evaluator group and method (SU = Single User, CD = Co-Discovery).

Evaluator	N	Tech	Lang	Graphic	Proc	Total
SU naïve AMT	5	14	47	34	67	162
CD total	0	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

The values in Table 2 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, procedural and language errors were again the most frequently cited problems. 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?)

CONCLUSIONS

The results of this investigation show that 1) Usability Testing and Cognitive Walkthrough evaluations are complementary techniques for evaluating maintenance documentation, 2) the

errors identified by individual participants varied in significant ways according to familiarity (naïve vs. expert) and training (AMTs vs. engineers), and 3) procedure and language errors were the most commonly cited errors in the maintenance documentation.

Usability Testing 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 user's interpretation differs from the intent of the writer. A proofreader cannot know if his/her interpretation of the procedure is in error unless queried about his/her understanding. Also, ambiguous language becomes more salient when a user is confronted with the task of translating written statements into specific actions. Of the two usability testing 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.

Results from this analysis suggest that the two techniques should be viewed as complementary. The errors identified using each overlap only partially, indicating that both are effective in identifying different kinds of user problems. Application of a cognitive walkthrough 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.

Note: This research was presented at the 13th International Symposium on Aviation Psychology (ISAP) Oklahoma City, OK, April 18-21, 2005.

REFERENCES

1. Hobbs, A. and A. Williamson (2003). Associations between errors and contributing factors in aircraft maintenance. *Human Factors*, 45(2): p. 186-201.
2. Reason, J. and A. Hobbs (2003). *Managing Maintenance Error*, Aldershot, U.K.: Ashgate.
3. Dekker, S. (2003). *Field Guide to Human Error Investigations*, Aldershot, U.K.: Ashgate.
4. McDonald, N., Corrigan, S., Daly, C., & Cromie, S. (2000). Safety management systems and safety culture in aircraft maintenance organisations. *Safety Science*, 34: p. 151-176.
5. Ricci, K. (2003). Human factors issues in maintenance publications design. in DOD Maintenance Symposium & Exhibition, King of Prussia, PA: SAE International.
6. Chaparro, A. Groff, L. Chaparro, B. & Scarlett, D. (2002). Survey of aviation technical manuals Phase 2 report: User evaluations of maintenance documentation, Federal Aviation Administration: Washington D.C.
7. Hobbs, A. and A. Williamson (2000). Aircraft maintenance safety survey-results. Department of Transportation and Regional Services, Australian Safety Bureau.

SUBSCRIBE to *Usability News*!