Usability Evaluation of a Fault Tree Software User Documentation

by

Samuel S. Lee

Thesis Submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in Partial Fulfillment of the Requirements for
the Degree of

MASTER OF SCIENCE

in

Industrial and Systems Engineering

APPROVED:

D.L. Price, Chairman

K.H.E. Kroemer

P.T. Kemmerling

Blacksburg, Virginia

February 21, 1992

Usability Evaluation of a Fault Tree Software User Documentation

by

Samuel S. Lee

Committee Chairman: Dr. Dennis L. Price
Industrial and Systems Engineering

(ABSTRACT)

To incorporate users' opinions into the evaluation phase early in the software documentation development stage, the critical incident technique was used to identify usability problems in a fault tree software user document. The critical incidents were used to modify the original document to improve its usability. To test whether the modified document was better in usability than the original document, an experiment was conducted to compare objective measures and subjective ratings.

Four objective measures included number of errors, task completion time, document reading time, and number of personal helps requested. In addition, subjective ratings on ease of use, accuracy of information, inconsistencies, ease of learning, completeness, helpfulness of figures, and ease of understanding were compared between the two documents.

The analyses showed that for 3 of 4 objective measures and 7 out of 9 subjective ratings, the new document was better and easier to use. In some cases, this difference was task specific. Generally, easier tasks accounted for better objective measures and more favorable subjective ratings.

ACKNOWLEDGEMENTS

First, I would like to thank Dr. Dennis L. Price for serving as the chairman of my faculty committee. From the very beginning of the research proposal to the last word written on this thesis, he has been thoroughly supportive of my effort. His critical yet encouraging feedback throughout the research has been invaluable and is much appreciated. The National Institute for Occupational Safety and Health provided me with the funds to conduct this research. EG&G provided the Integrated Reliability and Risk Analysis System (IRRAS) software and user documentation and permitted this research to be conducted.

I would also like to thank Dr. K.H.E. Kroemer for his genuine interest and support as a member of my faculty committee. Without the research experience I have had under his guidance in the past couple of years, I would not have approached this project with as much confidence and zeal. His encouragement over the years is greatly appreciated.

My thanks is also extended to Professor Paul T. Kemmerling for willingly serving on my committee in the midst of his loaded schedule. His flexibility is much appreciated.

I would also like to thank Jun, Andrew, Joon, Charlie, and J.C. for putting up with me over the past several months.

Lastly, my greatest appreciation goes to my father, Choon.S. Lee, my mother, Myunghi. K. Lee, and my sister, Christina for their strong support of my education.

This work is dedicated to the memory of my grandfather Ji-Rak Kim and his example of enduring faith and service.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Overview

Good user documentation is critical to the success of most computer softwares. User documentation will make the difference as to whether a program can be used effectively or not. In some cases, if the program is not documented properly, it may not be usable at all. Despite this awareness, inadequate software documentation for users has been and still is a widely acknowledged problem in the computer industry. In recent years, this awareness has stimulated document producers to consider the usability of software documents in their design and evaluation phases. Designing for usability is to produce a software document that is useful and easy to use. The principle objective of this thesis was to address useful methods to evaluate usability of software documents.

1.2. Rationale

User documentation is the hard-copy or on-line documentation provided with a software application program. It tells the user how to get the program up and running, trains the user in its operation, and provides reference information once the user becomes experienced. Its main objective is to help the user make the program work and to use it as easily as possible. For this reason, it is only sound to give as much attention to user documentation as to the software program itself.

In the computer industry, good documentation promotes sales. Because good user documentation makes programs usable, there is more interest in making it good, or at least making it look good. Simpson and Casey (1988) have observed that user documentation is increasingly becoming a selling point for software products.

On the other hand, poor user documentation costs software publishers in time, in extensive service, and in customer dissatisfaction. Even though the documentation provided with software products is much better today than it was in the past, many technical writers, editors, and researchers (e.g., Brockman, 1990; Schriver, 1989;

Simpson and Casey, 1988; Mills and Dye, 1985; McGhee, 1984; Price, 1984) are of the opinion that the documentation provided with most computer programs has not been distinguished by its completeness, quality, or ease of use and that there is much room for improving usability of user documents. Therefore, human factors evaluations of software products and accompanying user documentation must be conducted so that developers can be certain that the target user population can learn to use the product with minimum of difficulty and be able to perform the intended tasks efficiently.

Several methods of evaluating usability of software and its documentation are found in the literature. The objective of this thesis was to review these methods and apply several of them to evaluate the usability of a hard-copy user documentation of a fault tree computer program, IRRAS (Integrated Reliability and Risk Assessment System), which was developed for the Nuclear Regulatory Commission. This program is intended to be used by safety professionals to identify faults which may lead to an undesirable event such as an accident within a system function. The reader is referred to the Fault Tree Handbook by the Nuclear Regulatory Commission (1981) for basic fault tree concepts.

1.3. Organization

First, a literature review is presented on the history and evolution of software documentation, causes of poor documentation, usability of user documentation, and the latest usability testing methods. Next, the method section describes a two-phase experiment to evaluate the usability of the IRRAS user documentation. The purpose of the first phase was to evaluate the existing IRRAS user documentation using the critical incident technique to assess problematic areas of the document by testing subjects on designed subtasks. Then, the results from this first phase were implemented to the redesign of that particular part of the user documentation to improve its usability. The second phase of the experiment compared the two versions of the IRRAS documentation (draft IRRAS version and the modified version) for differences in their usability. A 2X8 mixed factor experimental design was used to test for any difference in task completion time and number of errors in user performance. Additionally, a questionnaire was used to assess any difference in subjective evaluation of the two user's guides. The between-subjects factor represents the two versions of the IRRAS user documentation. The within-subject (or repeated measures) factor represents eight subtasks which each user performed. The following sources were tested for statistical significance: Documentation (D), Subject (S/D), Subtask (T), Interaction of documentation and subtask (D x T), and Interaction of Subtask and Subject (T x S/D).

1.4. Experimental Objectives

In phase 1 of this study, I investigated whether the critical incident technique is useful to identify documentation problems in the draft IRRAS user documentation. These findings were implemented in the redesign of the IRRAS user documentation. In phase 2 of this study, I investigated whether the usability of the modified IRRAS user documentation is better than that of the draft IRRAS user documentation.

When usability measures for two versions of the IRRAS user documentation (draft version and the modified version) were compared along eight different subtasks, the following five alternative hypotheses were assumed for testing.

$\mathbf{H_{1a}}$) The modified version of the IRRAS documentation is expected to result in fewer errors.

$\mathbf{H_{1b}}$) Some subtasks are expected to result in fewer errors than other subtasks.

$\mathbf{H_{2a}}$) The modified version of the IRRAS user documentation is expected to result in shorter subtask performance times.

$\mathbf{H_{2b}}$) Some subtasks are expected to result in shorter task performance times than other subtasks.

H_{3a}) The modified version of the IRRAS user documentation is expected to result in shorter document reading times (as a portion of a subtask completion time).

H_{3b}) Some subtasks are expected to result in shorter document reading times than other subtasks.

H_{4a}) The modified version of the IRRAS user documentation is expected to result in fewer help requests sought by the subjects from the experimenter.

H_{4b}) Some subtasks are expected to result in fewer help requests than other subtasks.

H_{5a}) The modified version of the IRRAS user documentation is expected to receive better subjective evaluations on usability.

H_{5b}) Some subtasks are expected to receive better subjective evaluations on usability than other subtasks.

2. LITERATURE REVIEW

2.1. Overview

The literature review covers information relevant to the stated objective of this study. A brief history and evolution of software user documentation is covered. Then, modern test methods used to evaluate software documentation are reviewed.

2.2. History and Evolution of User Documentation

Documentation design (in general, not just for computer softwares) is an interdisciplinary area of inquiry with rich historical roots. It draws on a matrix of theory and research about how people produce and use text, particularly how they read, write, understand, and are motivated by text. At its heart, document design is concerned with how writers can most effectively find ways to provide readers with texts they can understand and use easily, effectively, and efficiently.

According to Schriver (1989), the knowledge that document designers need to draw on has been expanding rapidly in the past ten years. Figure 1 shows Schriver's (1989) characterization of the relevant theory and research influencing document design as described earlier by Felker (1980). As shown, early conceptions of document design highlighted the importance of composition, cognitive psychology, instructional design, readability, human factors, typography and graphic design, and psycholinguistics. Figure 2 shows the characterization of current theory, research, and practice as described by Schriver (1989). The field has been enlarged by also drawing on rhetoric, social psychology, reading comprehension, human-computer interaction, computer technologies, discourse analysis, and cultural studies.

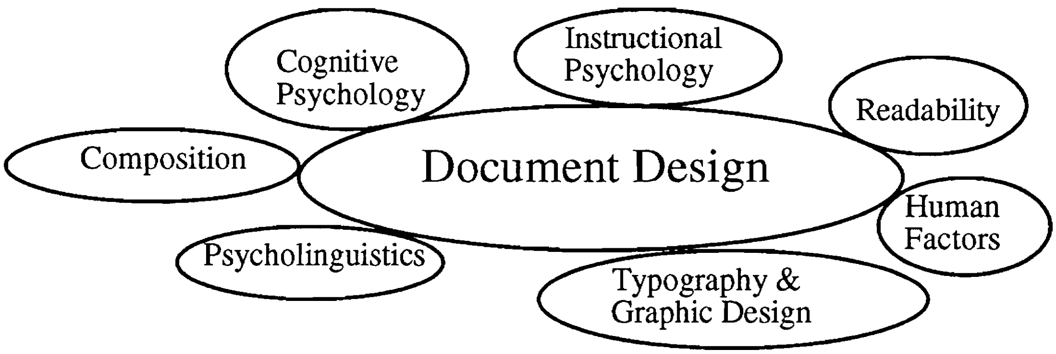


Figure 1. Relevant theory and research influencing document design ten years ago (derived from Schriver, 1989)

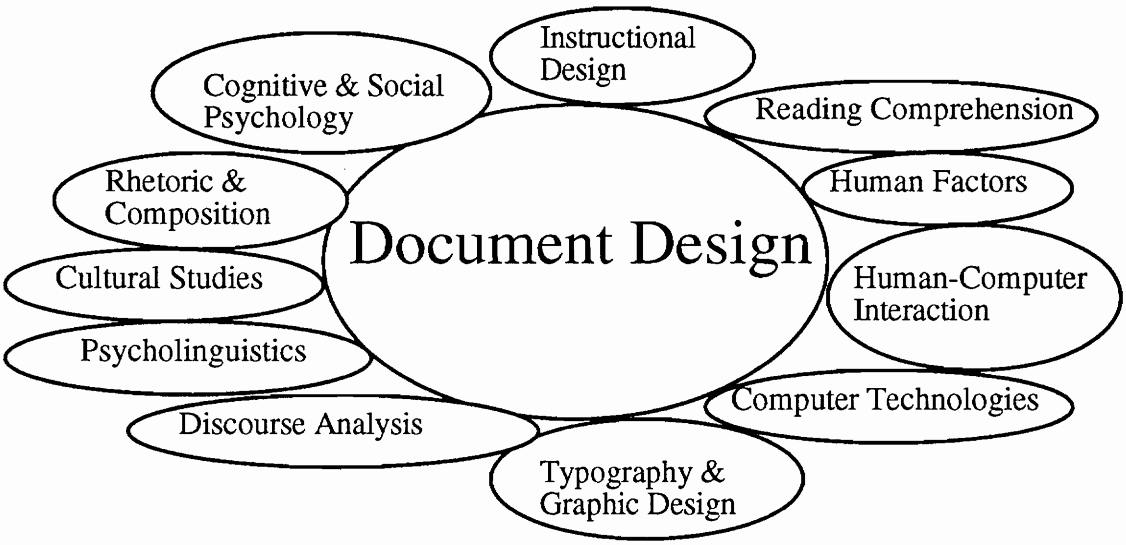


Figure 2. Relevant theory and research influencing document design today (derived from Schriver (1989)

For the interest of this thesis, the growth of computing technologies altering document design over the past ten years is notable, particularly the influence of human factors and instructional design issues. The explosion of the computer industry has not only allowed innovative ways to store, transfer, and retrieve information, but it has opened up new ways to think about the structure of information itself. Emerging technologies have simultaneously created challenges for research in human-computer interaction and user-interface design, spotlighting the importance of developing sophisticated theories of user-centered design.

This challenge, however, has not been enthusiastically embraced by the software and documentation producers in the industry as evidenced by many poor documents still being produced today (Brockmann, 1990; Simpson and Casey, 1988; Wright, 1987; McGehee, 1984). There are several common reasons why the creator of a computer software may develop poor supporting documentation.

2.3. Causes of Poor Documentation

From a documentation editor's perspective, the causes for producing poor documents are as follows. First, much user documentation is written at a wrong level. The usual problem is that the documentation tends to be too technical, too complex, and assumes the user to know too much, especially if documentation is prepared by programmers or other technical professionals. Often, the document is written in the technical language of the author, thus making it unsuitable for the end user. Secondly, many user documents suffer from "multiple-author syndrome" causing confusion and frustration for the readers. The authors of different parts of the document may have contrasting styles, use terms differently, discuss the same topics, cover topics at different depths and breadths, and so forth. Lastly, inconsistent terminology, the practice of describing the same thing with different words, is found in many documents. Similarly, the use of same word to mean different things is also used.

Another common reason for producing poor user documentation is that it is often developed late in the product development cycle or even after the software itself has been released (Mills and Dye, 1985; Mosier, 1984). Due to the pressures of schedules and deadlines, factors such as adequate planning, quality control, user field testing can easily be ignored in the process of developing documentation.

Additionally, a piece of documentation, for example, a user's manual, is too often thought of in isolation and regarded solely as a writing project. To develop proper documentation, considerable front-end work is required beforehand; writing should be the last thing done (Brockmann, 1990; Simpson & Casey, 1988; McGehee, 1984; Price, 1984). Objectives of documentation must be defined. A plan must be developed. The types of documentation relevant for the software project must be chosen. Various user groups must be specified, and documentation must be tailored to meet their needs. Then, finally, the actual writing should be done.

Simpson and Casey (1988) emphasize that when developers assume sole responsibility for developing documentation without ever consulting end-users, the likelihood of producing an inadequate documentation is high. In recent years, as software has reached the mass marketplace and has come into the hands of many novice users, there has been an increasing design emphasis on ease of use. Inadequate documentation ignores this whole philosophy. Common problems are the use of unnecessarily complex language, incomplete coverage of topics, and a tendency to assume that the reader knows more than should be expected. These problems can also result from developer's view of such documentation as secondary and not worthy of monetary expenditure, personnel, and time resources necessary to produce a first-rate product.

2.4. The Value of Usability Testing

A software system undergoes extensive testing as it is being developed. The three most common forms of testing are component or module testing, system or integration testing, and usability testing. In component or model testing, programmers test the code to see if each component of the system is working correctly. In system or integration testing, the entire system is tested for quality of performance, in keeping with its design objectives. In usability testing, test subjects are used as end users to exercise the major functions of the system. Its test results are weighed against the usability requirements that were established in the early stages of system development. Its main objective is to have the system meet the needs and expectations of the people who will be using it.

The focus of usability testing differs from that of component or system testing. It evaluates the system from the outside in, through the end user's eyes. Component and system testing evaluate the system from the inside out, through the programmer's eyes. With this distinction, software systems can be thought of as consisting of two subsystems - the user subsystem and the programming subsystem as shown in Figure 3.

The user subsystem consists of many elements, including user training, the user/software interface, and documentation. The programming subsystem consists of program code and internal documentation such as requirements and design specifications. Figure 3 shows the interaction between the two subsystems. The focal point is the user/software interface. Through this interface, users input instructions to the software, and the software reciprocates results such as error messages, prompts for additional information, and documentations.

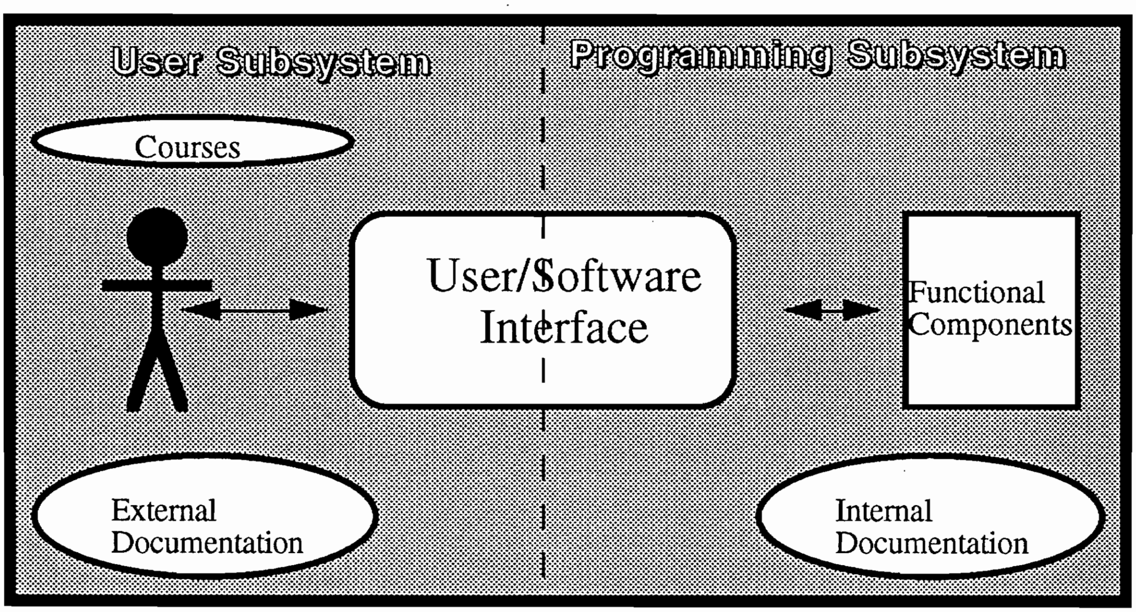


Figure 3. Software subsystems (Adopted from Casey, 1984)

Usability testing is the vehicle by which software or information developers can determine whether or not a product's software and associated documentation are easy to use. The end-user interface is that part of the software which is evaluated for its ease of use characteristics. The documentation associated with the software can be on-line documentation and/or hard copy documentation. Usability tests are conducted for each of these areas of a product. For example, a usability evaluation may be conducted to assure the ease of using the hard copy documentation.

Research in text evaluation is geared toward developing, refining, and testing alternative methodologies for assessing the effectiveness of text. Text-evaluation studies began in the 1930s, mainly focusing on generating formulas to assess the relative readability of text (Spivey, 1987). Today, the gaining popularity of usability testing in industry is fostering the development of ways to incorporate text evaluation into all phases of the document-development cycle, and sophisticated formative and summative evaluation procedures are now becoming commonplace (Schriver, 1989). Formative evaluation consists of series of evaluations throughout various stages in the development of a documentation. Summative evaluation is performed on a tentatively finished document with the intent of improving it through an iterative process.

Carefully designed and conducted usability tests can yield information about everything from particular words and sentences that are hard for the user to comprehend to difficulties in the organization of the document. For example, Winbush and McDowell (1980) performed usability testing on 6 computer manuals in the International Business Machines (IBM) labs in Rochester and identified 12 different types of problems. These included procedures inadequately explained, undefined terms, ambiguous explanations, confusing organization, an inadequate index, and missing examples. When these problems in the IBM manuals were corrected, Winbush and McDowell (1980) stated that the usability of the manuals was significantly improved.

2.5. Content of Usability Testing

The type of tasks that users perform in a usability test depends on the objectives of the test, the type of documentation being tested, and the potential users.

2.5.1. Objectives of the Test

Testing at different times in the development cycle (e.g., using formative evaluations) has different purposes and yields different information. A test can involve a user reading documentation one page at a time while following the instructions. This may be particularly helpful early in the development cycle since it will reveal problems of content or of style in the documentation. Alternatively, tests may have the user perform a typical task using the documentation. This test should be most informative later in the documentation development, since it will reveal whether the documentation is complete and easy to use.

2.5.2. Types of Documentation Tested

The type of tasks the users perform also depends on the purpose of the documentation. According to Simpson and Casey (1988), the larger and more complex the program, the more likely it will need several separate documentation components to fulfill the various needs of its audience. Usually, these documentation components serve two functional purposes: tutorial and reference.

A tutorial is a teaching tool whose intended audience is a group of readers who have never used the program being documented but who will need to use the program in the near future. The tutorial starts with the assumption that the reader knows nothing about the topic, and then it explains the topic simply and logically, building on the reader's growing comprehension (Grimm, 1982). It does not detail every option; instead, it provides a foundation on which the reader can build skills. This journey dictates the structure of the tutorial from simple to complex. The main objective of the tutorial is for the user to read it to *learn.*

As Major (1989) points out, a reference manual is a collection of information organized so that specific information can be located as quickly and effortlessly as possible. It is not meant to be read from cover to cover. The intended audience of the reference is a group of readers who are familiar with the topic and who have a working knowledge of at least the context of what is being documented. This audience needs specific information about some detail of their work.

Whereas the tutorial is structured to lay a foundation and then build upon that foundation, becoming increasingly complex, the reference manual is designed for ready access to a specific detail or group of details. The reader of a reference manual is reading to *do*.

Tutorials generally lend themselves to tests in which the user reads through the documentation while performing a task that is described in the documentation. For example, the instruction for constructing an entire fault tree diagram can be tested by observing a user actually perform this on the computer screen. For this task, the user needs to read through the instructions.

Reference materials, on the other hand, lend themselves to task oriented activities. For example, a typical task such as constructing only a tree branch or a simple event can easily be tested.

2.6. Types of Usability Testing

The choice of a method to test for the usability of a documentation depends on what type of data are desired. Many types of data can be collected in usability tests, and generally more than one type is collected in any given test (Wright, 1985). The type of data collected depends on what is going to be done with the data and how they are going to be analyzed. In other words, if a criterion has been established for the test, data should be collected to determine whether the criterion has been met. For example, if constructing a fault tree on a computer display is supposed to take no more than 30 minutes, then in the test of the documentation for constructing a fault tree, time has to be recorded.

Three main classes of information are collected in usability tests. They are logs, objective measures, and subjective measures. In addition to these measures, the critical incident technique has been used in the literature to evaluate software documentation.

2.6.1. Logs

Several different kinds of logs can be recorded in usability testing: observational logs, keystroke logs, and verbal protocols.

Observational logs are useful for measuring many different types of human behavior (Hutt and Hutt, 1970). Observational logs are written or coded records of the users' actions that are pertinent to performing a task. For example, an observational log could contain information about when and for how long the users looked at the documentation, the keyboard, and the computer display. The amount of detail in the log would depend on the purpose of the test. Observational logs can be made during the test or after the test is completed, from videotapes of the test.

Keystroke logs have been used extensively to find patterns of responding in skilled typewriting (Mills and Dye, 1985; Gentner, 1983). Keystroke can be used for the same purpose in usability testing and can be recorded automatically in tests that involve computers. These logs contain the exact sequence of keystrokes as well as the timing of the strokes.

Neal and Simons (1984) developed the Playback experimental methodology which uses keystrokes to evaluate usability of software and software documentation. The central idea of Playback is that, while a user is working with the system, the keyboard activity is timed and recorded by a second computer. This stored log of activity is later played back through the host system for observation and analysis. Thus, Neal and Simons (1984) claim that Playback methodology is noninvasive in that the data-collection programs are external to the product being evaluated, and the method is nonintrusive because the data collection does not infringe upon the user's activities.

The Playback method was used to evaluate IBM BASIC computer language and its documentation (Neal and Simons, 1984). Six iterations of this study were conducted, with improvements in both the program system and tutorial being made after each test. In almost all cases, improvements in user performance (e.g., fewer errors and less time) were achieved with new versions of the system and tutorial.

Verbal protocols, "thinking aloud", provide useful information in some situations about how an individual is performing a task (Ericcson and Simon, 1984; Nisbet and Wilson, 1977). For example, Carroll and Mack (1983) used verbal protocols to study how people learn to use a word processor. Carroll and Mack (1983) found that users actively learn by constructing rules and interpretations for what occurs. In usability tests, verbal protocols can be provided by users while they are performing a task. These protocols can be recorded on audio or videotape so that they can be studied later. Verbal protocols provide information about what the user is thinking while reading a document or performing a task, information that is unavailable from performance measures alone.

Logs are useful for finding patterns of behavior that may not have been anticipated prior to the study. These patterns may indicate problems with the product or indicate problems with the product or indicate how the users are accomplishing their tasks. For example, if users continually flip between two different pages in the documentation or if they get "hung up" in a fixed set of keystrokes, this might indicate problems with the product. Patterns in verbal protocols may reveal how the user is performing the task. For example, through protocol analysis, Flower, Hayes, and Swarts (1983) found that in order to understand a set of government regulations, readers restructured the information in certain ways. These restructurings revealed ways the regulations could be rewritten to make them easier to understand.

These different types of logs can be used as sources of information themselves, or they can be used as means to derive the objective and subjective measures.

2.6.2. Objective Measures

In usability tests, the objective data collected are standard performance measures used by human factors specialists and experimental psychologists (Brockmann, 1990; Simpson and Casey, 1988; Foehr, 1986; Casey, 1984). These include error measures, time measures, and measures of the amount of help required from the documentation. These measurements can be recorded either during the test or later, from the logs or recordings.

For the error measures, both the number of errors committed while performing a task and when the errors occurred are important for evaluating the usability of a document. The time taken to complete individual parts of a task is helpful for isolating difficult aspects of the task. Times are particularly helpful if performance is error free. One measure of the amount of help required is the number of "help" calls (when user asks the evaluator for help). This measure provides an index of how difficult a task is. More importantly, the help calls indicate where difficulties are in the task.

The objective data are analyzed by determining when and how often task errors and help calls occurred. The number of errors and help calls attributable to different parts of the software and documentation could depict difficulty levels of those parts. The average time taken to complete different parts of the test can also reveal relative difficulty of certain parts, especially if number of errors and help sought are few.

2.6.3. Subjective Measures

Subjective data from usability tests are measures of users' opinions about the software documentation. These data come from questionnaires and interviews, both of which are widely used methods in consumer research (Bethke, 1983). Written questionnaires can contain rating scales or questions about user opinions. Interviews and oral debriefings can also be used to find users' opinions about the documentation. These are usually conducted after the the usability task is completed by the subject.

From the subjective data, individual rating scales can be averaged across users to get an impression for that scale. This will reveal areas that users found satisfactory and unsatisfactory in the documentation. For example, users may like the general organization but dislike a particular section. The data from the interviews or oral debriefings may also serve to be valuable. Users may point out specific problems or recommendations that are unanticipated.

2.6.4. Critical incident technique

Gould and Lewis (1985) propose a software design and evaluation method that incorporates three design principles: early focus on users, empirical measurement, and an iterative process into the design of software. Del Galdo, Williges, Williges, and Wixon (1986a) state that the two most popular methods of design and evaluation of software products, namely logging and questionnaires, fail to take into account all the design principles specified by Gould and Lewis (1985). Although logging method incorporates end-users and can be used iteratively in the design process, it forces designers to speculate the cause of user problems. Questionnaires can also be used iteratively, but are given to the user at the end of a task. This may burden the user with greater memory load since the user is forced to verbalize information collected in retrospect. In addition, since the questions may reflect designer's concerns only and be presented in a leading context, information on missing features or unique problems might be omitted.

Another method that has been used, but less frequently than logging or questionnaires, is the critical incident technique. Del Galdo and others (1986a) suggest that this method shows promise for interface design and evaluation as recommended by Gould and Lewis (1985). This technique has been used to identify common features or elements of the 'critical incidents' in order to classify design deficiencies. Critical incidents are defined as human error or equipment failure that did or could have had unsatisfactory results. Conversely, critical incidents can also be categorized into number of successes.

As reviewed by del Galdo and others (1986a), this method was first used by Fitts and Jones (1947) to analyze pilot error. Cooper (1982) used it to identify and classify errors made as well as good practices performed during anesthesiology procedures. Andersson and Nilsson (1964), in a study to determine the job and training requirements of store managers, showed that critical incident technique is both reliable and valid for certain applications. Dzida (1978) claims that the critical incident technique is a feasible method to evaluate human-computer interfaces and to apply that information into design requirements.

Del Galdo and others (1986a) reported critical incidents encountered in using the on-line and hardcopy documentation of a prototype electronic conferencing utility. Fourteen hardcopy problems, 12 on-line problems, 4 hardcopy assets, and 5 on-line assets were found and arranged in descending order from most critical to least critical by frequency of critical incidents associated with each problem and assets. Ties were separated by employing an average severity index which was calculated by averaging the incident severity ratings given by users at the time each incident was reported. This result was presented to the software design team of Digital Equipment Corporation to guide the redesign process. No further experiment was conducted to test the usability of the modified documentation. Although the information collected from this technique may have identified certain usability problems and assets, the reader can only speculate as to what effects these findings might have had on the usability of the redesigned on-line and hardcopy documentation.

2.7. Usability Research

Usability, as discussed so far, concerns how useful a given software documentation is. Usability testing involves evaluation of that specific product as it is being developed or after it has been developed. The evaluation methods that have been discussed up to this point are quantitative and non-experimental. Another way of analyzing usability is in a research rather than an evaluation framework (Wright, 1980). In research, usability can be addressed by comparing two or more different versions of a documentation to determine if there is any difference in usability among the versions that are tested.

Redish, Felker, and Rose (1981) conducted a comparative study of two different versions (draft and new) of Federal Communications Commission rules concerning two-way radios on recreational boats. The draft rules were rewritten and reorganized to devise the new rules to contain only the information that users needed. The results showed that comprehension was superior for the new rules and that they were subjectively rated as much easier to use.

Carroll (1984) compared a minimal manual for word processing with a comparable state-of-the-art commercial manual. The minimal manual was changed in many ways from the commercial manual. For example, repetitions, summaries, reviews, and the index were removed. Vocabulary was simplified, and error-recovery information was added. The results of the study showed that the users of the minimal manual were 40% faster on given tasks than users of the commercial manual.

Comparative studies such as these may demonstrate that one version of a document is easier to use than another version. However, they do not reveal why a particular document is easier to use than another. For example, in the minimal manuals, it is difficult to assess whether the differences in usability between the minimal manual and the commercial manual were due to information that was added, deleted, or both.

To understand what particular aspect of a document facilitated better or worse performance, a systematic research on usability is needed. For example, with the minimal manual, research is needed to determine whether adding or dropping information (e.g., added error recovery or deleted summaries and reviews) facilitated better performance.

In addition to a systematic approach for research, the modifications of the draft versions need to be justified from the user's perspective. The modifications (i.e., rewriting, reorganizing, adding information, and deleting information) of the draft versions in each of the two studies (Redish et al., 1981 and Carroll, 1984) were done without consulting users. This neglects one of Gould and Lewis' (1985) principles of incorporating user's opinion into the evaluation phase early in the product development stage.

For this thesis, a new experiment was conducted to evaluate a draft documentation for the fault tree software, IRRAS. The experiment consisted of two phases.

In the first phase, the critical incident technique was used to identify and categorize poor aspects of the software documentation. The critical incident technique was chosen over logging and questionnaire methods for the similar reasons del Galdo (1986a) chose to use it to identify hard copy and on-line successes and failures. These reasons concur with the design and interface guidelines given by Gould and Lewis (1985). These principles include early focus on users, the use of empirical measurement, and an iterative design process in the design and evaluation of human-computer interfaces. The critical incidents identified and categorized from this first part of the experiment were used to make improved changes to the draft IRRAS documentation.

The second phase compared the draft version of the IRRAS user documentation to the modified version of the IRRAS user documentation. For this part of the experiment, both objective and subjective measures were recorded for further analyses. Four objective measures, namely frequency of errors, subtask completion time, document reading time, and number of personal help requests from the experimenter were recorded on the video tapes. In addition, a questionnaire with nine specific questions was used to assess subjective measures on perceived ease of use.

3. PHASE ONE: METHOD AND RESULTS

In section 3.1, the method to evaluate the draft document using the critical incident technique is described. In section 3.2, the results of the phase one experiment is presented.

3.1. Method

3.1.1. Subjects

Eight subjects who learned basic fault tree concepts in the System Safety Engineering course (ISE 4614) of Virginia Polytechnic Institute and State University (VPI & SU) were used as subjects. The total time commitment by each subject was about two hours. Each subject was paid at a rate of $5.00 per hour, for a total of $10.00. All were required to sign an informed consent form and initialize a sheet of experimental instruction (Appendices A & B)

3.1.2. Materials and Apparatus

The experiment was conducted in the Safety and Environmental Engineering Laboratory of the VPI & SU Human Factors Center. The following equipment was used to conduct the first phase of the experiment. The equipment layout is shown in Figure 4.

Zenith 386-SX computer system
IRRAS fault tree software
Draft IRRAS fault tree software user documentation (Appendix C)
Test room with one-way window
RCA color television & Mitsubishi VCR
Two split-image video cameras
Two camera mounts
Critical incident technique (CIT) questionnaire (hardcopy)
3X5 Cards to answer CIT questionnaire

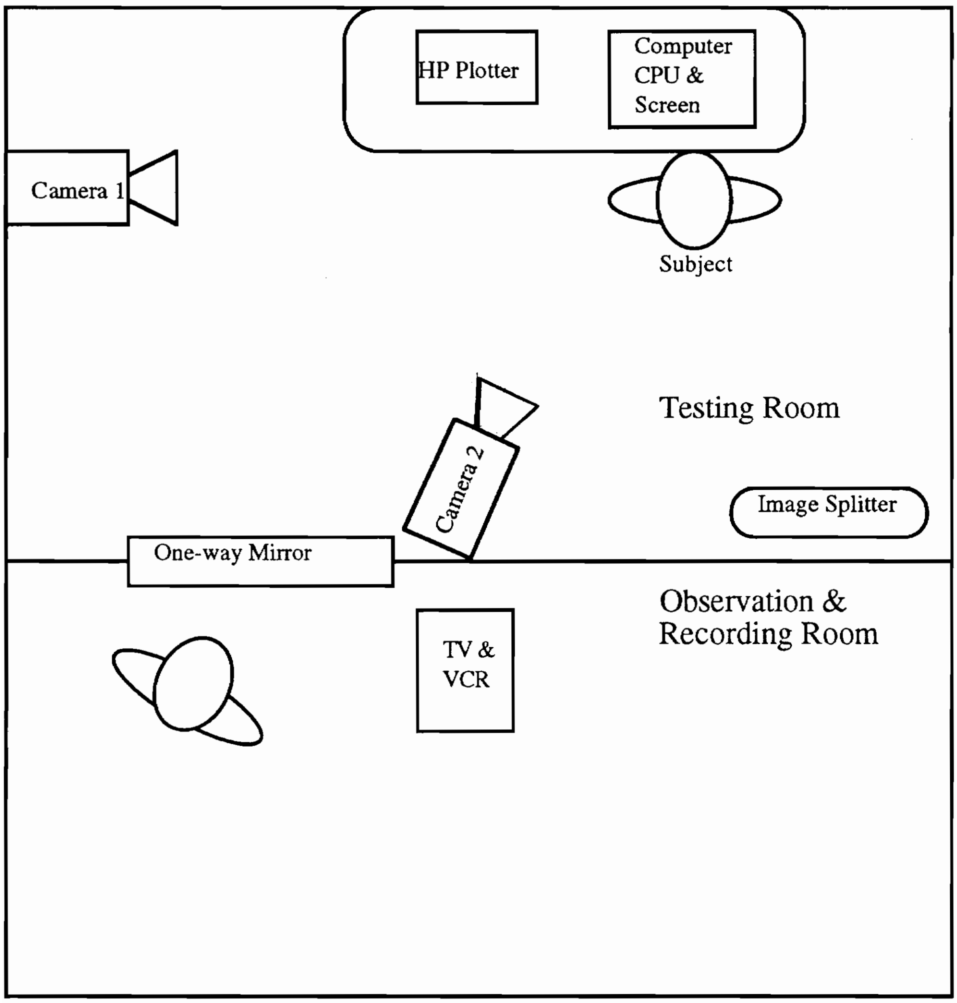


Figure 4. Equipment layout

Each subject used the user documentation in conjunction with operating the computer hardware and software. The critical incident tool that was used is illustrated in Figure 5. The subject recorded his or her response to the questions in the critical incident tool on 3X5 cards. The one-way window allowed the test administrator to make in-the-room observations of the test subject. The television and the video cameras were used to monitor and record the subject's activities and the computer display.

3.1.3 Procedure

The subjects were given a set of instructions from the draft IRRAS user documentation for an 8-part benchmark task to be completed using the IRRAS software. The 8 subtasks were performed in the order of most basic operation to the more complex type since the subjects had to perform these tasks as they learned to use the software. The following is the ordered list of 8 subtasks:

1) picking and placing 8 (out of 16) assigned fault tree symbols from the pop-up menus onto the computer screen,

2) picking and placing multiple copies of a symbol on the screen for any 3 of the 8 fault tree symbols,

3) drawing 6 sets of a horizontal line under a gate symbols,

4) writing symbol description under or next to four assigned symbols on a predrawn fault tree,

5) copy editing 3 sections or symbols on a predrawn fault tree,

6) move editing 3 sections or symbols on a predrawn fault tree,

7) deleting 7 sections or symbols on a predrawn fault tree, and

8) drawing an entire fault tree.

Appendix D shows the products of the subtasks as shown on the computer screen after the subtasks were completed. Each subject was instructed to report incidents (as many as the subject observed) involving the user documentation immediately after the completion of each of the 8 subtasks on 3X5 cards. One card was used for each incidence. An incident was defined as any feature of the document specified for the given subtask

Critical Incident Tool

Indicate the features of this section of the user document which you thought were NOT easy to understand and did NOT readily help you to accomplish your subtask. An undesirable feature can be any aspect of the information presented on the document (e.g., confusing word(s), sentence, phrase, format, wording, structures, appearance, and etc.) Indicate also the features you thought were easy to understand and readily help you to accomplish the given subtask.

Procedure:
1. Write down one desirable or one undesirable feature on the back side of a 3X5 card.

2. On the front side of the card, answer the following question by writing down the number you chose:

How critical was the information of this feature for relatively easy completion of the task? Indicate the criticalness of your incident by circling the number which corresponds to your rating.

CRITICAL--NON-CRITICAL

Extremely	Quite	Slightly	Neutral	Slightly	Quite	Extremely
1	2	3	4	5	6	7

3. If you could and want to change this particular feature of this section of the documentation, how would you change it? Respond on the same side of the card you answered question #2.

4. Repeat steps 1 through 3 as many times as you feel necessary on other cards

Figure 5. Critical incident tool (Adopted and modified from del Galdo et. al., 1986)

which the subject perceived to be easy to understand and readily usable to accomplish the subtask. Or, conversely, an incident was any feature of that same part of the document which the subject preferred to change to make it more understandable and readily usable. Sometimes, subjects reported critical incidents relating to the software instead of the document. Whenever this occurred, they were reminded to report critical incidents that only pertained to the document. The subjects were allowed to work on the benchmark subtasks up to two hours. All subject activities were monitored by the experimenter and by the video cameras.

3.2. Results

According to del Galdo (1986a), the analysis of critical incident data includes three steps. First, the critical incidents are divided into the outcome categories, success or failure. Second, elements (problems and assets) of the critical incidents are identified. Finally, problems and assets are prioritized.

In this study, all of the collected critical incidents were characterized as failure. Not a single success incident was reported by the subjects.

After the outcome of incidents were labeled, incident descriptions for each failure category were examined to determine features of the documentation that led to or caused the failure incident. The incidents that had similar meaning but were worded differently were judged by the experimenter to be in a same problem category.

Some single incident descriptions included multiple problems. For these incident descriptions, the multiple features were identified under the heading of a representative incident. Some incident descriptions were worded in such a manner that no useful information could have been extracted. These incident reports were discarded. For example, one critical incident card read "can't redraw." First, the experimenter could not understand what relevance this "problem" had with the task of picking and placing symbols. Second, the experimenter could not decipher whether the "problem" pertained to the lack of information about "redrawing" in the document or the lack of software's capability to "redraw."

The priority of the problems were first determined by the frequency of the incidents classified within the problem. The problem categories with the highest frequency of occurrence were considered most important. When two or more problem categories were tied in frequency, order was determined by averaging the severity ratings of incidents assigned by the subjects. The severity ratings of incidents within a problem were averaged to determine the average severity rating for each problem.

The summary of critical incidents (problem categories) is shown in Table 1. A total of 150 critical incidents were identified. Among these, the most frequently appearing features included 44 unclear and wrong procedural instructions, 42 ambiguous, undefined, and poorly chosen words and phrases, 27 unfriendly format problems, and 14 confusion over the functions of the three buttons on the mouse.

After the critical incidents were identified and categorized, they were used as springboards to make modifications on the draft document. The features that contained error were corrected. For example, the name of the undeveloped event symbol was corrected from C-Event to U-Event to keep it consistent throughout the document. The paragraph format was changed to step-by-step format with numbered instructions. Even the steps that were assumed to be obvious, and therefore neglected in the draft document, were stated and presented in the modified document. The language of the step-by-step instructions was changed from passive voice to active voice. The statements describing subject's input and the computer's output were presented separately to reduce confusion The ambiguous words and phrases were defined and clarified. The functions of the mouse buttons were more clearly presented with pictures and highlighted identifiers. The subtask sections were separated with their respective headings. These represent the major changes made to the draft IRRAS document. The modified IRRAS document is shown in Appendix E.

TABLE 1
Summary of Critical Incident Findings

Subtask 1 (Picking and Placing Symbols) Problems:

Problems	Frequency	Average Severity
1. Ambiguous terms and phrases (poor word choice, undefined words, incorrect tenses)	9	4.78
2. Information regarding 3-button functions is confusing (function of each button not clear)	7	3.29
3. "C Event symbol should be U Event" (document error)	5	1.6
4. Irrelevant information (e.g., delete, icon placement) is unneeded or should be given later	4	2
5. Unfriendly format (cluttered paragraph, no highlight) Should use step-by-step format	3	3
6. No warning about not to drag the mouse with a button held down (Macintosh stereotype)	2	2
7. No warning about location of menu obstructing the view of symbols	2	2.5
8. Hyphen is awkward and grammatically incorrect in "right- button". Badly needs editing	2	5
9. Figures are not near the text	1	3

Subtask 2 (Placing Multiple Copies of Symbols) Problems:

Problems	Frequency	Average Severity
1. Unfriendly format (cluttered paragraphs, ambiguous wording, too verbose,no heading) Should use step-by-step format	6	3.33
2. Ambiguous terms and phrases (poor grammar, poor word choice, redundant phrases, "ghost: is undefined)	5	3.6
3. Information regarding 3-button functions is confusing (should provide a diagram)	1	1
4. No warning about what will appear on the screen (warning should appear before a discussion on feedback	1	3
5. No information on how to get out of a menu	1	7

Subtask 3 (Drawing Lines Adjacent to Symbols) Problems:

Problems	Frequency	Average Severity
1. Information regarding 3-button functions for drawing a line is confusing (e.g., finish drawing a line) Should provide a diagram and be explained clearly	6	2.5
2. Ambiguous terms and phrases (e.g., "rubber band") Should use active not passive voice	5	2.8
3. Unfriendly format (e.g., cluttered paragraph) Should use step-by-step format	3	3
4. Irrelevant information (e.g., "delete" and "copy" are mentioned too soon and not explained)	1	4

TABLE 1 (Continued)
Summary of Critical Incident Findings

Subtask 4 (Writing Texts Near Symbols) Problems:

Problems	Frequency	Average Severity
1. Poor instructions on how to end a line of text and begin a next line of text (e.g., wrap around)	7	1.55
2. Unfriendly format (cluttered paragraph, no title) Should use step-by-step format Should highlight important words. Should begin with description of task followed by procedure	7	2.57
3. Ambiguous terms and phrases (Should use active voice)	6	3
4. No waning that the text will be centered (Should inform this before the procedure)	3	2.33
5. Irrelevant information (e.g., "right button also puts up text menu, use of quotes around menu)	3	3.33

Subtask 5 (Copying Symbols, Lines, and Texts) Problems:

Problems	Frequency	Average Severity
1. "Reference Point" is undefined, unclear, and misleading (Should define it and instruct how to use it)	9	1.11
2. Unfriendly format (e.g., cluttered paragraph, too verbose, no title) Should use step-by-step format	4	3
3. The selection box, or "rubber-band" box, is unclear. The procedure to create this box is unclear	3	2
4. Parameters to be included in the "region to be copied" is not clearly explained (Should warn the user to create the box with sufficient space to include the desired features)	2	1
5. User is not informed ahead that bottom left of the screen displays feedback information	1	1
6. User is not informed ahead that if button is pressed too hard or too long, the desired response does not occur	1	1

Subtask 6 (Moving Symbols, Lines, and Texts) Problems:

Problems	Frequency	Average Severity
1. "Reference Point" is undefined and how it's used is unclear	8	1.62
2 The selection box is unclear	7	2
3. Unfriendly format (e.g., cluttered paragraph, too verbose, no title) Should use step-by-step format	6	2
4. Parameters to be included in the "region to be copied" is not clearly explained (Should warn the user to create the box with sufficient space to include the desired features)	1	2
5. No warning about pressing button too long which causes editing problems (Should emphasize depress & release)	1	3

TABLE 1 (Continued)
Summary of Critical Incident Findings

Subtask 7 (Deleting) Problems:

Problems	Frequency	Average Severity
1. Parameters to be included in the "region to be copied" is not clearly explained (Should warn the user to "box" in all of the the symbol including stems or else the symbol will not be deleted	6	2
2. Unfriendly format (cluttered paragraph, too verbose, no title) Should use step-by-step format	3	2.25
3. No information on how to cancel a delete function after a box has been drawn	3	2
4. No warning that when a portion of a menu is deleted that it can be redrawn (Should inform the user ahead of time so that the user can expect it)	3	4
5. User is not told to look at the status displayed on the lower left corner of the screen	1	3
6. Unclear that second click of the right button erases the boxed part of the fault tree	1	6

4. PHASE TWO: METHOD AND RESULTS

After the modified IRRAS user document was developed the second phase of the experiment was conducted to compare the usability of the draft document and the modified document.

4.1 Method

4.1.1. Subjects

Sixteen new subjects (none from the first phase) were used for this phase of the experiment. All of these subjects, students at Virginia Tech, were recruited because of their familiarity with basic fault tree principles learned in the VPI & SU System Safety Engineering course (ISE 4614). The total time commitment by each subject was about two hours. Each subject was paid at a rate of $5.00 per hour, for a total of $10.00. All were required to sign an informed consent form and initialize a sheet of experimental instruction (Appendices A & F)

4.1.2. Materials and Apparatus

This phase of the experiment was also conducted in VPI & SU Safety and Environmental Engineering Laboratory. All of the equipment listed in section 3.1.2, except for critical incident technique questionnaire and 3X5 cards, were used. Also, the following additional equipment was used.

Hewlett Packard 7475A Plotter
IRRAS fault tree software
Modified IRRAS fault tree software user documentation
Timex chrono/lap watch and Meylan stop watch
Questionnaires for perceived ease of use

Each subject used the user documentation in conjunction with operating the computer hardware and software. The one-way window allowed the test administrator to make in-the-room observations of the test subject. The television and the video cameras were used to monitor and record the subject's activities and the computer display. The video tapes were played back at later times for the analyses of task completion times and errors. Questionnaires were used to assess subjective ratings on the usability of the documentation. Figure 6 shows the questionnaire used to assess subjects' perceived ease of use.

4.1.3. Procedure

4.1.3.1. Experimental Tasks. First, to avoid subject confounding due to the between-subjects factor, document (D), the subjects were matched with the two versions of the IRRAS document. This was done two weeks prior to this phase of the experiment. To match the subjects' reading comprehension level, the subjects were matched according to their performance in the reading-ability test. The reading-ability test consisted of reading several passages and answering 17 comprehension questions (Appendix G). The reading-ability performance was based on the number of correct answers. The correct number of answers ranged from 8 to 15. Two subjects who had the same or nearly the same number of correct answers (different by no more than one correct answer) were assigned to two separate groups. As a result, the two groups were matched in number of correct answers (the mean number of correct answers for one group was 10.8 and 10.9 for the other group). All of the subjects reported having at least five years of experience with computers and mouse. One group used the draft document and the other group used the modified document.

Each of the 8 subjects was given a set of instructions from the draft version of the IRRAS user documentation for an 8-part benchmark task (same as the ones specified in section 3.1.3.) to be completed using the IRRAS software. Each of the other 8 subjects was given a set of instructions from the modified version of the IRRAS user documentation for the same 8-part benchmark task to be completed on the computer screen using the IRRAS software. Four objective measures, number of errors, subtask completion time, document reading time, and number personal helps requested, were taken from recorded

1. How easy to use was this part of the user documentation? Choose a number from 1 to 6. One denotes most difficult to use and 6 denotes easiest to use.______

For statements 2 through 8, use the following scale to express your opinion of the statement:

Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree
1	2	3	4	5	6

2. I have found the information in this part of the document to be very accurate. _______

3. This section of the document provided complete information to perform the sub-task. _______

4. This section of the document has extra, unnecessary information. ________

5. It took extra effort to understand the language used in this section of the document. ________

6. The examples and figures were helpful. ________

7. The information was developed in a way that made it easy to learn. ________

8. I have found annoying inconsistencies in presentation or coverage in this section of the document. ________

9. How easy to use is this section of the user document? Circle one of the six choices.

Extremely Easy

Very Easy

Fairly Easy

Fairly Difficult

Very Difficult

Extremely Difficult

Figure 6. Questionnaire for perceived ease of use

video tapes of each experimental session. As soon as the subject started to read the given section of the document, the subject was timed until the subtask was completely and correctly performed on the computer screen. If the subject made an error, the subject was expected to correct the error. After the correct completion of each of the 8 subtasks, the subject was given a questionnaire as shown in Figure 6 to evaluate the part of the document that was just used. After all subtasks and questionnaires were completed, the subjects were asked to rank order the subtasks from easiest to most difficult to perform. In addition, the subjects were given both versions of the IRRAS document and were asked to compare the two versions. All subject activities were monitored by the experimenter and recorded on the video cameras. Appendix D shows the finished subtasks (same as the subtasks in phase 1 of the experiment) as shown on the computer screen.

4.1.3.2. Experimental Design. The experimental design used was a 2X8 mixed two-factor design. The between-subjects factor represents the two versions of the IRRAS user documentation. A between-subjects design was preferred over a within-subject design for the document factor to eliminate confounding caused by document familiarity and previous knowledge. The within-subject (or repeated measures) factor represents the eight subtasks. The experimental matrix is shown in Table 2.

The five dependent variables consisted of the following for each experimental cell:

1) number of errors,

2) subtask completion time,

3) document reading time (portion of the subtask completion time),

4) number of personal helps requested by the subjects from the experimenter, and

5) subjective ratings on 9 questions (Only subtasks 1 through 7 were examined for subjective ratings since subtask 8 (last subtask) did not require learning of new material. The subjects were given the freedom to use the document if they wished.

TABLE 2

Experimental design matrix.

		Documentation (D)	
		Original IRRAS	New IRRAS
Subtasks (T)	1. Picking and Placing Symbols		
	2. Placing Multiple Symbols		
	3. Drawing Lines		
	4. Writing Text		
	5. Copy Editing		
	6. Move Editing		
	7. Delete Editing		
	8. Constructing a Fault Tree		

4.2 Results

A 2 X 8 mixed two factorial analysis of variance (ANOVA) was performed for each of the 5 dependent variables: number of errors, subtask completion time, document reading time, number of helps sought, and subjective ratings (9 questions) of preferred usability. The results of these analyses are presented under the names of their respective dependent variables. As explained earlier, the subjects were matched according to their reading-ability scores and number of years of computer experience to avoid subject confounding due to the use of a between-subjects document factor (see section 4.1.3.1, page 31).

4.2.1. Errors (Hypothesis 1)

Many different types of errors occurred. The most common errors were the following: getting into a wrong menu menu, selecting wrong choices in the menu, not following task instructions in the document, placing more or fewer symbols than instructed, and clicking wrong buttons on the mouse. Hypothesis 2 is restated as two sub-hypotheses as they apply to this dependent variable:

1a) The modified IRRAS document is expected to result in fewer errors than the draft document.

1b) Some subtasks are expected to result in fewer errors than other subtasks.

The mean number of errors is shown in Figure 7. To test hypotheses 1a and 1b, a 2X8 mixed factor ANOVA was performed to determine the significance of the main effect of document and the main effect of subtask. The results are shown in Table 3.

When hypothesis 1a was tested, the between-subjects main effect of document was found to be significant, $\underline{F}(1,14)=23.03$, $p<0.05$. The use of the modified document resulted in fewer errors than the use of the draft document.

When hypothesis 1b was tested, the within-subject main effect, subtask, was found to be significant also, $\underline{F}(7,98)=3.84$, $p<0.05$. As shown in Figure 7, the subjects made more errors in some subtasks than in others. The Newman-Keuls post-hoc analysis ($p<0.05$) revealed that subjects made more errors in subtasks 5 and 4 than in subtasks 1, 2, 6, 7, and 8.

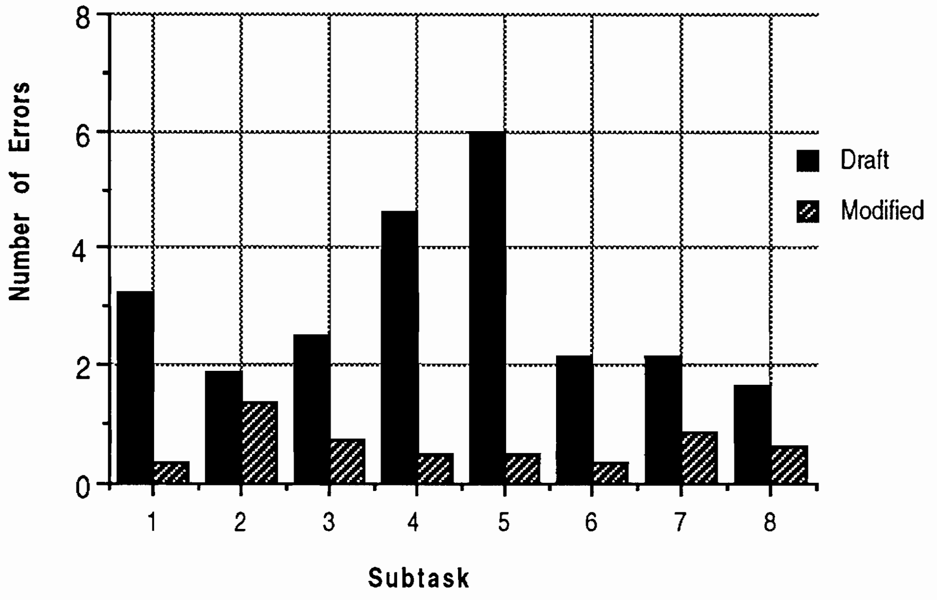


Figure 7. Mean number of errors (Subtasks: 1=Symbols, 2=Multiple Symbols, 3=Lines, 4=Text, 5=Copy Edit, 6=Move Edit, 7=Delete Edit, 8=Fault Tree)

TABLE 3.

Document (2) X Subtask (8) ANOVA Summary Table

Number of Errors

Source	df	SS	MS	F	p
Between					
Document (Doc)	1	175.78	175.78	23.03	0.0003
Subject (S)/Doc	14	106.84	7.63	-----	
Within					
Subtask (T)	7	52.25	8.18	3.84	0.001
Doc X T	7	82.09	11.73	5.5	0.0001
T X S/Doc	98	208.91	2.13	-----	

The interaction effect (Document X Subtask) was found to be significant, $F(7,98)=5.50$, $p<0.05$. Although the main effect of document was found to be significant, Newman-Keuls sequential range test ($p<0.05$) showed that the significant difference only occurred between the document sections for subtasks 1, 4 and 5.

4.2.2. Subtask Completion Time (Hypothesis 2)

Hypothesis 2 is restated as two sub-hypotheses as they apply to the dependent variable, subtask completion time:

2a) The modified IRRAS document is expected to result in shorter subtask completion time than its draft IRRAS document.

2b) Some subtasks are expected to result in shorter subtask completion time than other subtasks.

The mean subtask completion times for 8 subtasks using both versions of the IRRAS manual are shown in Figure 8. To test hypotheses 2a and 2b, a 2X8 mixed factor ANOVA was performed to determine the significance of the main effect of document and the main effect of subtask. The results are shown in Table 4.

The between-subjects main effect of document was not found to be significant, $F(1,14)=0.32$, $p>0.05$. Thus, hypothesis 2a was not supported. As shown in Figure 8, the subjects who used the modified document took just as much time to complete the subtasks as the subjects who used the draft document.

However, when hypothesis 2b was tested, the within-subject main effect of subtask was found to be significant , $F(7,98)=78.31$, $p<0.05$. The Newman-Keuls post-hoc analysis revealed that subjects took more time to complete some subtasks than other subtasks. As shown in Figure 8, the subjects took the most time to accomplish subtask 8 (drawing an entire fault tree). Among subtasks 1 through 7, subtasks 4, 5, 1, and 3 required more time for completion than subtasks 7, 2, and 6.

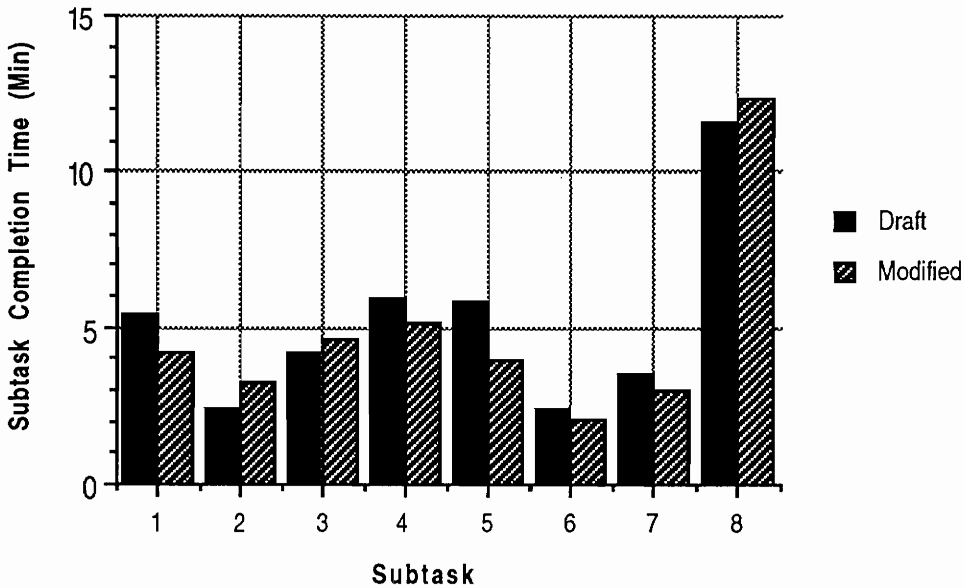


Figure 8. Mean subtask completion time (Subtasks: 1=Symbols, 2=Multiple Symbols, 3=Lines, 4=Text, 5=Copy Edit, 6=Move Edit, 7=Delete Edit, 8=Fault Tree)

TABLE 4.

Document (2) X Subtask (8) ANOVA Summary Table

Subtask Completion Time

Source	df	SS	MS	F	p
Between					
Document (Doc)	1	3.80	3.80	0.32	0.5826
Subject (S)/Doc	14	168.18	12.01	-----	
Within					
Subtask (T)	7	1031.04	147.29	78.31	0.0001
Doc X T	7	27.47	3.92	2.09	0.052
T X S/Doc	98	184.32	1.88	-----	

The interaction effect (Document X Subtask) was not observed to be significant, $\underline{F}(7,98)=2.9$, $p>0.05$.

4.2.3. Document Reading Time (Hypothesis 3)

Hypothesis 3 is restated as two sub-hypotheses as they apply to the dependent variable, subtask document reading time:

3a) The modified IRRAS document is expected to result in shorter document reading time than its draft IRRAS document.

3b) Document sections for some subtasks are expected to result in shorter document reading time than those for other subtasks

The mean document reading times for 8 subtasks using both versions of the IRRAS document are shown in Figure 9. To test hypotheses 3a and 3b, 2X8 mixed factor ANOVA was performed to determine the significance of the main effect of document and the main effect of subtask. The results are shown in Table 5.

When hypothesis 3a was tested, the between-subjects main effect of document was found to be significant, $\underline{F}(1,14)=17.76$, $p<0.05$. As shown in Figure 9, the subjects who used the modified document took less time to read and understand the information necessary to perform the respective subtask than the subjects who read the draft document. This result was observed even though the modified document contained more than twice the number of words (a total of 2110 words) than the draft document (1039 words). All of the seven modified document sections contained nearly twice or more than twice as many words than the draft document. In addition, as shown in the previous section, significantly fewer errors were observed from those who used the modified document than from those who used the draft document. These observations show that the subjects comprehended the information in the modified document in a shorter time although more words were contained in it.

The testing of hypothesis 3b also revealed significant difference in the within-subject main effect, subtask, $\underline{F}(7,98)=78.31$, $p<0.05$. The Newman-Keuls post-hoc

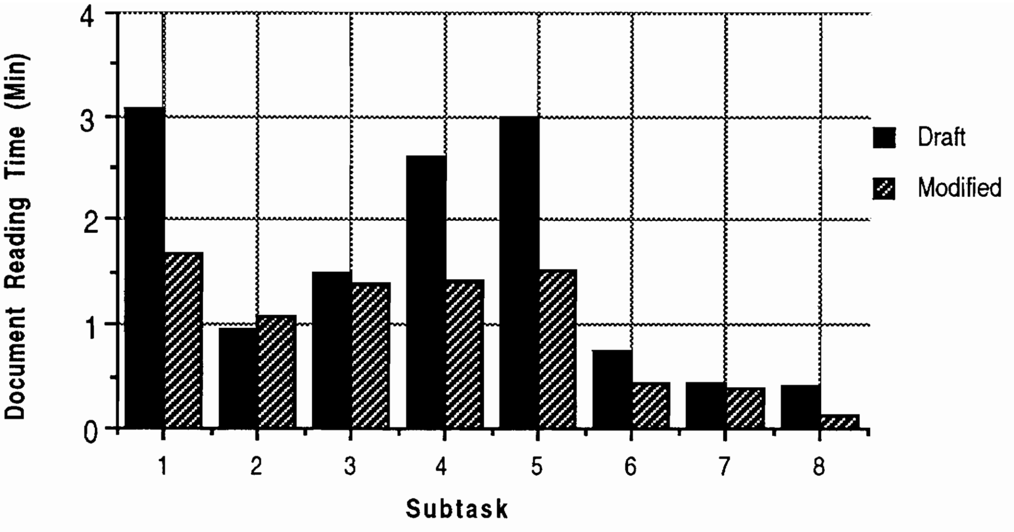


Figure 9. Mean document reading time (Subtasks: 1=Symbols, 2=Multiple Symbols, 3=Lines, 4=Text, 5=Copy Edit, 6=Copy Move, 7=Copy Delete, 8=Fault Tree)

TABLE 5.
Document (2) X Subtask (8) ANOVA Summary Table
Document Reading Time

Source	df	SS	MS	F	p
Between					
Document (Doc)	1	10.69	10.69	17.76	0.0009
Subject (S)/Doc	14	8.43	0.60	-----	
Within					
Subtask (T)	7	80.21	11.46	46.81	0.0001
Doc X T	7	11.92	1.70	6.96	0.0001
T X S/Doc	98	23.99	0.24	-----	

analysis ($p<0.05$) revealed that subjects took more time to read document sections for some subtasks than for those of other subtasks. The subjects took the most time to read document sections for subtasks 1, 5, and 4, then subtask 3, then subtask 2, then subtasks 6 and 7, and then subtask 8.

The subjects took the least amount of time to read the document for accomplishing subtask 8 (drawing an entire fault tree) even though they were given the entire document which included all 7 sections pertaining to subtasks 1 through 7.

The interaction effect (Document X Subtask) was found to be significant, $F(7,98)=6.96$, $p<0.05$. Although the main effect of document showed significant difference, the Newman-Keuls sequential range test ($p<0.05$) further revealed that significant difference in the reading time between the two documents occurred for document sections only pertaining to subtasks 4, 5, and 1. In other words, reading times were not significantly different between the two documents for document sections respective to subtasks 2, 3, 6, and 7 even though the modified document sections contained nearly twice or more than twice as many words as the draft document. However, for document sections pertaining to subtasks 1, 4, and 5, the modified document sections were read in significantly shorter time than the draft document sections even though the modified document sections contained approximately twice as many words as the draft document sections.

4.2.4. Number of Personal Helps Requested (Hypothesis 4)

If the subjects could not accomplish an assigned subtask using only the given section of a document, they were given the freedom to call for personal help from the experimenter. Hypothesis 4 is restated as two sub-hypotheses as they apply to the dependent variable, subtask document reading time:

4a) The modified IRRAS user document is expected to result in fewer personal help requests sought by the subjects from the experimenter than the draft IRRAS document.

4b) Also, some subtasks are expected to result in fewer personal help requests than other subtasks.

The total number of helps sought for 8 subtasks using both versions of the IRRAS document are shown in Figure 10. To test hypotheses 4a and 4b, 2X8 mixed factor ANOVA was performed to determine the significance of the main effect of document and the main effect of subtask. The results are shown in Table 6.

When hypothesis 4a was tested, the between-subjects main effect of document was found to be significant, $\underline{F}(1,14)=4.76$, $p<0.05$. As shown in Figure 10, the subjects who used the modified document sought fewer helps than the subjects who used the draft document. In fact, not a single help request was sought by subjects who used the modified version of the IRRAS document.

The testing of hypothesis 4b also revealed significant difference in the within-subject main effect, subtask, $\underline{F}(7,98)=2.64$, $p<0.05$. The Newman-Keuls post-hoc analysis did not reveal any significant difference among the subtasks. However, the less conservative Least Significant Difference (LSD) post-hoc analysis ($p<0.05$) revealed that subjects sought significant number of helps to accomplish subtasks 1, 4, 5, and 7 while they did not seek any help to perform subtasks 2, 3, 6, and 8.

The interaction effect (Document X Subtask) was found to be significant, $\underline{F}(7,98)=2.64$, $p<0.05$. The subjects who used the modified document did not request a single help from the experimenter for any of the 8 subtasks. Likewise the subjects who used the draft document did not seek any help while they were performing subtasks 2, 3, 6, and 8. However, for subtasks 1, 4, 5, and 7, these same subjects sought help from the experimenter. The Newman-Keuls sequential range test ($p<0.05$) showed that the significant difference in the document effect occurred only between the document sections of subtasks 1 and 4.

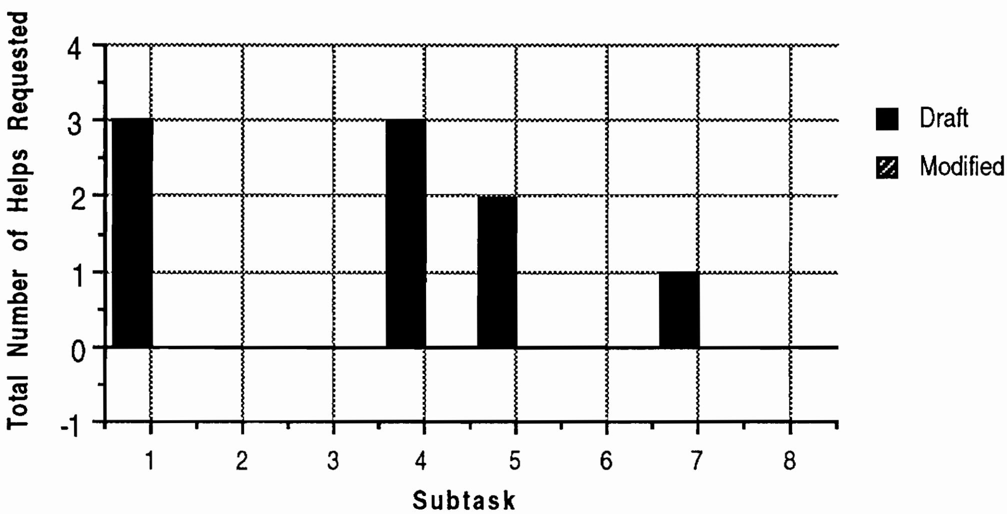


Figure 10. Total number of personal helps requested (Subtasks: 1=Symbols, 2=Multiple Symbols, 3=Lines, 4=Text, 5=Copy Edit, 6=Move Edit, 7=Delete Edit, 8=Fault Tree

TABLE 6.

Document (2) X Subtask (8) ANOVA Summary Table

Number of Personal Helps Requested

Source	df	SS	MS	F	p
Between					
Document (Doc)	1	0.63	0.63	4.76	0.0466
Subject (S)/Doc	14	1.86	0.13	-----	
Within					
Subtask (T)	7	0.80	0.11	2.64	0.0152
Doc X T	7	0.80	0.11	2.64	0.0152
T X S/Doc	98	4.27	0.04	-----	

4.2.5 Preferred Subjective Ratings (Hypothesis 5)

After the subjects finished each subtask, they were asked to fill out a questionnaire with nine questions. These questions were asked in the form of a direct question or a statement. The subjects answered each question by choosing one score from a six point rating scale. The description of the anchored rating scale is illustrated in Figure 6. The data for each of the nine questions were analyzed separately. A 2 X 8 (Document X Subtask) mixed factor ANOVA was performed for each question. These results are presented separately under their respective subheadings.

4.2.5.1. Question One (Hypothesis 5.1). Question one asked the following: "How easy to use was this part of the document?" The subjects selected 1 out of 6 possible choices in the rating scale ranging from 1 to 6. Only the two extreme scores were anchored. A score of 1 denoted most difficult to use and 6 denoted easiest to use. It was left up to the subject's judgment to attach meaning (anchor) to the other four scores. Hypothesis 5.1 (hypothesis 5 pertaining to question 1) is restated as two sub-hypotheses as they apply to the dependent variable, rating score:

5.1a) The modified IRRAS document is expected to result in better subjective rating than the draft IRRAS user document. For question 1, the specific meaning of this hypothesis is that subjects are expected to rate the modified document to be easier to use than the draft document.

5.1b) Document sections for some subtasks are expected to be perceived as easier to use than those of other subtasks.

The mean subjective rating scores of both documents are shown in Figure 11. The results of the ANOVA are shown in Table 7.

When hypothesis 5.1a was tested, the between-subjects main effect of document was found to be significant, $\underline{F}(1,14)=10.51$, $p<0.05$. The modified document was rated to be easier to use than the draft document. The mean score for the modified document was slightly higher than five. The mean score for the draft document was slightly higher than

Question 1: How easy to use was this part of the document?
(Rating Scores: 1=Most Difficult, 6=Easiest)

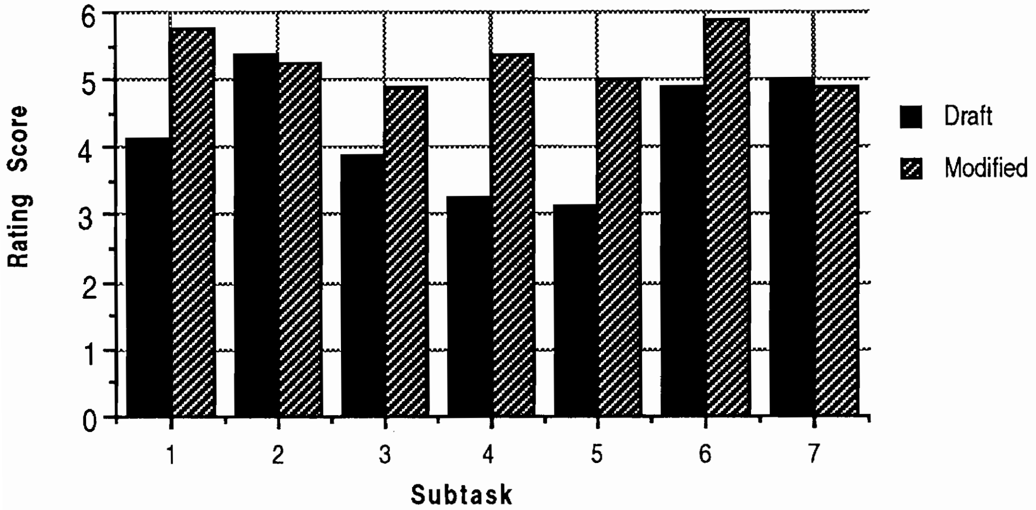


Figure 11. Mean rating scores for question 1 (Subtasks: 1=Symbols, 2=Multiple Symbols, 3=Lines, 4=Text, 5=Copy Edit, 6=Move Edit, 7=Delete Edit)

TABLE 7.

Document (2) X Subtask (8) ANOVA Summary Table

Question 1: How easy to use was this part of the document?

Source	df	SS	MS	F	p
Between					
Document (Doc)	1	31.08	31.08	10.51	0.0059
Subject (S)/Doc	14	41.41	2.96	-----	
Within					
Subtask (T)	6	25.30	4.22	4.60	0.0004
Doc X T	6	19.73	3.29	3.59	0.0032
T X S/Doc	84	76.96	0.92	-----	

four. Since these scores were not anchored, they could not be interpreted clearly. However, almost the same scores were observed for question nine which basically asked for the same information (See section 4.2.5.9.). The scores for question nine were anchored for better interpretation of the data.

The testing of hypothesis 5.1b also revealed significant difference in the within-subject main effect, subtask, $\underline{F}(6,84)=4.6$, $p<0.05$. The Newman-Keuls post-hoc analysis ($p<0.05$) revealed that subjects rated the document sections respective to subtasks 1, 2, 6, and 7 easier to use than those of subtasks 3, 4, and 5.

The interaction effect (Document X Subtask) was found to be significant, $\underline{F}(6,84)=3.59$, $p<0.05$. The Newman-Keuls sequential range test further revealed that only the modified document sections for subtasks 1, 4, and 5 were preferred in their usability (east to use) over the respective sections of the draft document.

4.2.5.2. Question Two (Hypothesis 5.2). Question two asked the subjects how much they agreed or disagreed with the following statement: "I have found the information in this part of the document to be very accurate." Each subject selected 1 out of 6 possible choices in the rating scale ranging from 1 to 6. For this question, all six scores were anchored. Their respective descriptions are as follows:

(1) = Strongly disagree
(2) = Disagree
(3) = Somewhat Disagree
(4) = Somewhat Agree
(5) = Agree
(6) = Strongly Agree

Hypothesis 5.2 is restated as two sub-hypotheses as they apply to the dependent variable, rating score:

5.2a) The modified IRRAS document is expected to result in better subjective rating than the draft IRRAS document. For question 2, the specific meaning of this hypothesis is that subjects are expected to rate the modified document to contain more accurate information than the draft document.

5.2b) Document sections for some subtasks are expected to be perceived as more accurate than those of other subtasks.

The mean subjective rating scores of both versions of the IRRAS document are shown in Figure 12. The results of the ANOVA are shown in Table 8.

When hypothesis 5.2a was tested, the between-subjects main effect of document was found to be significant, $F(1,14)=10.67$, $p<0.05$. The modified document was rated to be more accurate than the draft document. The Figure 12 shows that sections of the modified document pertaining to all subtasks are scored on the average of over 5, agreeing or strongly agreeing with the statement that the document is very accurate. For the draft document, the answers varied greatly, ranging from disagreeing to strongly agreeing with the statement. Their average score was slightly over 4, somewhat agreeing with the statement that the draft document was very accurate.

The testing of hypothesis 5.2b also revealed significant difference in the within-subject main effect, subtask, $F(6,84)=4.6$, $p<0.05$. The Newman-Keuls post-hoc analysis ($p<0.05$) revealed that subjects rated the document sections respective to subtasks 2, 3, 6, and 7 to be more accurate than those of subtasks 5, 4, and 1.

The interaction effect (Document X Subtask) was found to be significant, $F(6,84)=3.59$, $p<0.05$. The Newman-Keuls sequential range test further revealed that although there was a difference in accuracy between the two documents for subtasks 1, 4, and 5, there was no difference in preference for document sections pertaining to subtasks 2, 3, 6, and 7.

4.2.5.3. Question Three (Hypothesis 5.3). Question three asked the subjects how much they agreed or disagreed with the following statement: "This section of the document provided complete information to perform the subtask." For this question, all six scores were anchored with the same descriptions listed in the previous subsection.

Question 2: Information in this part of the document is very accurate. (Rating Scores: 1=Strongly Disagree, 2=Disagree, 3=Slightly Disagree, 4=Slightly Agree, 5=Agree, 6=Strongly Agree)

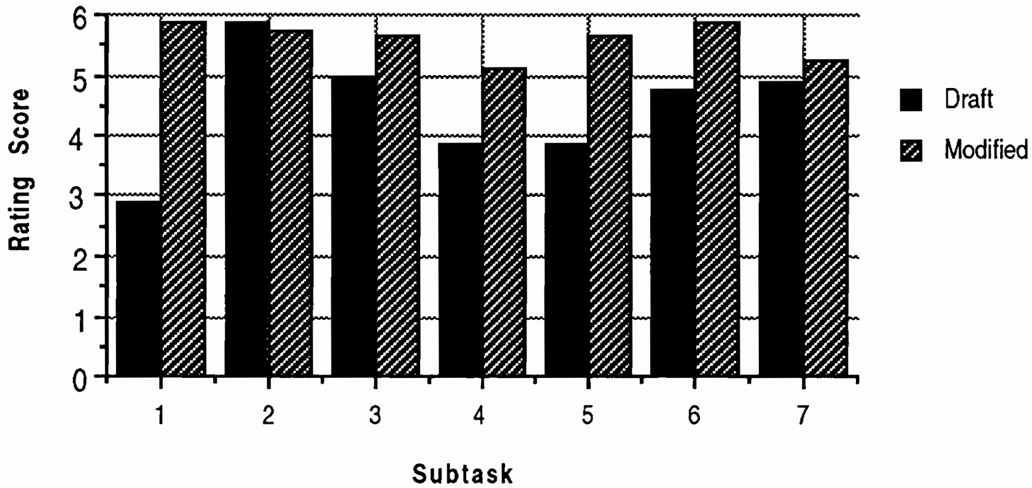


Figure 12. Mean rating scores for question 2 (Subtasks: 1=Symbols, 2=Multiple Symbols, 3=Lines, 4=Text, 5=Copy Edit, 6=Move Edit, 7=Delete Edit)

TABLE 8.

Document (2) X Subtask (8) ANOVA Summary Table

Question 2 (in a statement form): Information in the document is very accurate.

Source	df	SS	MS	F	p
Between					
Document (Doc)	1	36.57	36.57	10.67	0.0056
Subject (S)/Doc	14	47.96	3.43	-----	
Within					
Subtask (T)	6	24.96	4.16	5.90	0.0001
Doc X T	6	25.18	4.20	5.95	0.0001
T X S/Doc	84	59.29	0.71	-----	

Hypothesis 5.3 is restated as two sub-hypotheses as they apply to the dependent variable, rating score:

5.3a) The modified IRRAS document is expected to result in better subjective rating than the draft IRRAS document. For question 3, the specific meaning of this hypothesis is that subjects are expected to rate the modified document higher than the draft document in containing complete information necessary to accomplish their respective subtasks.

5.3b) Document sections for some subtasks are expected to be rated higher in containing complete information necessary to accomplish the subtasks than document sections for other subtasks.

The mean subjective rating scores of both versions of the IRRAS document are shown in Figure 13. The results of the ANOVA are shown in Table 9.

When hypothesis 5.3a was tested, the between-subjects main effect of document was found to be significant, $F(1,14)=6.72$, $p<0.05$. The modified document was rated better than the draft document for providing complete information to accomplish the assigned subtasks. Figure 13 shows that sections of the modified document pertaining to all subtasks are scored on the average of over 5, agreeing or strongly agreeing with the statement that the document provides sufficient and complete information to carry out the subtasks. For the draft document, the subjects' opinions varied from slightly disagreeing to strongly agreeing with the statement.

The testing of hypothesis 5.3b also revealed significant difference in the within-subject main effect, subtask, $F(6,84)=4.11$, $p<0.05$. The Newman-Keuls post-hoc analysis ($p<0.05$) revealed that subjects rated the document sections respective to subtasks 2, to provide more sufficient information than those of subtasks 7, 5, 4, and 1. However subtask 2 did not differ with subtasks 6 and 3.

The interaction effect (Document X Subtask) was not found to be significant, $F(6,84)=2.17$, $p>0.05$.

Question 3: This document section provided complete information to perform the subtask. (1=Strongly Disagree, 2=Disagree, 3=Slightly Disagree, 4=Slightly Agree, 5=Agree, 6=Strongly Agree)

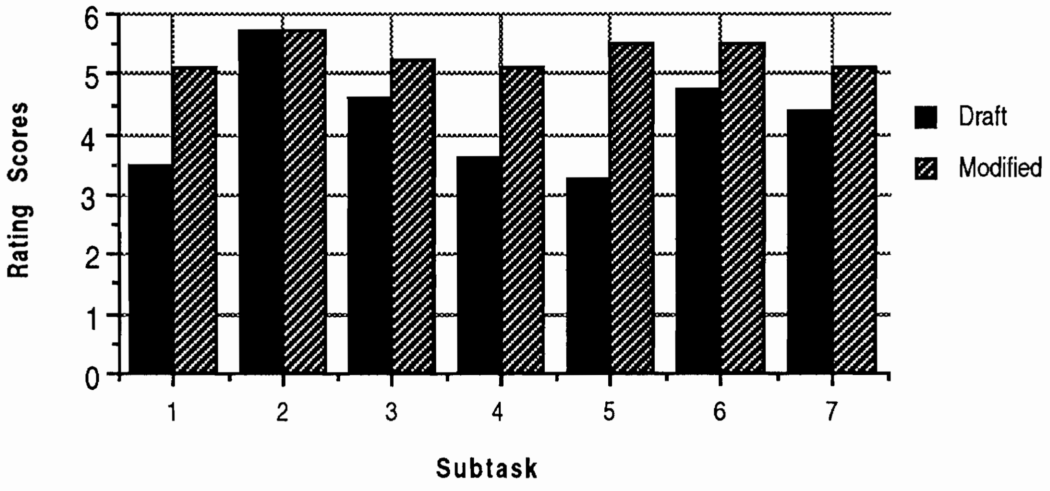


Figure 13. Mean rating scores for question 3 (Subtasks: 1=Symbols, 2=Multiple Symbols, 3=Lines, 4=Text, 5=Copy Edit, 6=Move Edit, 7=Delete Edit)

TABLE 9.

Document (2) X Subtask (8) ANOVA Summary Table

Question 3 (in a statement form): The document provided complete information to perform the subtask.

Source	df	SS	MS	F	p
Between					
Document (Doc)	1	32.14	32.14	6.72	0.0213
Subject (S)/Doc	14	66.96	4.78	-----	
Within					
Subtask (T)	6	26.05	4.34	4.11	0.0012
Doc X T	6	13.73	2.29	2.17	0.0544
T X S/Doc	84	88.79	1.06	-----	

4.2.5.4. Question Four (Hypothesis 5.4). Question four asked the subjects how much they agreed or disagreed with the following statement: "This section of the document has extra, unnecessary information." For this question, all six scores were anchored with the same descriptions listed in the previous subsection.

Hypothesis 5.4 is restated as two sub-hypotheses as they apply to the dependent variable, rating score:

5.4a) The modified IRRAS document is expected to result in better subjective rating than the draft IRRAS document. For question 4, the specific meaning of this hypothesis is that the modified document is expected to be perceived as containing less unnecessary information than the draft document.

5.4b) Also, document sections for some subtasks are expected to be perceived as containing less unnecessary information than those of other subtasks.

The mean subjective rating scores of both documents are shown in Figure 14. The results of the ANOVA are shown in Table 10.

When hypothesis 5.4a was tested, the between-subjects main effect of document was not found to be significant, $\underline{F}(1,14)=0.63$, $p>0.05$. There was no difference between the draft and modified documents in containing unnecessary information. Thus hypothesis 5.4a was not supported. In general, the subjects thought that neither document contained unnecessary information.

However, the testing of hypothesis 5.4b revealed significant difference in the within-subject main effect, subtask, $\underline{F}(6,84)=3.84$, $p<0.05$. The Newman-Keuls post-hoc analysis ($p<0.05$) revealed that subjects rated the document sections respective to subtasks 7 and 2 to contain more unnecessary information than document section respective to subtask 1. Also, document section for subtask 2 was believed to contain more unnecessary information than that of subtask 3.

The interaction effect (Document X Subtask) was found to be significant, $\underline{F}(6,84)=2.90$, $p<0.05$. Although the document main effect did not show a significant difference, the Newman-Keuls sequential range test ($p<0.05$) revealed that the modified

Question 4: This document section has extra, unnecessary information.
(1=Strongly Disagree, 2=Disagree, 3=Slightly Disagree,
4=Slightly Agree, 5=Agree, 6=Strongly Agree)

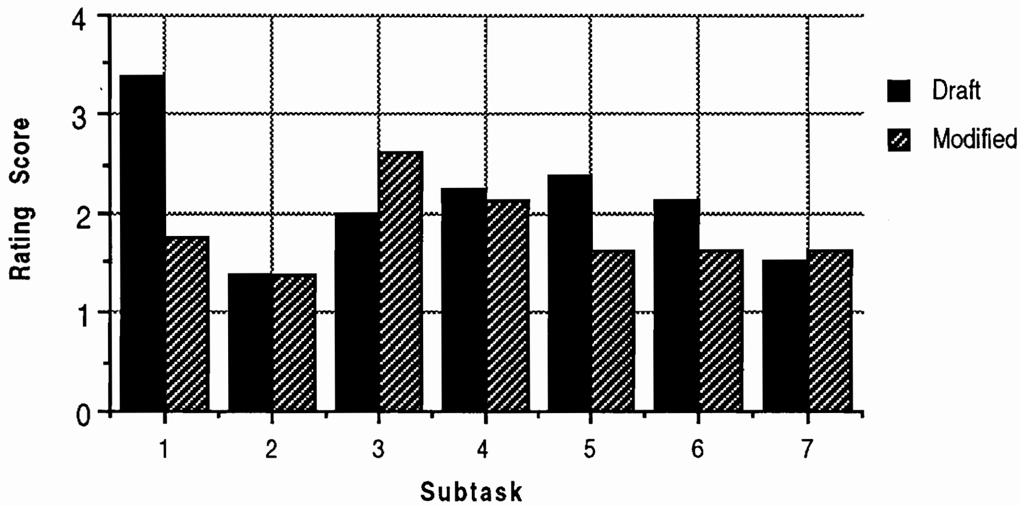


Figure 14. Mean rating scores for question 4 (Subtasks: 1=Symbols, 2=Multiple Symbols, 3=Lines, 4=Text, 5=Copy Edit, 6=Move Edit, 7=Delete Edit)

TABLE 10.

Document (2) X Subtask (8) ANOVA Summary Table

Question 4 (in a statement form): This document has extra, unnecessary information

Source	df	SS	MS	F	p
Between					
Document (Doc)	1	2.89	2.89	0.63	0.4423
Subject (S)/Doc	14	64.79	4.63	-----	
Within					
Subtask (T)	6	16.71	2.79	3.84	0.002
Doc X T	6	12.60	2.10	2.90	0.0129
T X S/Doc	84	60.96	0.73	-----	

document section for subtask 1 was perceived to contain less amount of unnecessary information than that of the draft document.

4.2.5.5. Question Five (Hypothesis 5.5). Question five asked the subjects how much they agreed or disagreed with the following statement: "It took extra effort to understand the language used in this section of the document" For this question, all six scores were anchored with the same descriptions listed in the previous subsection.

Hypothesis 5.5 is restated as two sub-hypotheses as they apply to the dependent variable, rating score:

5.5a) The modified IRRAS document is expected to result in better subjective rating than the draft IRRAS document. For question 5, the specific meaning of this hypothesis is that the subjects took more effort to understand the language used in the draft document than the modified document.

5.5b) Also, document sections for some subtasks are expected to be perceived as requiring more effort to understand the language used than those of other subtasks.

The mean subjective rating scores of both documents are shown in Figure 15. The results of the ANOVA are shown in Table 11.

When hypothesis 5.5a was tested, the between-subjects main effect of document was found to be significant, $\underline{F}(1,14)=6.64$, $p<0.05$. The subjects ranged from those who disagreed to strongly disagreed with the statement that it took extra effort to understand the language used in the modified document. The subjects ranged from those who slightly disagreed to disagreed with the same statement for the draft document. However, when the two documents were compared, the subjects thought that it took more effort to understand the language used in the draft document than the one used in the modified document.

The testing of hypothesis 5.5b revealed significant difference in the within-subject main effect, subtask, $\underline{F}(6,84)=5.67$, $p<0.05$. The Newman-Keuls post-hoc analysis ($p<0.05$) revealed that subjects rated the document section respective to subtasks 4 and 5 to

Question 5: It took extra effort to understand the language used in this document section.
(Rating Scores: 1=Strongly Disagree, 2=Disagree, 3=Slightly Disagree,
4=Slightly Agree, 5=Agree, 6=Strongly Aree)

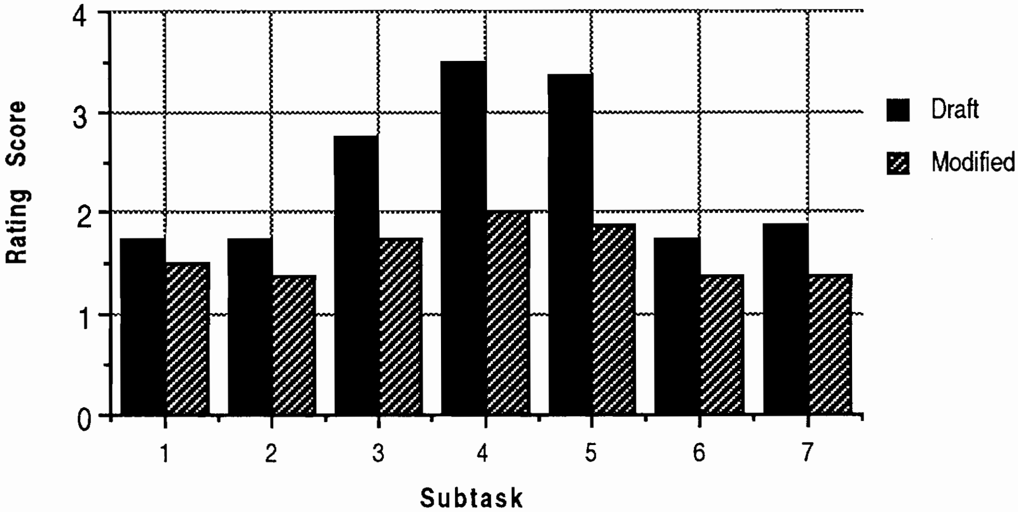


Figure 15. Mean rating scores for question 5 (Subtasks: 1=Symbols, 2=Multiple Symbols, 3=Lines, 4=Text, 5=Copy Edit, 6=Move Edit, 7=Delete Edit)

TABLE 11.

Document (2) X Subtask (8) ANOVA Summary Table

Question 5 (in a statement form): It took an extra effort to understand the language used.

Source	df	SS	MS	F	p
Between					
Document (Doc)	1	17.29	17.29	6.64	0.0219
Subject (S)/Doc	14	36.43	2.60	-----	
Within					
Subtask (T)	6	26.88	4.48	5.67	0.0001
Doc X T	6	7.09	1.18	1.5	0.1893
T X S/Doc	84	66.32	0.79	-----	

require more effort to understand the language than those respective to subtasks 1, 7, 2, and 6.

The interaction effect (Document X Subtask) was not found to be significantly different, $F(6,84)=1.5$, $p>0.05$.

4.2.5.6. Question Six (Hypothesis 5.6). Question six asked the subjects how much they agreed or disagreed with the following statement: "The examples and figures were helpful" For this question, all six scores were anchored with the same descriptions listed in the previous subsection.

Hypothesis 5.6 is restated as two sub-hypotheses as they apply to the dependent variable, rating score:

5.6a) The modified IRRAS document is expected to result in better subjective rating than the draft IRRAS document. For question 6, the specific meaning of this hypothesis is that the examples and figures in the modified document are expected to be perceived as more helpful than those of the draft document.

5.6b) Also, the examples and figures in document sections for some subtasks are expected to be perceived as more helpful than those of other subtasks.

The mean subjective rating scores of both versions of the IRRAS document are shown in Figure 16. The results of the ANOVA are shown in Table 12.

When hypothesis 5.6a was tested, the between-subjects main effect of document was not found to be significant, $F(1,14)=3.90$, $p>0.05$. The graph in Figure 16 indicates that subjects slightly disagreed with the statement that the examples and figures used in the draft document were helpful. On the other hand, subjects slightly agreed with the statement that the examples and figures used in the modified document were helpful. Even though there seems to be a qualitative difference in the evaluation of the two documents, the ANOVA revealed that there was no significant difference. Thus, hypothesis 5.6a was not supported.

Question 6: The examples and figures in this document section were helpful. (1=Strongly Disagree, 2=Disagree, 3=Slightly Disagree, 4=Slightly Agree, 5=Agree, 6=Strongly Agree)

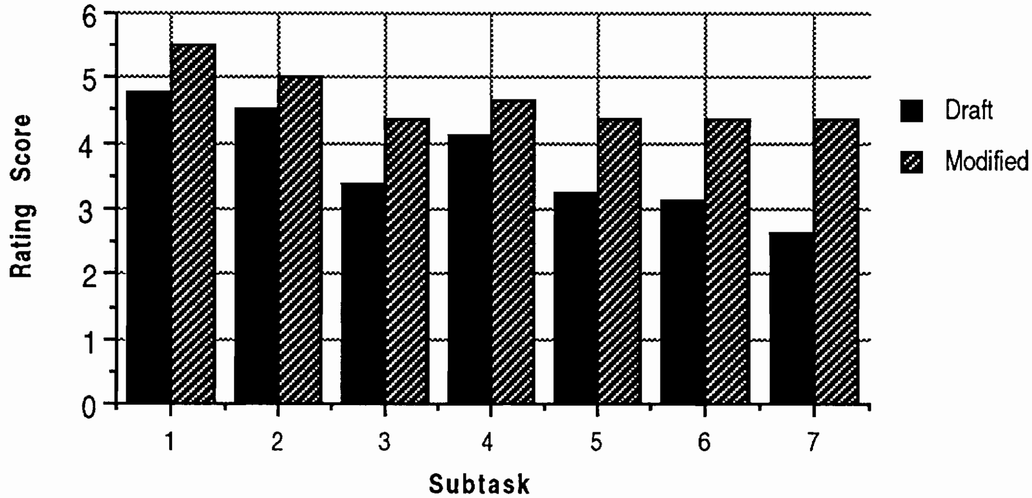


Figure 16. Mean rating scores for question 6 (Subtasks: 1=Symbols, 2=Multiple Symbols, 3=Lines, 4=Text, 5=Copy Edit, 6=Move Edit, 7=Delete Edit)

TABLE 12.

Document (2) X Subtask (8) ANOVA Summary Table

Question 6 (in a statement form): The examples and figures were helpful.

Source	df	SS	MS	F	p
Between					
Document (Doc)	1	27.01	27.01	3.90	0.0685
Subject (S)/Doc	14	97.05	6.93	-----	
Within					
Subtask (T)	6	34.09	5.68	5.14	0.0002
Doc X T	6	4.80	0.80	0.72	0.6310
T X S/Doc	84	92.82	1.11	-----	

The testing of hypothesis 5.6b revealed significant difference in the within-subject main effect, subtask, $F(6,84)=5.14$, $p<0.05$. The Newman-Keuls post-hoc analysis ($p<0.05$) revealed that subjects rated the figures and examples in the document section respective to subtask 1 as more helpful than those of subtasks 3, 5, 6, and 7. Also, the figures and examples for subtask 2 were more helpful than that of subtask 7.

The interaction effect (Document X Subtask) was not found to be significantly different, $F(6,84)=0.72$, $p>0.05$.

4.2.5.7. Question Seven (Hypothesis 5.7). Question seven asked the subjects how much they agreed or disagreed with the following statement: "The information was developed in a way that made it easy to learn." For this question, all six scores were anchored with the same descriptions listed in the previous subsection.

Hypothesis 5.7 is restated as two sub-hypotheses as they apply to the dependent variable, rating score:

5.7a) The modified IRRAS document is expected to result in better subjective rating than the draft IRRAS document. Specifically, the subjects are expected to perceive the information developed in the modified document as easier to learn than that of the draft document.

5.7b) The subjects are expected to perceive the information developed in the document sections for some subtasks as easier to learn than those of other subtasks.

The mean subjective rating scores of both documents are shown in Figure 17. The results of the ANOVA are shown in Table 13.

When hypothesis 5.7a was tested, the between-subjects main effect of document was found to be significant, $F(1,14)=7.26$, $p<0.05$. The subjects agreed with the statement that the information in the modified document was developed in a way that made it easy to learn. However, subjects ranged from slightly disagreeing to slightly agreeing with the statement for the draft document. The ANOVA shows that the information developed in the modified document was easier to learn than that of the draft document.

Question 7: Information was developed in a way that made it easy to learn.
(1=Strongly Disagree, 2=Disagree, 3=Slightly Disagree,
4=Slightly Agree, 5=Agree, 6=Strongly Agree)

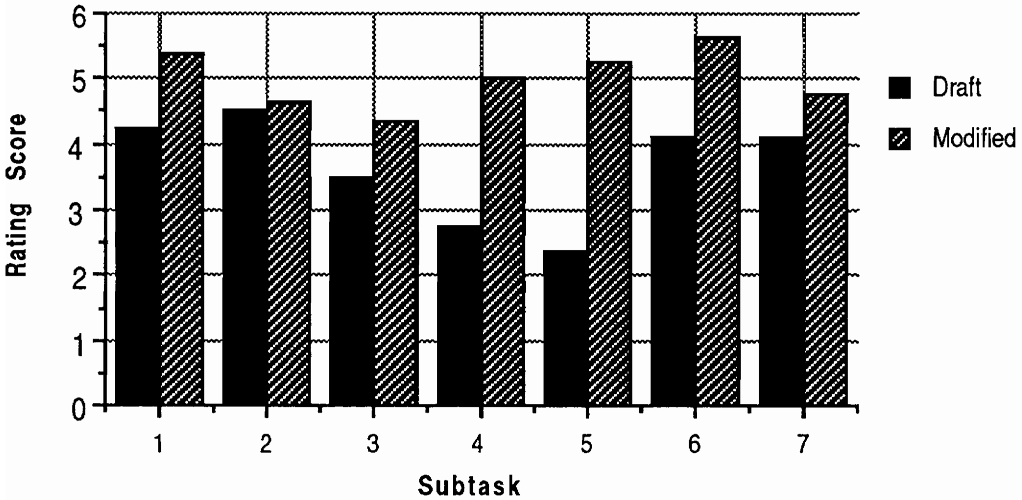


Figure 17. Mean rating scores for question 7 (Subtasks: 1=Symbols, 2=Multiple Symbols, 3=Lines, 4=Text, 5=Copy Edit, 6=Move Edit, 7=Delete Edit)

TABLE 13.

Document (2) X Subtask (8) ANOVA Summary Table

Question 7 (in a statement form): Information was developed in a way that made it easy to learn.

Source	df	SS	MS	F	p
Between					
Document (Doc)	1	50.22	50.22	7.26	0.0174
Subject (S)/Doc	14	96.84	6.92	-----	
Within					
Subtask (T)	6	19.59	3.26	3.60	0.0032
Doc X T	6	21.84	3.64	4.01	0.0014
T X S/Doc	84	76.29	0.91	-----	

The testing of hypothesis 5.7b revealed significant difference in the within-subject main effect, subtask, $F(6,84)=3.60$, $p<0.05$. The Newman-Keuls post-hoc analysis ($p<0.05$) revealed that the information developed in the document sections respective to subtasks 6 and 1 as easier to learn than those of subtasks 3, 4 and 5. Also, information developed in the document section for subtask 2 was easier to learn than those of subtasks 4 and 5.

The interaction effect (Document X Subtask) was found to be significantly different, $F(6,84)=4.01$, $p<0.05$. The Newman-Keuls sequential range test revealed that only the information contained in the modified document sections for subtasks 4 and 5 were rated to be significantly easier to learn than the information contained in the respective sections of the draft document.

4.2.5.8. Question Eight (Hypothesis 5.8). Question eight asked the subjects how much they agreed or disagreed with the following statement: "I have found annoying inconsistencies in presentation or coverage in this section of the document." For this question, all six scores were anchored with the same descriptions listed in the previous subsection.

Hypothesis 5.8 is restated as two sub-hypotheses as they apply to the dependent variable, rating score:

5.8a) The modified IRRAS document is expected to result in better subjective rating than the draft IRRAS document. Specifically, the subjects are expected to find more annoying inconsistencies in presentation or coverage of the draft document than the modified document.

5.8b) The subjects are expected find more annoying inconsistencies in presentation or coverage of the document sections for some subtasks than those of other subtasks.

The mean subjective rating scores of both documents are shown in Figure 18. The results of the ANOVA are shown in Table 14.

Question 8: Annoying inconsistencies were present in this document section.
(1=Strongly Disagree, 2=Disagree, 3=Slightly Disagree,
4=Slightly Agree, 5=Agree, 6=Strongly Agree)

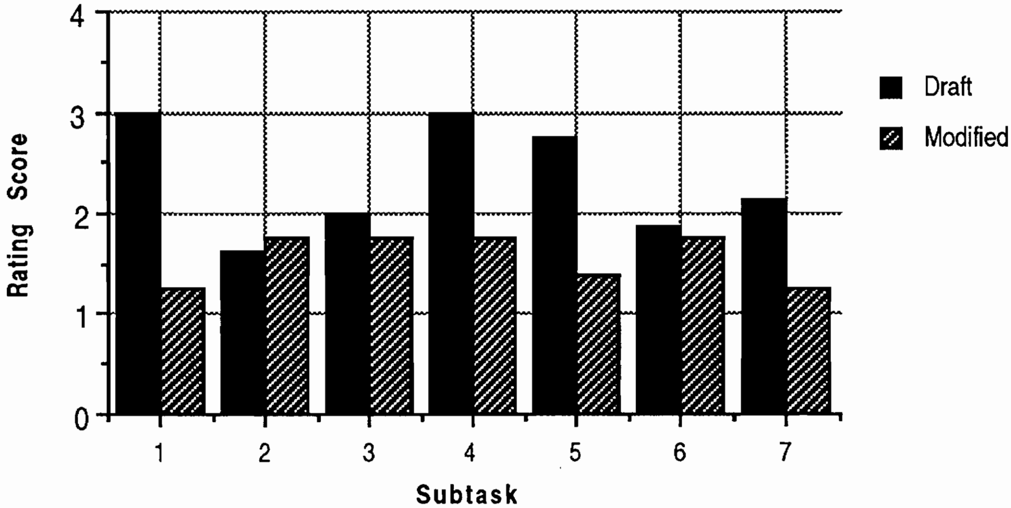


Figure 18. Mean rating scores for question 8 (Subtasks: 1=Symbols, 2=Multiple Symbols, 3=Lines, 4=Text, 5=Copy Edit, 6=Move Edit, 7=Delete Edit)

TABLE 14.

Document (2) X Subtask (8) ANOVA Summary Table

Question 8 (in a statement form): Annoying inconsistencies were present.

Source	df	SS	MS	F	p
Between					
Document (Doc)	1	17.29	17.29	4.7	0.048
Subject (S)/Doc	14	51.54	3.68	-----	
Within					
Subtask (T)	6	6.18	1.03	0.96	0.46
Doc X T	6	12.21	2.04	1.89	0.092
T X S/Doc	84	90.46	1.08	-----	

When hypothesis 5.8a was tested, the between-subjects main effect of document was found to be significant, $F(1,14)=4.70$, $p<0.05$. The subjects ranged from strongly disagreeing to disagreeing with the statement that annoying inconsistencies were found in the modified document. For the draft document, the subjects ranged from disagreeing to slightly disagreeing with the statement. The ANOVA showed that subjects perceived the modified document to contain fewer inconsistencies in its presentation or coverage than the draft document.

The testing of hypothesis 5.8b did not reveal significant difference in the within-subject main effect, subtask, $F(6,84)=0.96$, $p>0.05$. The subjects did not find any significant difference in annoying inconsistencies among the document sections of all 7 subtasks.

Also, the interaction effect (Document X Subtask) was not found to be significantly different, $F(6,84)=1.89$, $p>0.05$.

4.2.5.9. Question Nine (Hypothesis 5.9). Question nine asked the following to the subjects: "How easy to use is this manual (for this section)? This question was basically same as question one, but phrased differently. The difference is that, for question 9, the subjects were asked to choose 1 of the 6 descriptive answers that were provided. These answers which were not supported with a respective numerical scale included the following:

extremely easy,
very easy,
fairly easy,
fairly difficult,
very difficult, and
extremely difficult

Since these choices were not anchored with rating scores, a score was assigned to each of the six descriptive answers for analysis. A score of 6 was assigned for extremely easy, 5 for very easy, 4 for fairly easy, 3 for fairly difficult, 2 for very difficult, and 1 for extremely difficult.

Hypothesis 5.9 is restated as two sub-hypotheses as they apply to the dependent variable, rating score:

5.9a) The modified IRRAS document is expected to result in better subjective rating than the draft IRRAS document. Specifically, the modified document is expected to be rated easier to use than the draft document.

5.9b) The document sections for some subtasks are expected to be rated easier to use than those of other subtasks.

The mean subjective rating scores of both documents are shown in Figure 19. The results of the ANOVA are shown in Table 15.

When hypothesis 5.9a was tested, the between-subjects main effect of document was found to be significant, $\underline{F}(1,14)=10.03$, $p<0.05$. The modified document was rated as very easy to extremely easy to use. The draft document was rated as fairly easy to use. The ANOVA showed that the modified document was rated as easier to use than the draft document.

The testing of hypothesis 5.9b showed significant difference in the within-subject main effect, subtask, $\underline{F}(6,84)=6.31$, $p<0.05$. The Newman-Keuls post-hoc analysis ($p<0.05$) showed that the document sections for subtasks 6, 2, 1, and 7 were rated to be easier to use than the document sections for subtasks 4 and 5.

The interaction effect (Document X Subtask) was not found to be significantly different, $\underline{F}(6,84)=1.31$, $p>0.05$.

4.2.6 Perceived Subtask Difficulty

Several factors that could have contributed to the differences observed in both objective measures as well subjective ratings for the subtask main effect. One obvious reason could be the differences in the lengths of the document sections for the eight subtasks. Another reason for the observed differences could be that certain subtasks are longer and therefore requires more work and time. Another related reason for the subtask

Question 9: How easy to use is this manual?
(Rating Scores: 1=Extremely Difficult, 2=Very Difficult, 3=Fairly Difficult, 4=Fairly Easy, 5=Very Easy, 6=Extremely Easy)

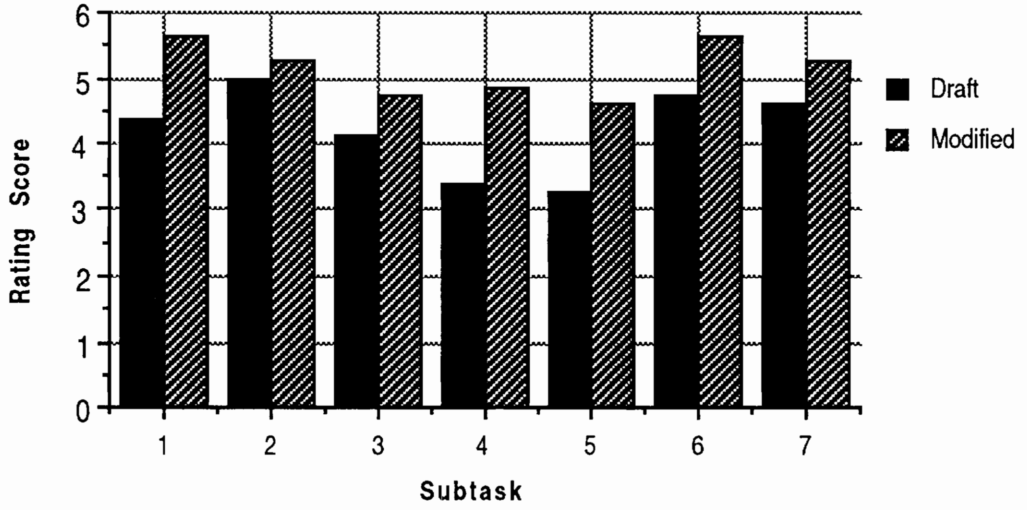


Figure 19. Mean rating scores for question 9 (Subtasks: 1=Symbols, 2=Multiple Symbols, 3=Lines, 4=Text, 5=Copy Edit, 6=Move Edit, 7=Delete Edit)

TABLE 15.

Document (2) X Subtask (8) ANOVA Summary Table

Question 9: How easy to use is this section of the document?

Source	df	SS	MS	F	p
Between					
Document (Doc)	1	24.14	24.14	10.03	0.0069
Subject (S)/Doc	14	33.71	2.41	-----	
Within					
Subtask (T)	6	24.68	4.11	6.31	0.0001
Doc X T	6	5.11	0.85	1.31	0.2639
T X S/Doc	84	54.79	0.65	-----	

differences is that some subtasks may be perceived as more difficult to perform than others. To determine if there is any correlation between the subtask differences in the observedmeasures (objective and subjective) and the difficulty level of the subtasks, the following information was collected and analyzed.

To establish this order of difficulty, the subjects were asked to rank order the difficulty of the first seven subtasks. The ANOVA showed significant difference among the perceived subtask difficulty, $F(6,90)=143.18$, $p<0.001$ (Table 16). The following is the order of perceived difficulty from least difficult to most difficult when mean scores were compared:

1) Subtask 6, Move editing (least difficult)
2) Subtask 7, Delete editing
3) Subtask 2, Placing multiple copies of fault tree symbols
4) Subtask 1, Selecting and placing various fault tree symbols
5) Subtask 3, Line drawing
6) Subtask 4, Text writing
7) Subtask 5, Copy editing (most difficult)

The Newman-Keuls post-hoc analysis ($p<0.05$) showed that the two most difficult subtasks, subtask 5 and subtask 4, were not significantly different from each other but were significantly different from the rest of the subtasks. Additionally, the two least difficult subtasks, subtask 7 and subtask 6, were not significantly different from each other but were significantly different from the rest of the subtasks. All other subtasks were significantly different from each other.

To investigate the relationship between the rank order of the difficulty levels of the subtasks and the rank orders of the 13 observed values (4 objective and 9 subjective), the Spearman rank-order correlation coefficient values were calculated. A summary of correlated items is shown in Table 17.

TABLE 16.

Subtask Difficulty Rating ANOVA Summary Table

Source	df	SS	MS	F	p
Between					
Subjects (S)	15	0			
Within					
Subtask (T)	6	405.5	67.58	143.18	<0.001
T X S	90	42.5	0.472	-----	

TABLE 17

Correlation Analysis between Subtask Difficulty Levels and Observed Values

Correlation between Subtask Difficulty Level and ________________	r_s
(Objective Measures)	
Number of Errors	0.882**
Subtask Completion Time	0.893***
Document Reading Time	0.75*
Number of Helps Requested	0.524
(Subjective Ratings)	
Q1: Difficulty of use	0.893***
Q2: Inaccurate information	0.393
Q3: Incomplete information	0.538
Q4: Extra, unnecessary information	0.500
Q5: Extra effort to understand the language	0.929****
Q6: Examples and figures not helpful	0.321
Q7: Information not developed for easy learning	0.857**
Q8: Annoying inconsistencies	0.714*
Q9: Difficulty of use	0.893***

* p<0.05

** p<0.025

*** p<0.01

**** p<0.005

4.2.7. Document Comparison

The last analysis compared the subjects' opinion of the usability of one document versus the other. After the completion of eight subtasks using one document, subjects were given the other document which they did not use when previously performing the subtasks. After reading and examining the other document, they were asked to compare and rate the two documents. The modified document was compared to the draft document using the rating scale shown in Figure 20. The subjects were not told which of the two documents was the modified version.

The subjects who used the modified document to perform their subtasks rated the modified document as easier or much easier (mean score of 2.7) to use than the draft document. The subjects who used the draft document for their subtasks also rated the modified document as easier or much easier to use (mean score of 2.5). The ANOVA for the comparison ratings showed that the two groups of subjects did not disagree that the modified document was easier or much easier to use than the draft document, $F(1,14)=2.739$, $0.1<p<0.25$ (Table 18). In other words, the subjects who were more acquainted with the draft document and the subjects who were more acquainted with the modified document both thought that the modified document was easier or much easier to use than the draft document.

Much More Difficult	More Difficult	Slightly More Difficult	Same	Slightly Easier	Easier	Much Easier
-3	-2	-1	0	+1	+2	+3

Figure 20. Document comparison rating scale

TABLE 18.

Document Comparison ANOVA Summary Table

Source	df	SS	MS	F	p
Between					
Document (D)	1	0.5625	0.5625	2.739	<0.25
Subjects (S)/D	14	2.875	0.20536	-----	

5. DISCUSSION

The purpose of the first phase of this experiment was to determine poor aspects of the draft IRRAS user document using the critical incident technique. The identified features were used to modify and develop the draft IRRAS user document to improve its usability. To determine if the changes made any difference in the usability of the modified document, both the draft and modified documents were tested and compared for 4 objective measures and 9 subjective measures in the second phase of this experiment.

In order to significantly improve the usability of a product, Gould and Lewis (1985) emphasized that any modification has to be justified from a user's perspective. Gould and Lewis also stated that user's opinions should be incorporated in the early phase of the product development. To apply these principles into modifying the draft version of the IRRAS user document, the critical incident technique was used to help users identify the draft document's usability problems. The most frequently appearing incidents were unclear and wrong procedural instructions, ambiguous and undefined terms and phrases, unfriendly format, and confusion over multi-functional mouse buttons. These problems seem to have hindered the subjects from easily learning the instructions from the draft document. Particularly, the unclear and undefined words and phrases seem to have caused the most annoyance among the subjects.

As discussed in section 3.2, the critical incidents were identified and categorized and then used as springboards to make modifications on the draft document. Many changes were made on the draft document to make it easier to use. To test whether these changes significantly improved the usability of the document, experimental tests were conducted to measure objective scores and subjective ratings. Moreover, these tests were conducted to determine if the critical incident technique was a reliable method to aid users in identifying document usability problems.

In the second phase of the experiment, both objective and subjective measures generally supported the hypothesis that the modified document is superior in usability over the draft document. For some cases, this difference was only subtask specific. In other words, the difference was observed for some subtasks but not for others.

5.1 Summary and Interpretation of Findings for Hypothesis 1

Hypothesis 1a was sufficiently supported by the data: the use of the modified document resulted in fewer errors than the use of the draft document. Hypothesis 1b was also supported in that for some subtasks, more errors were made than others. Specifically, more errors were made in subtask 5 (copy editing) and 4 (text writing) than subtasks 1 (copying and placing symbols), 2 (placing multiple copies of symbols), 6 (move editing), 7 (delete editing), and 8 (drawing a fault tree).

The instructions in the draft document sections for subtasks 4, 5, 6, and 7 are longer and more complex than the instructions for subtasks 1, 2, and 3. This might account for more errors made for subtasks 4 and 5. The subjects may not have made as many mistakes in subtasks 6 and 7 because they involve instructions very similar to those of subtask 5 which is performed earlier (All subtasks were performed in their numerical order). However, even though the instructions for subtasks 4 and 5 were longer and more complex than the instructions for other subtasks in the modified document, hardly any difference was observed in the number of errors across all of the subtasks.

In addition, the high Spearman rank-order correlation coefficient of 0.88 between the difficulty level of the subtasks and the errors made in subtasks may also explain the observation that subjects made more errors in more difficult subtasks.

The significance of the interaction effect showed that the difference in number of errors made in using the modified and the draft documents was significant for more complex subtasks, specifically 4 and 5, than for other subtasks. As pointed out by the subjects in phase 1 study, more unclear instructions, specifically undefined terms and their functions, were found for document sections for subtasks 4 and 5 than other document sections.

5.2. Summary and Interpretation of Findings for Hypothesis 2

Hypothesis 2a was not supported by the data: the modified document did not result in faster subtask completion than the draft document. Usability is generally measured in terms of how easy it is to use the product by an ordinary user. Although significant differences were observed for other objective and subjective measures, it seems as though subtask time did not show any indication that there was any difference in the usability between the two documents. This observation did not support the idea that if a product is found to be better in its usability, that it will require less time to accomplish a given task. However, task completion time should not be a sole indicator to measure usability. In this study, even though a significant difference in time was not observed, users' subjective ratings and other objective measures (errors, personal helps, and reading time) show significant improvement of usability in the modified document. When using the draft document, the subjects may have resorted to figuring out the information in a trial and error strategy to compensate for the lack of well written instructions. Thus, the time span to complete the subtask may not have been different; however, more errors and poorer subjective ratings were observed for the draft document

However, hypothesis 2b was found to be significant. Subtask 8 (drawing a fault tree) required the most time to accomplish, followed by subtasks 4, 5, 1, 3, and then subtasks 7, 2, and 6. The subtask of drawing a whole fault tree naturally took the most time as expected. Subtasks 4, 5, 3 required more time than subtasks 7, 2, and 6 for similar reasons as mentioned in the previous section. Subtask 1 required more time than subtasks 7, 2, and 6 since it was the first subtask to be performed. The subjects may have taken more time during this subtask because of their unfamiliarity with the software and the document. Also, instructions for subtasks 7, 2, and 6 were similar to the instructions of their previously similar subtasks

In addition, the high Spearman rank-order correlation coefficient of 0.89 between the difficulty level of the subtasks and the subtask completion time may also explain the observation that difficult subtasks required more time for completion than the less difficult subtasks.

5.3. Summary and Interpretations for the Findings for Hypothesis 3

Hypothesis 3a was supported by the data: the modified document resulted in shorter document reading time than the draft document. The significant interaction effect showed that this main effect difference was mainly due to the differences in reading times for document sections pertaining to subtasks 1, 4, and 5. This result is not due to the difference in the number of words contained in the documents. In fact, if it was due to the number of words, the result should have been reversed. In this case, differences observed for document sections for subtasks 1, 4, and 5 seem to be due to difference in the time of comprehending the information in the document. For the modified document, even though its text was longer, the information was presented in a clearer format and the information itself was easier to understand.

Hypothesis 3b was also supported by the data: the subjects took the most time to read document sections for subtasks 4 and 5, followed by subtask 3, then subtask 2, then subtask 6 and 7, and then subtask 8.

As mentioned in the previous section, the subtasks perceived to be the most difficult and complex, 4 and 5, required the most reading time. Subtasks 2, 6, and 7 did not require as much reading time since the nature of the subtasks were similar to those performed earlier. Subtask 8 (drawing a whole fault tree) did not consume as much reading time since the majority of the subjects did not choose to use the document. Subtask 8 required only the skills and knowledge introduced to the subjects in the document necessary for performing first 7 subtasks.

In addition, the high Spearman rank-order correlation coefficient of 0.75 between the difficulty level of the first seven subtasks and document reading time for subtasks may also explain the observation that subjects took more time to read document sections pertaining to more difficult subtasks than those of less difficult subtasks.

5.4 Summary and Interpretations of the Findings for Hypothesis 4

Hypothesis 4a was supported by the data: the modified document resulted in fewer personal help requests than the draft document. In fact, the data for the modified document showed that not a single help was requested. For the draft document, helps were sought for subtasks 5, 4, 1, and 7 only. This accounted for the significant main effect of subtask and significant interaction effect. Thus, hypothesis 4b was also supported.

The subjects who used the modified document found the document to contain sufficient information to accomplish the subtasks. Those who used the draft document had trouble understanding the document sections of subtasks that were longer and more complex than those of other subtasks. Although the Spearman rank-order correlation coefficient of 0.52 is low, the subtasks that were perceived to be more difficult than others seem to have affected the subjects who used the draft document to seek for personal help.

5.5 Summary and Interpretations of the Findings for Hypothesis 5

Seven out of nine subjective opinions showed that the modified document was rated to be better and easier to use than the draft document. For questions 1 and 9, the subjects were asked to give their opinions about the documents' ease of use. Even though significant results were observed for both questions, it is noted that their rating scores were based on scales different from each other and different from the rating scale used for questions 2 through 8 (see Figure 6, page 32). For question 1, only two extreme scores, 1 and 6, and their respective anchors, most difficult and easiest, were provided to the subjects. For question 9, although 6 descriptive choices ranging from extremely difficult to extremely easy were given in a vertical column, they were not supported with respective scores. Since the construction features of these two rating scales were different from the rating scale used for questions 2 through 8, it is possible that the results of questions 1 and 9 could have been affected differently if the rating scale used for questions 2 through 8 was also used for questions 1 and 9.

5.5.1. Hypothesis 5.1

Hypothesis 5.1a was supported by the data: the modified document was rated easier to use than the draft document. The modified document sections for all 7 subtasks were rated above the score of 5 on a 6 point scale. The draft document sections for the 7 subtasks ranged from the score of slightly over 3 to slightly over 5. Since the 6-point scale was not anchored for this rating, it's difficult to assess what qualitative meaning these scores have. The same question of the document's ease of use was asked again in different words in question 9 (section 5.5.9).

Hypothesis 5.1b was also supported by the data: the document sections for subtasks 5, 4, and 3 were rated to be easier to use than those for subtasks 6, 2, 1, and 7. As stated in the previous sections, this difference seems to be due to the fact that the document sections for subtasks 5, 4, and 3 are longer and more complex Additionally, a high Spearman correlation coefficient of 0.89 shows that the perceived ease of use of for subtasks is correlated with perceived difficulty levels of the subtasks.

The significant interaction effect showed that the preference difference between the two documents was significantly different only for subtasks 1, 4, and 5. This result shows that for relatively easier subtasks with shorter document sections, there was no difference in the ratings of documents for perceived ease of use.

5.5.2. Hypothesis 5.2

Hypothesis 5.2a was supported by the data: the modified document was rated to contain more accurate information than the draft document. The subjects agreed or strongly agreed with the statement that the modified document contained accurate information. For the draft document, subjects' ratings ranged from disagreeing to strongly agreeing with the statement, depending on the document sections of the subtasks.

Hypothesis 5.2b was also supported by the data: the document sections respective to subtasks 2, 3, 6, and 7 were rated to be more accurate than those of subtasks 5, 4, and 1. This significance is mainly due to the single effect of the draft document. The Spearman correlation coefficient of 0.39 did not seem to show a significant relationship between the difficulty level of the subtasks and the subjective ratings for this question. The critical incidents collected from subtasks 4 and 5 show that there were relatively many ambiguous and undefined terms and phrases in their respective document sections. This may explain why the document sections for subtasks 5 and 4 were rated to be less accurate. The document section for subtask 1 was rated to be least accurate of all. The critical incidents also reveal that a key symbol was wrongly labeled in this document section.

The significant interaction effect showed that the difference in accuracy rating between the two documents was only significant for subtasks 1, 4, and 5. This shows that incorporating the findings of critical incident technique in the modification of the document improved its accuracy.

5.5.3. Hypothesis 5.3

Hypothesis 5.3a was supported by the data: the modified document was rated be better than the draft document in providing complete information to accomplish the assigned subtasks. The subjects agreed or strongly agreed that the modified document provided complete information. For the draft document, subjects varied in their opinion from slightly to strongly agreeing.

Hypothesis 5.3b was also supported by the data: the document section for subtasks 5, 1, 4, and 7 were rated to provide less amount of information than the document sections for subtask 2. This result is also explained by relatively large number of critical incidents identified as unclear and wrong procedural instructions contained in their respective document sections. For subtask 4 (writing text), the draft document did not provide clear information on how to end a text. For subtasks 4 and 5, the function of the reference point was undefined and was not shown how to use it.

5.5.4. Hypothesis 5.4

Hypothesis 5.4a was not supported by the data: the subjects disagreed and strongly disagreed with the statement that the documents contained extra, unnecessary information. There was no significant difference between the two documents

However, hypothesis 5.4b was supported by the data: the subjects rated the document sections respective to subtasks 7 and 2 to contain less amount of unnecessary information than the document section respective to subtask 1. This result is difficult to explain since the subjects generally disagreed with the statement that the documents contained unnecessary information. Also, since some of the critical incidents were described as document sections containing some irrelevant information, it was expected that the subjects would disagree with the statement for document sections for some of the subtasks. However, this was not observed. With the Spearman correlation coefficient value of 0.53 it is also difficult to conclude that the subtask difficulty level is associated with the significant difference among the document sections.

The difference in the rating scores between the two documents for subtask 1 might account for the significant interaction effect even though the main effect of the document was not found to be significant. The difference here, though, is between strongly disagreeing and slightly disagreeing with the statement that the documents contain unnecessary information.

5.5.5. Hypothesis 5.5

Hypothesis 5.5a was supported by the data: the subjects took more effort to understand the language used in the draft document than the modified document. The subjects strongly disagreed with the statement that it took extra effort to understand the language used in the modified document. However, for the draft document, the opinions ranged from disagreeing to slightly agreeing with the statement. The critical incidents showed that some users had trouble understanding certain words to describe a function. For example, the users are not informed what the reference point is in subtasks 4, 5, 6, and 7, let alone how it is used. A simple definition would have prevented the problem. This problem was corrected for the modified document , thus the users did not seem to have a problem understanding nor using the reference point in their subtasks.

Hypothesis 5.5b was also supported by the data: document sections for subtasks 4 and 5 were rated to require more effort to understand the language than document sections for subtasks 1, 7, 2, and 6. The Newman-Keuls sequential difference test ($p<0.05$) showed that this difference was mainly due to the single effect of the draft document since subjects, in general, strongly disagreed with the statement for the modified document. The critical incidents found for subtasks 4 and 5 contained unclear procedural instructions. Since the nature of these subtasks were more complex than others, they required more and clearer explanations. However, this result supported by the critical incidents show that the draft document did not provide easily understandable instructions. The Spearman correlation coefficient of 0.93 may explain that the subject's perception of the subtask difficulty level is also associated with how they evaluate the comprehensibility of the language used.

5.5.6. Hypothesis 5.6

Hypothesis 5.6a was not supported by the data: the subjects found the figures and examples in both documents to be slightly helpful without significant difference. This observation is due to the fact that same figures were used for both documents. Not a single critical incident was observed regarding figures. Thus, they were not modified for the modified document.

Hypothesis 5.6b was supported by the data: the figures and examples in the document section for subtask 1 were rated to be more helpful than the document sections for subtasks 3, 5, 6, and 7. Additionally, the figures and examples for subtask 2 were more helpful than those of subtask 7. The Spearman correlation coefficient of 0.32 showed that this difference did not have much to do with the difficulty levels of the subtasks. However, the fact that there were figures used for subtasks 1 and 2 and not for other subtasks may account for the significant difference among certain subtasks.

5.5.7. Hypothesis 5.7

Hypothesis 5.7a was supported by the data: the information developed in the modified document was perceived to be easier to learn than that of the draft document. In general, the subjects agreed with the statement that the information was developed in a way that made it easy to learn for the modified document. However, for the draft document, subjects' opinions varied from disagreeing to agreeing with the statement.

Hypothesis 5.7b was also supported by the data: the information developed in the document sections respective to subtasks 6 and 1 to be easier to learn than those respective to subtasks 3, 4, and 5. Also, the information developed in the document section for subtask 2 was rated to be easier to learn than those for subtasks 4 and 5. The Spearman correlation coefficient of 0.857 helps to explain that this significant difference among the subtasks is related to the subtask difficulty levels. In other words, subjects tended to think that if a subtask was more difficult to accomplish than others, the information development in its corresponding document section was rated to be not as easy to learn as those of less difficult subtasks.

The significant interaction effect may be partially to the relatively larger difference between the ratings of the two documents for subtasks 5 and 4. The subjects thought that the draft documents sections for subtasks 5 and 4 were not easy to learn. However, the ratings for the same sections for the modified document showed that subjects thought that the information was easy to learn. This relatively large difference in opinions shows that the modified document was significantly improved by incorporating critical incidents found for these subtasks.

5.5.8. Hypothesis 5.8

Hypothesis 5.8a was supported by the data: the subjects perceived the modified document to contain fewer inconsistencies in its presentation than the draft document. Even though there was this significant difference between the documents, subjects generally disagreed with the statement that annoying inconsistencies were present in the documents. The difference was in the degree of how strongly they disagreed. In this case, subjects disagreed more strongly with the statement for the modified document.

Hypothesis 5.8b was not supported by the data: there were no significant differences in the subjective opinions regarding annoying inconsistencies among the document sections for the seven subtasks.

5.5.9. Hypothesis 5.9

Hypothesis 5.9a was supported by the data: the modified document was rated to be easier to use than the draft document although this difference is between easy to use and fairly easy to use. This result is consistent with that observed for hypothesis 5.1.

Hypothesis 5.9b was also supported by the data: the document sections for subtasks 6, 2, 1, and 7 were rated to be easier to use than the document sections for subtasks 4 and 5. The Spearman correlation coefficient of 0.89 shows that subtasks that were perceived to be relatively more difficult were associated with document subsections that were not as easy to use as others. The change from the rating of slightly difficult to the rating of easy from the draft to modified document for document sections pertaining to subtasks 5 and 4 shows that the input of the critical incidents contributed to this improvement.

5.6. Conclusion

The results supported the overall hypothesis that the modified IRRAS document is easier to use than the draft IRRAS document. The first time task and complex task sections of the modified document allow for fewer user errors. These document sections also require less document reading time even though they are longer in text length. In some cases, faster reading time does not necessarily mean faster comprehension time. In this case, the test of comprehension was successfully completing the assigned subtask. Thus for the modified document, shorter reading time accounts for faster comprehension. The modified document contains sufficient amount of information to allow the user to complete the subtasks without additional help. For the draft document, some users may need additional help to accomplish complex subtasks.

Additionally, the sections of modified document for first time tasks and complex and difficult tasks are more accurate, more complete, easier to understand, easier to learn, easier to use, and contains fewer inconsistencies than those of the draft document.

The results show that the changes made in the draft document to develop the modified document significantly improved its usability. The critical incidents found in the first phase of this experiment clearly identified the problematic areas of the draft document. This is supported by the similar observations made in the comparison tests between the draft and modified documents for both objective and subjective measures. The significant differences in the objective and subjective measures of usability showed that the differences were due to changes made in the parts of the draft document using the observed critical incidents.

Moreover, this study showed that the critical incident technique is suitable for identifying critical problems associated with document usability. It was first used by Fitts and Jones (1947) to analyze pilot error. It has been used commonly since then, for example, by Cooper (1982) to investigate errors made by anesthesiologists. Recently, the critical incident technique has shown to be a useful tool in usability studies. For this particular study, this technique was used to find problems with a fault tree software document. Thus, the usefulness of critical incident technique extends throughout the field of safety in helping users identify potential hazards.

5.7. Future Studies

Few studies have experimentally compared the usability of evaluation methods. Therefore, usability studies are needed not only for determining the usability of documents but also for determining the usability of evaluation methods.

To examine how the critical incident technique compares to other techniques for its usefulness in identifying documentation usability problems, comparison studies are suggested. For example, a critical incident technique may be used to identify documentation usability problems and another method, such as expert opinion, may be used for the same purpose. Then the two modified documents can be experimentally compared for differences in usability. Several techniques could be tested across different types of documentation to determine which techniques are superior for identifying usability problems for different types of documentation

In addition to the early focus on users, Gould and Lewis (1985) proposed to use an iterative process for improving the usability of software. For this experiment, only one iteration of critical incident technique was used. Applying the critical incident technique to the modified document might improve its usability even more. Another iteration may serve to identify more usability problems of the document or it may serve to confirm that conducting one iteration of critical incident technique is sufficient.

In the second phase of this experiment, two subjective rating scales to assess users' opinions were neither consistently nor correctly used as previously discussed. In future studies, all rating scales should be designed to include a continuum shown as a line with scores and their respective anchors equally spaced out along its continuum. Instead of using mixed scale designs, one scale should be employed throughout the same experiment.

Lastly, since documentation functions to support software in a system such as IRRAS, a combination study of software and documentation is suggested to improve the entire system. Some of the usability problems found in the documentation can easily be traced to a software problem, and vice versa. Thus, a change in only one aspect of the system may not significantly improve the usability of the whole system.

5. REFERENCES

Andersson, B.E., and Nilsson, S.G. (1964). Studies in the reliability and validity of the critical incident technique. *Journal of Applied Psychology, 48* (6), 398-403.

Barr, J.P., and Rosenbaum, S. (1990). Documentation and training productivity benchmarks. *Technical Communication,* (4), 399-408.

Benson, P.J. (1985). Writing visually: Design considerations in technical publications. *Technical Communications, 32* (4), 35-39.

Bethke, F.J. (1983). Measuring the usability of software manuals. *Technical Communication, 29* (2), 13-16.

Brockmann, R.J. (1990). *Writing Better Computer User Documentation* (2nd ed.). New York: John Wiley & Sons, Inc.

Carroll, J.M. (1984). Minimalist Training. *Datamation, 30* (18), 125-136.

Casey, B. (1984). An information developer's view of usability testing. *Proceedings of the 31st International Technical Communication Conference*, MPD-22-MPD-25.

Cooper, J.B., Long, L.D., Newbower, R.S., and Philip, J.H. (1982). Critical incidents associated with intraoperative exchanges of anesthesia personnel. *Anesthesiology, 56*, 456-461.

del Galdo, E.M., Williges, R.C., and Williges, B.H., and Wixon, D.R. (1986a). A critical incident evaluation tool for software documentation. Presented at *Third Mid-Central Ergonomics/Human Factors Conference*, Miami University, Oxford, OH, June 18-20, 1986.

del Galdo, E.M., Williges, R.C., and Williges, B.H., and Wixon, D.R. (1986b). An evaluation of critical incidents for software documentation design. In *Proceedings of the Human Factors Society 30th Annual Meeting* (pp. 19-23). Santa Monica, CA: Human Factors Society.

DeMillo, R.A., McCracken, W.M., Martin, R.J., and Passafiume, J.F. (1987). *Software testing and evaluation.* Menlo Park, California: The Benjamin/Cummings Publishing Co., Inc.

Dzida, W., Herda, S., and Itzfeldt, W.D. (1978). User-perceived quality of interactive systems. *IEEE Transactions on Software Engineering*, SE-4, 4, 270-276.

Ericcson, K.A., and Simon, H.A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Fitts, P.M., and Jones, R.E. (1947). Psychological aspects of instrument display 1: Analysis of 270 "pilot error" experiences in reading and interpreting aircraft instruments. In H.W. Sinaiko (Ed.), *Selected papers on human factors in the design and use of control systems* (pp. 1-38). New York: Dover Publications, Inc.

Flower, L., Hayes, J.R., and Swartz, H. (1983). Revising functional documents: The scenario principle. In P. Anderson, R.J. Brockman, and Miller (Eds.), *New Essays in Technical and Scientific Communications* (pp. 41-58). Farmingdale, N.Y.: Baywood Publishing.

Foehr, T., and Cross, T.B. (1986). *The soft side of software.* New York: John Wiley & Sons.

Gillihan, D., and Herrin, J. (1988). Evaluating product manuals for increased usability. *Technical Communication,* (3), 168-171.

Gould, J.D., and Lewis, C. (1985). Designing for usability: Key principles and what designers think. *Communications of the ACM, 28* (3), 300-311.

Grim, S.J. (1982). *How to write computer manuals for users.* Belmont, CA: Lifetime Learning. 315-340.

Hetzel, B. (1988). *The complete guide to software testing.* Wellesley, MA: QWE Information Sciences, Inc.

Houston, K., and Southard, S. (1988). Organization: The essential element in producing usable software manuals. *Technical Communication,* (3), 179-187.

Major, J.H. (1989). What should you write: A user's guide, tutorial, reference manual, or standard operating procedure? *Technical Communication,* (2), 130-134.

McGehee, B.M. (1984). *Writing software user manuals.* Cincinnati, OH: Writer's Digest Books.

Mills, C.B., and Dye, K.L. (1985). Usability testing: user reviews. *Technical Communication, 32* (4), 40-44.

Neal, A.S., and Simons, R.M. (1984). Playback: A method for evaluating the usability of software and its documentation. *IBM Systems Journal, 23* (1), 82-96.

Nisbet, R.E., and Wilson, T.D. (1977). Telling more than we can know: Verbal Reports on Mental Processes. *Psychological Review, 84*, 231-259.

Nuclear Regulatory Commission. (1981). *Fault tree handbook.* NUREG-0492. Washington, D.C.: Author.

Price, J. (1984). *How to write a computer manual: A handbook of software documentation.* Menlo Park, CA: The Benjamin/Cummings Publishing Co.

Redish, J.C., Felker, D.B., and Rose, A.M. (1981). Evaluating the effects of document design principles. *Information Design Journal, 2* (3&4), 236-243.

Schriver, K.A. (1989). Document design from 1980 to 1989: Challenge that remain. *Technical Communication,* (4), 316-328.

Simpson, H., and Casey, S.M. (1988). *Developing Effective User Documentation.* New York: McGraw-Hill Book Company.

Southard, S.G. (1988). Practical considerations in formatting manuals. *Technical Communication,* (3), 173-178.

Spivey, N. (1987). Theory building in rhetoric and composition: The role of empirical scholarship. *Rhetoric Review, 7* (2) (Spring 1989), 272-288.

Winbush, B., and McDowell, G. (1980). Testing: How to increase the usability of computer manuals. *Technical Communication, 27* (4), 17-19.

Wright, P. (1987). Manual dexterity: A user-oriented approach to creating computer documentation. In R.M. Baecker and A.S. Buxton (Eds.), *Readings in Human-Computer Interaction: A Multidisciplinary Approach.* (pp. 613-320). Los Alton, CA: Morgan Kaufmann, Inc.

Wright, P. (1985). Is evaluation a myth? Assessing test assessment procedures. In D.H. Jonassen (Ed.), *Designing, and Displaying Texts* (pp. 418-435). Englewood Cliffs, N.J.: Educational Technology Publications.

Wright, P. (1980). Usability: The criterion for designing written information. In P.A. Kolers, M.E. Wrolstad, and H. Bouma (Eds.), *Processing of Visible Language 2*, (pp. 183-205). New York: Plenum Press.

APPENDIX A: CONSENT FORM

I, ______________________________, am participating in this research study because I want to. The decision to participate is completely voluntary on my part. No one has coerced or intimidated me to participate.

The experimenter has adequately answered any and all questions I have asked about this study, my participation, and procedures involved, which are described in the attached "EXPERIMENTAL INSTRUCTIONS," which I have initialled.

I recognize the research team as Samuel S. Lee, graduate student (231-9087), and Dr. Dennis L. Price, Faculty Advisor (231-5635).

I understand that they will be available to answer any questions concerning procedures throughout this study. I understand that if significant new findings develop during the course of this research which may relate to my decision to continue participation, I will be informed. I further understand that I may withdraw this consent at any time and discontinue further participation in this study without prejudice to my entitlements. I also understand that the experimenter for this study may terminate my participation in this study if he feels this to be in my best interest.

I understand that the entire experiment will take about 2 hours. I also understand that I shall receive payment in the amount of $5.00 per hour. However, I further understand that if I withdraw from the experiment before it is completed, I will be paid only for the time I actually spent performing in the experiment at $5.00 per hour.

I understand that the results of my efforts will be recorded and that I will be audio/videotaped. I consent to the use of this information for scientific or training purposes, and I understand that any records of my participation in this study may be disclosed only according to federal law, including the Federal Privacy Act, and its implementing regulations. This means that personal information will not be released to an unauthorized third party without my permission.

I understand that if I have any further questions about my rights as a participant, I may contact Dr. Ernest R. Stout, Chairman of the Institutional Review Board at VPI&SU, at (703) 231- 9359.

I FULLY UNDERSTAND THAT I AM MAKING A DECISION WHETHER OR NOT TO PARTICIPATE. MY SIGNATURE INDICATES THAT I HAVE DECIDED TO PARTICIPATE UNDER THE CONDITIONS DESCRIBED ABOVE.

______________________	______________________	____________
Signature	Printed Name	Date

______________________________________	____________________
Address	Phone Number

Safety & Environmental Engineering Laboratory, Department of ISE
Virginia Polytechnic Institute & State University, Blacksburg, VA. 24061

APPENDIX B: EXPERIMENTAL INSTRUCTIONS
(For Phase 1 Study)

Thank you for participating in this research. The purpose of this experiment is to evaluate a user documentation of a fault tree computer software, Integrated Reliability Risk Assessment Software (IRRAS). You have been asked to participate because of your basic knowledge of fault tree concepts. You will be expected to participate for one two-hour session.

As a participant, you will be asked to perform an 8-part benchmark task on a computer using the IRRAS fault tree software and its user documentation. These 8 subtasks include basic operation of the pop-up menus using the Logitec mouse, choosing and placing fault tree symbols, moving a diagram, copying a diagram, deleting a diagram, writing texts, editing texts, and printing a finished fault tree. The IRRAS user documentation will serve as the sole source of instruction for you to perform these subtasks. However, if you absolutely need additional help, please call on the experimenter.

You will need to refer to a specific and relevant section of the user documentation to perform each of the 8 subtasks. You will be given specific instructions to follow for each of the 8 subtasks. After the completion of each subtask, you are expected to report incidents involving that specific section of the user document you used to perform the subtask. An incident can be any aspect of the document specified for the given subtask which you perceive to be easy to understand and readily usable to accomplish that specific subtask. Or, conversely, an incident can be any aspect of that same part of the document which you would prefer to change to make it more understandable and readily usable. A sheet containing 3 questions to help you identify these incidents will be given to you for each of the 8 subtasks. You will need to answer all 3 questions on a 3X5 card per incident. You will be reminded by the experimenter to report incidents if you inadvertently skip them.

This experiment concludes when you finish identifying incidents for your 8th subtask. Thank you again for your participation.

APPENDIX C: DRAFT VERSION of the IRRAS DOCUMENT

Build

The BUILD menu contains the predefined fault tree shapes. The BUILD menu is entered by picking "BUILD" in the primary menu using the center-button on the mouse (for this primary menu command, the right-button also puts up the BUILD submenu). The BUILD submenu is shown in Figure 10.

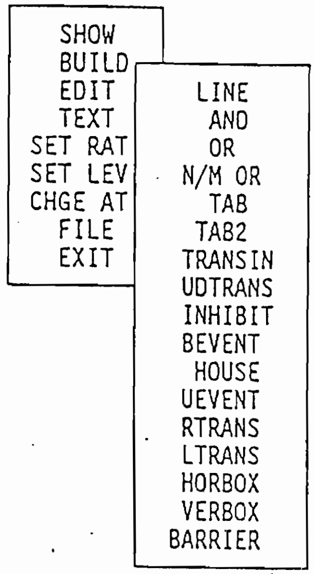


Figure 10. The BUILD submenu

The shapes corresponding to the names are shown in Figure 7. A shape is picked by moving the cursor to the shape name and pressing the right- button on the mouse. The center-button will put up another submenu and the left-button will take down the BUILD submenu (Table 1 lists the functions of the mouse buttons as used in the BUILD submenu). The picked shape is shown on the screen as a ghost which can be moved to the desired location. Once located, the ghost is permanently layed to rest by pressing the right-button a second time. The second depression of the right-button terminates the command. Another shape then can be picked from the menu. If the wrong shape is picked and detected before it is set in place, press the left-button on the mouse and the ghost is dropped, and another choice can be made. If a wrong shape is discovered after it has been set, it can be erased using the DELETE command in the EDIT menu.

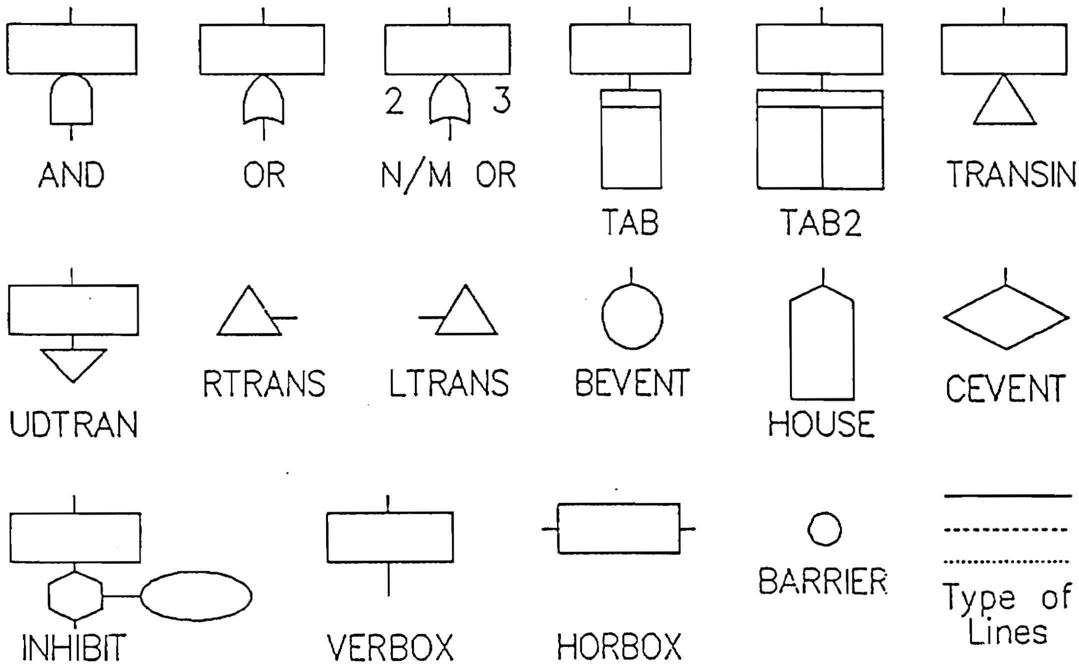


Figure 7. Fault tree basic shapes

Often it is desirable to put multiple copies of the same shape on the screen. To do this, the shape is again picked with the right-button. However, instead of placing the ghost with the right-button, the ghost is placed with the center-button. A permanent shape remains, but the ghost still exists to be placed elsewhere. The multiple copy option is terminated by placing the last shape using the right-button, or dropping the shape using the left-button.

The LINE command evokes the drawing of line segments. The line command is initiated by picking "LINE" using the right-button on the mouse. The screen cursor is moved to the start location of the first line segment, and the right-button is again pressed. This starts the drawing of line segments. Now as the cursor moves, a "rubber band" line is drawn from the last point. When the rubber band line is correct, press the right-button and the line is permanent. Continue drawing line segments in this fashion. When finished, depress the center or left button. The left-button will terminate the LINE command and another selection can be made. The center-button will terminate the current connected line segments, but will remain in the LINE command. A new line segment is started by pressing the right-button.

Text

Text and shape names are entered in the TEXT submenu. The TEXT submenu is entered by picking "TEXT" in the primary menu using the center-button on the mouse (for this primary menu command, the right-button also puts up the TEXT submenu). The TEXT submenu is shown in Figure 14.

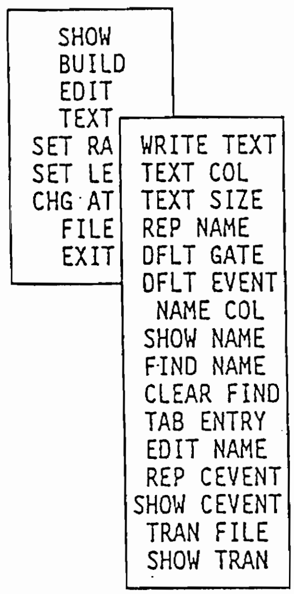


Figure 14. Text submenu

The WRITE TEXT command initiates entry of descriptive text. The WRITE TEXT command is selected using the right-button on the mouse. Move the cursor to the location for the center of the first line of text and press the right-button. Descriptive text can be placed anywhere on the screen. The text is entered after the prompt at the bottom of the screen. Multiple stacked lines of text can be entered. To terminate entry of text, press the "Enter" key as the only entry. The first descriptive text line is centered about the specified location point with additional lines placed under the previous line and centered.

Edit

In the EDIT submenu, parts of a diagram can be moved, copied or deleted and events in TAB or TAB2 gates can be modified. The EDIT submenu is entered by picking "EDIT" in the primary menu using the center-button on the mouse (for this primary menu command, the right-button also puts up the EDIT submenu). The EDIT submenu is shown in Figure 12.

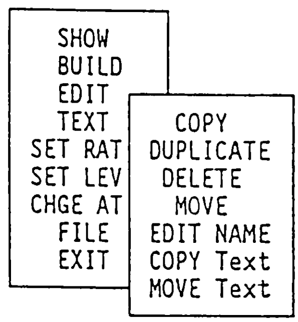


Figure 12. The Edit submenu

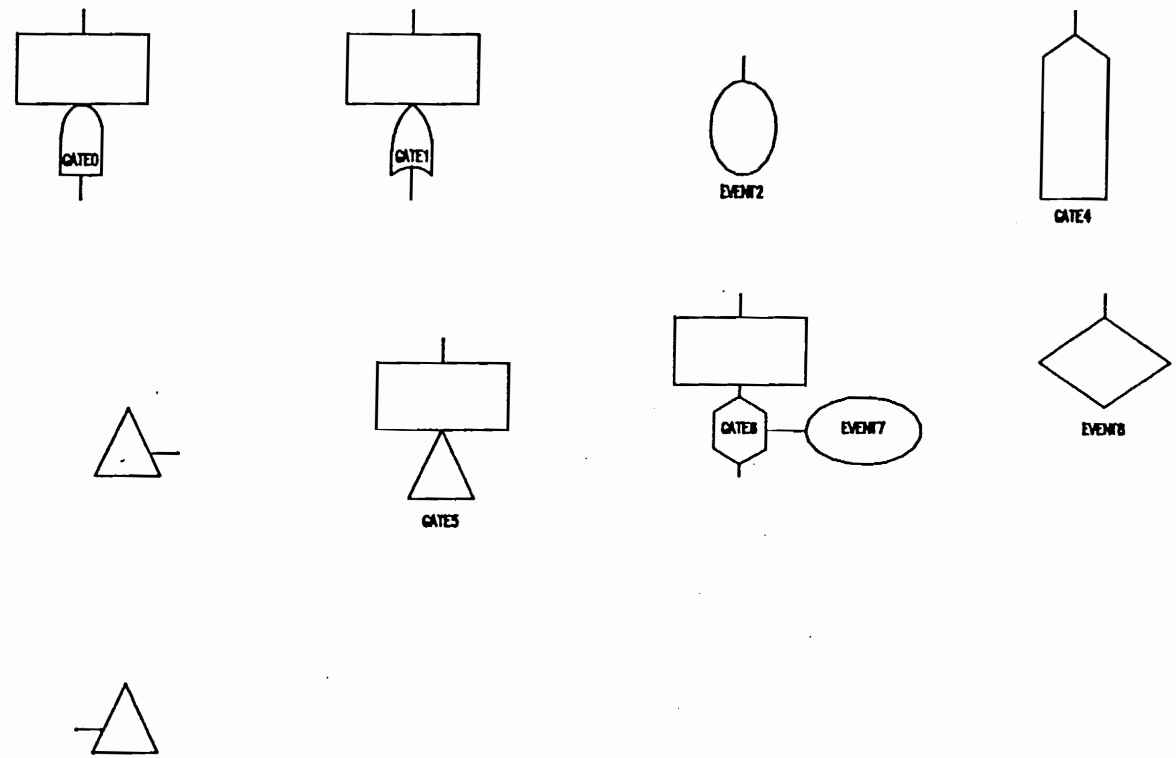
All selections in the EDIT submenu are made using the right-button on the mouse.

To copy a section of the drawing, pick "COPY" or "COPY Text" in the EDIT submenu. The COPY command copies all objects while the COPY Text command copies only text. Place the screen cursor at one corner of the box encompassing the region to be copied and press the right-button. Move to the other corner of the box. A rubber band box will appear as the cursor is moved. Once the box is properly located, press the right-button on the mouse. All items to be copied must lie completely within the box. Pick a reference point for copying. Move the cursor to the location to place the copy and again press the right-button. All parts of the chosen region will be copied to the new location with the reference point lying on the new point. The copied gates and events will be given new default names.

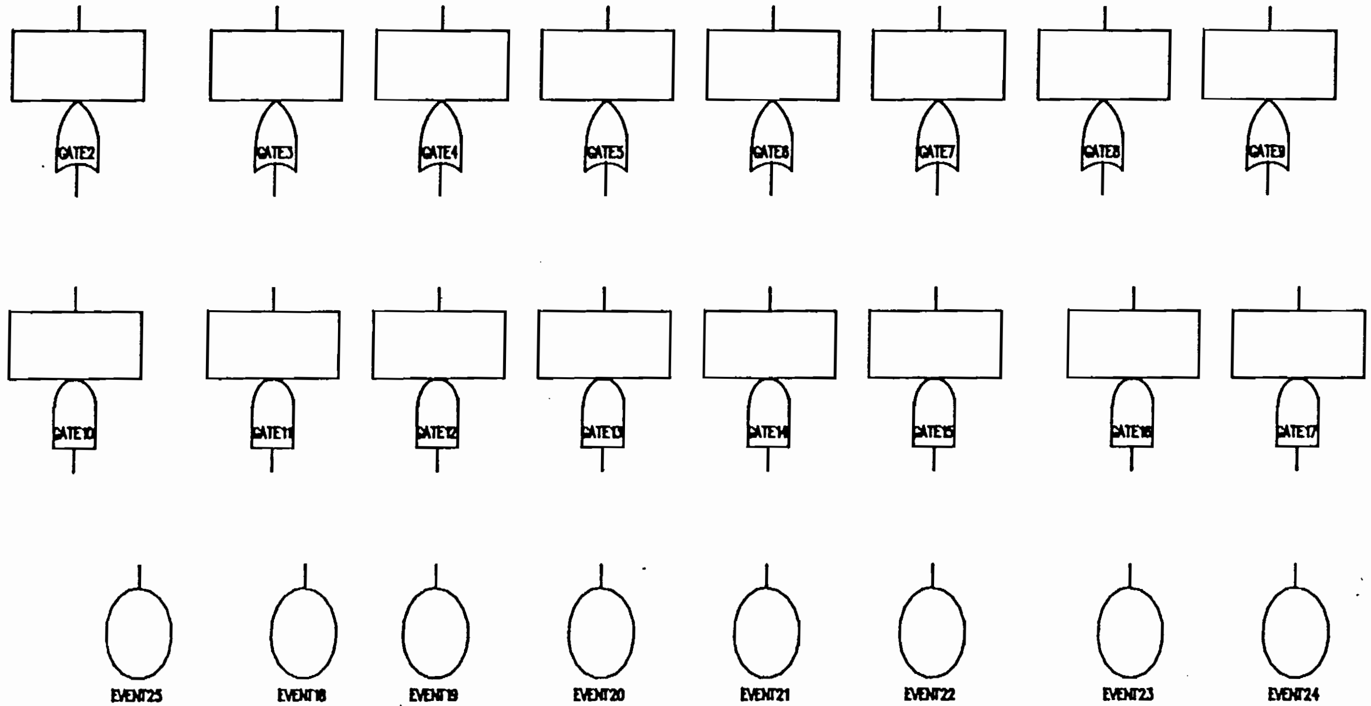
To move a section of the drawing, pick "MOVE" or "MOVE Text" in the EDIT submenu. The MOVE command moves all objects while the MOVE Text command moves only text. Place the screen cursor at one corner of the box encompassing the region to be moved. Press the right-button. Move to the other corner of the box. A rubber band box will appear as the cursor is moved. Once the box is properly located, press the right-button on the mouse. All items completely within the box will be moved. Using the screen cursor, pick a reference point for moving. Move the cursor to the new location to place the region and again press the right-button. All parts of the chosen region will be moved to the new location with the reference point lying on the new point.

To delete a section of the drawing, pick "DELETE" in the EDIT submenu. Place the screen cursor at one corner of the box encompassing the region to be deleted. Press the right-button. Move to the other corner of the box. A rubber band box will appear as the cursor is moved. Once the box is properly located, press the right-button on the mouse. All items completely within the box will be deleted.

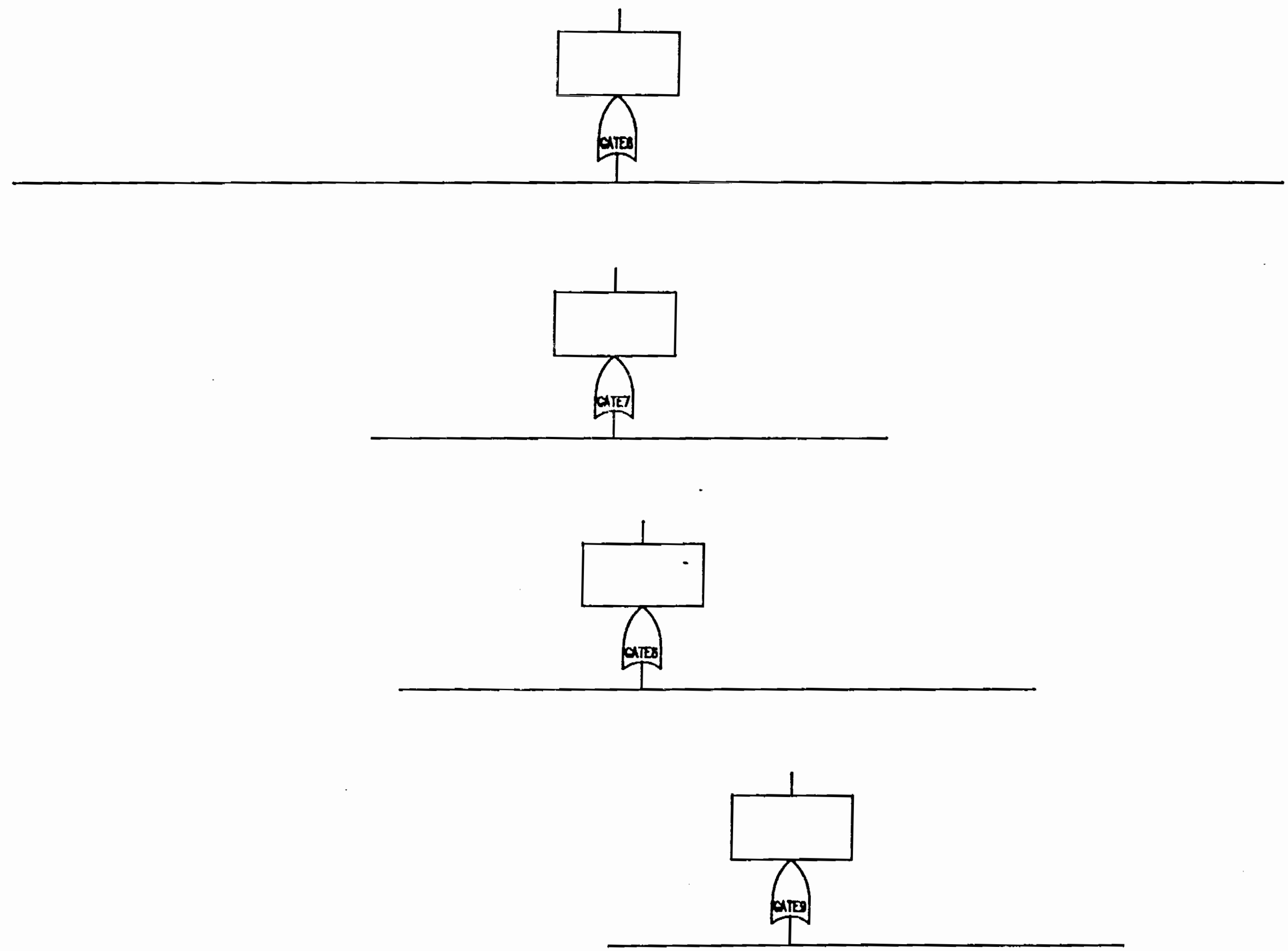
APPENDIX D. FINISHED SUBTASKS AS SHOWN ON THE SCREEN



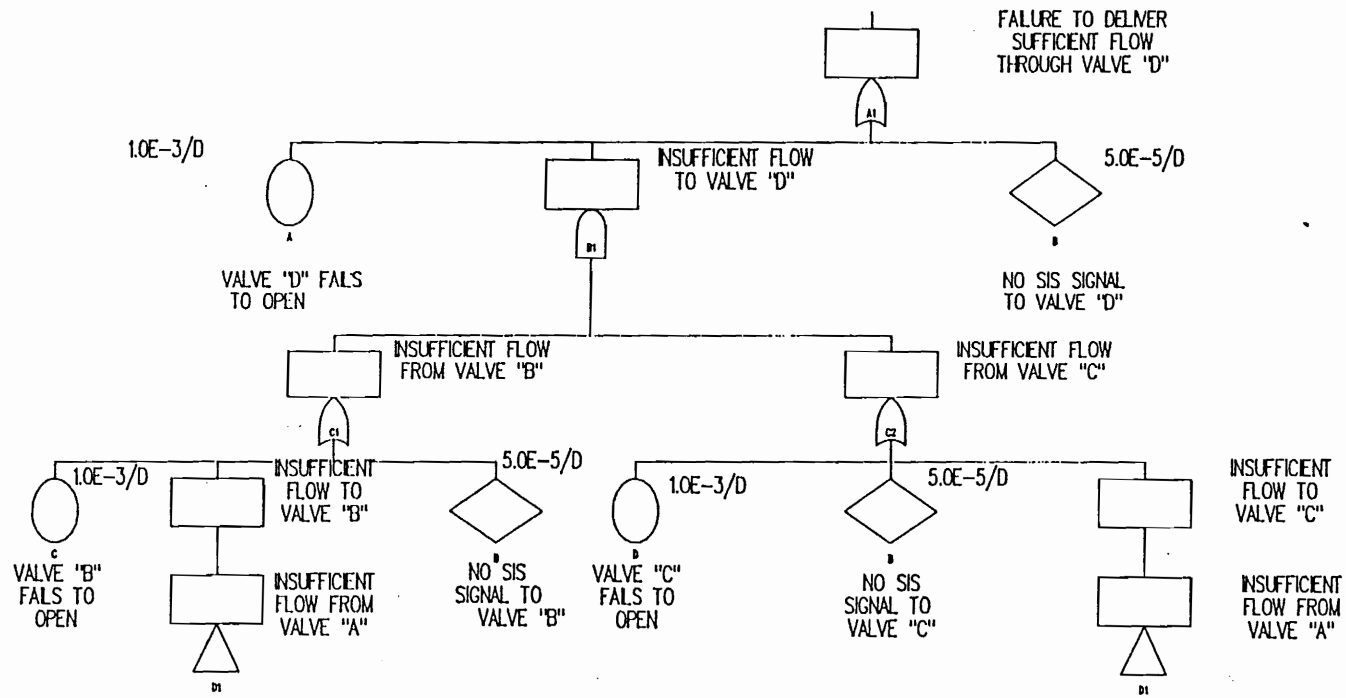
SUBTASK 1: PICKING AND PLACING SYMBOLS



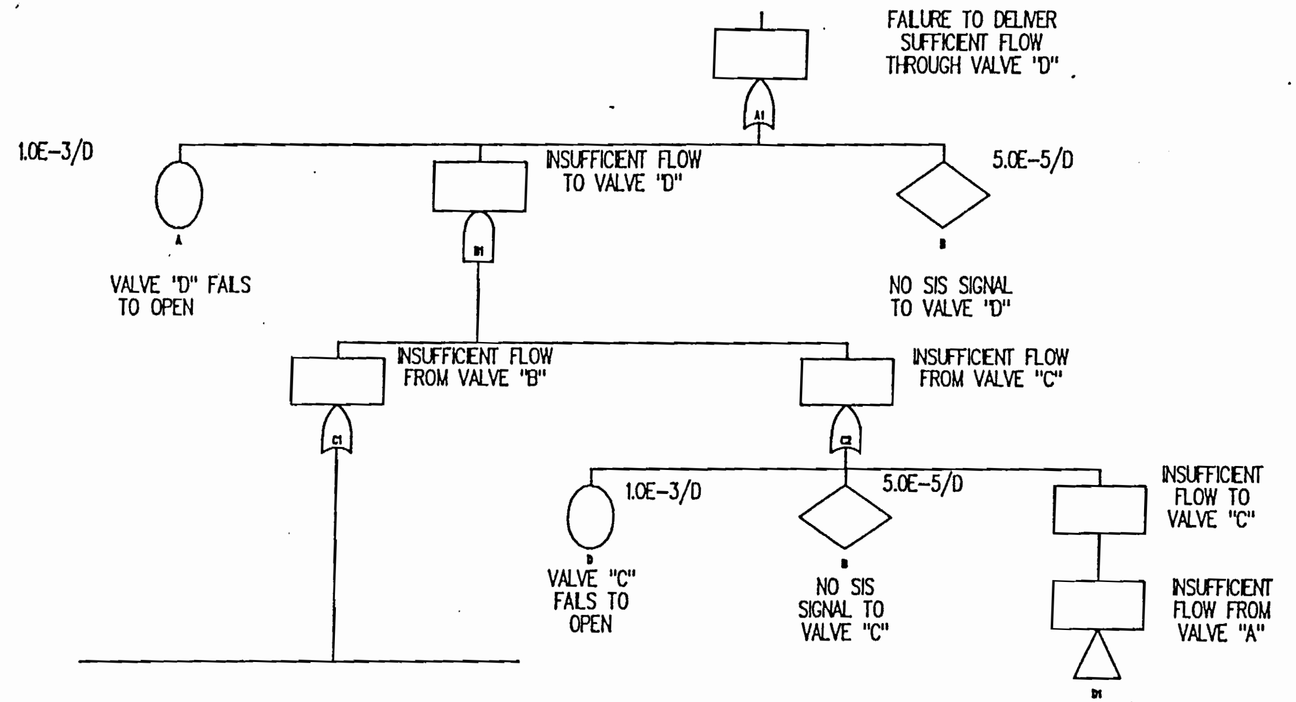
SUBTASK 2: PICKING AND PLACING MULTIPLE COPIES OF SYMBOLS

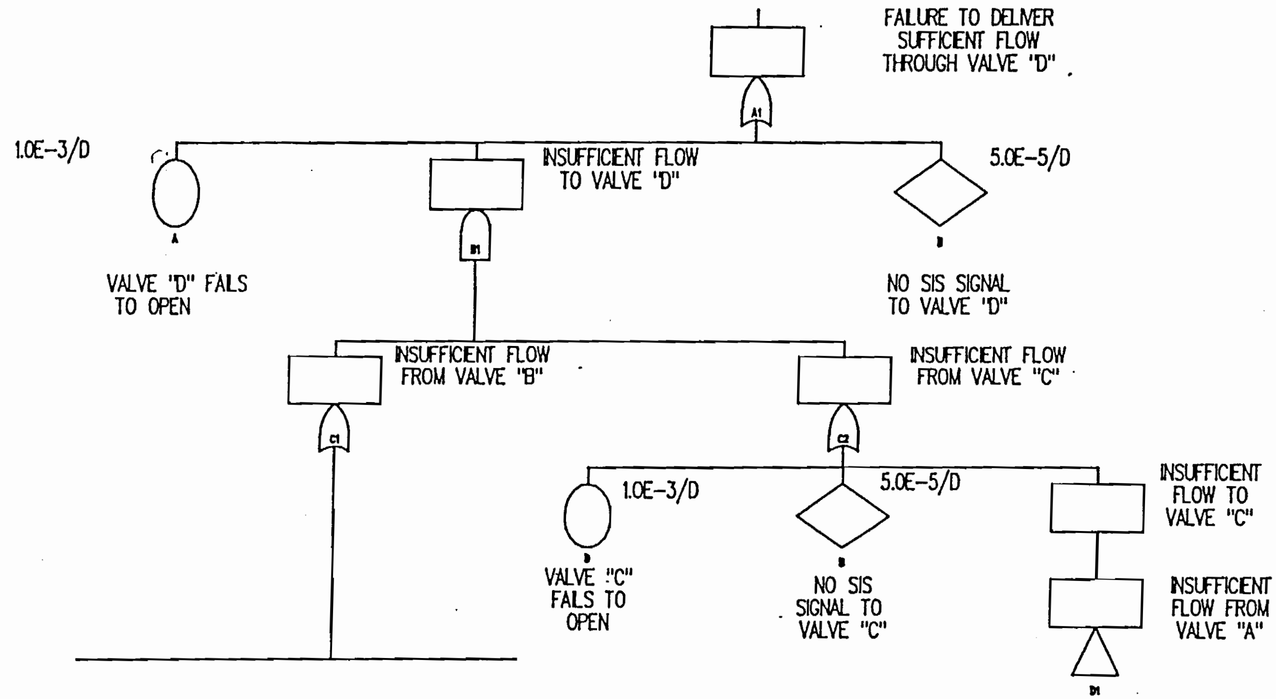


SUBTASK 3: DRAWING LINES UNDER THE SYMBOLS

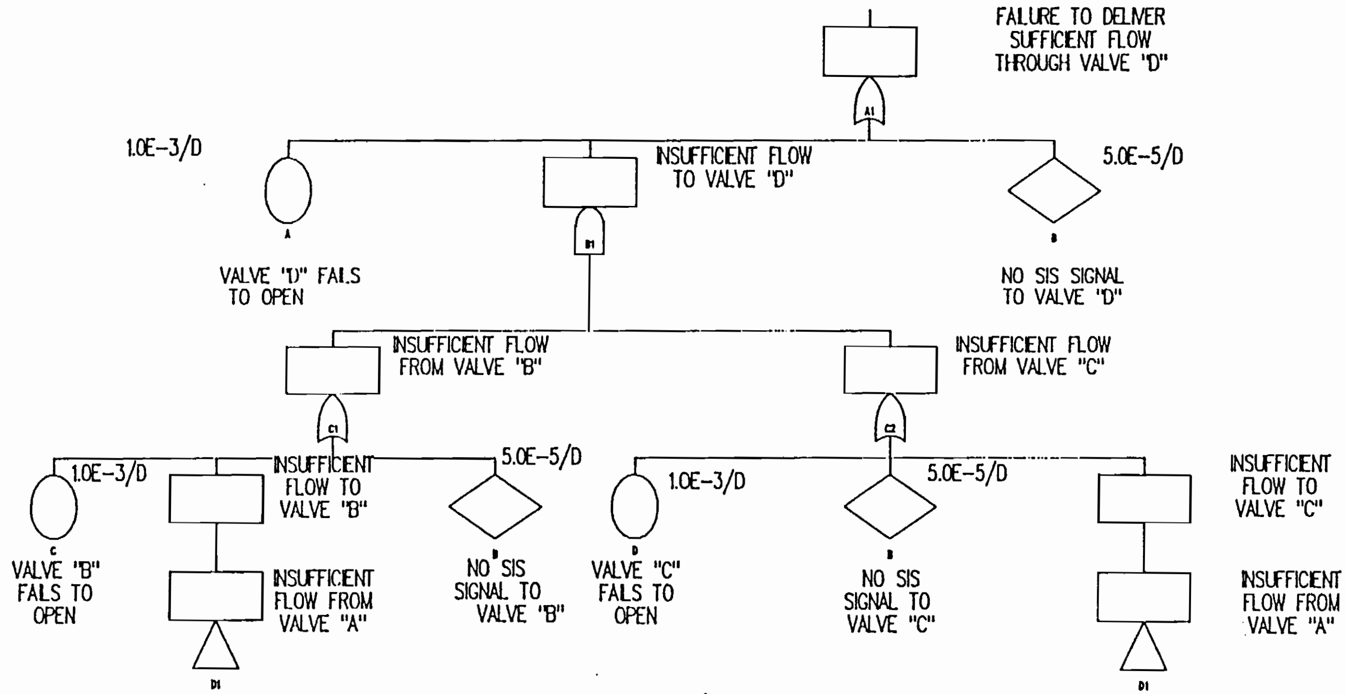


SUBTASK 4: WRITING TEXTS NEAR THE SYMBOLS

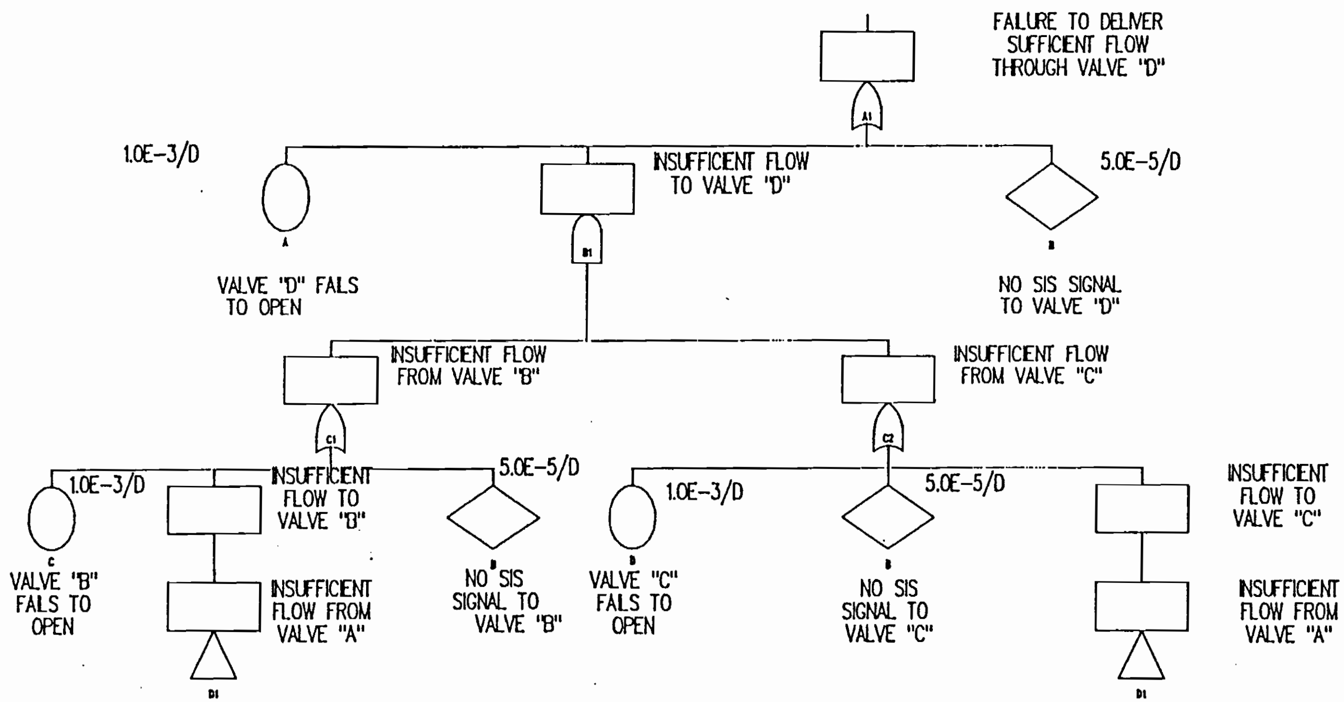

FALURE TO DELIVER
SUFFICENT FLOW
THROUGH VALVE "D"
A1
1.0E-3/D
VALVE "D" FALS
TO OPEN
INSUFFICIENT FLOW
TO VALVE "D"
B1
5.0E-5/D
NO SIS SIGNAL
TO VALVE "D"
INSUFFICIENT FLOW
FROM VALVE "B"
C1
INSUFFICIENT FLOW
FROM VALVE "C"
C2
1.0E-3/D
VALVE "C"
FALS TO
OPEN
5.0E-5/D
NO SIS
SIGNAL TO
VALVE "C"
INSUFFICIENT
FLOW TO
VALVE "C"
INSUFFICIENT
FLOW FROM
VALVE "A"



SUBTASK 6: MOVE EDITING



SUBTASK 7: DELETE EDITING



SUBTASK 8: CONSTRUCTING A FAULT TREE

APPENDIX E: MODIFIED VERSION of the IRRAS DOCUMENT

IRRAS USER DOCUMENTATION

The Integrated Reliability and Risk Assessment Software (IRRAS) is a menu driven software. Its primary menu is shown (opened) when the center button of the mouse is clicked. This kind of menu is called a "pop-up menu." Before "popping-up" (opening) this primary menu, the cursor can be positioned anywhere on the screen by moving the mouse. The position of the cursor is where the top left corner of the primary menu will appear when the center mouse button is clicked.

IMPORTANT!

The word "CLICK" as used for example in "Click the center button" ALWAYS means Press & Release the button immediately.

To "pop-up" (open) the primary menu,

1) **Move the cursor to a desired location on the screen by moving the mouse.** (Position of the cursor is where the top left corner of the primary menu will appear)

2) **Click** (press & release immediately) **the CENTER mouse button** (❑❑❑).

The primary menu should appear as shown in Figure 1.

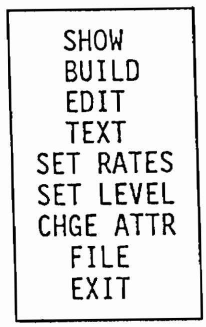


Figure 1. BUILD module primary menu

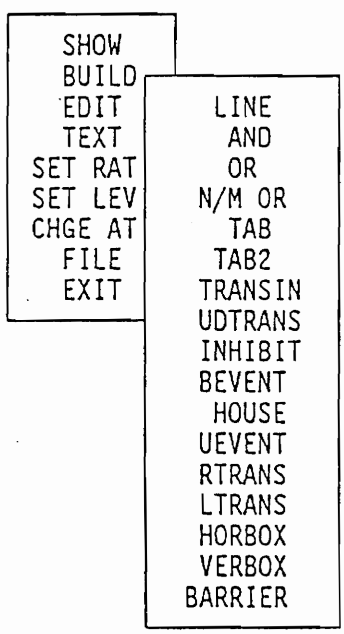


Figure 2. The BUILD menu

BUILD: Shapes

The Build submenu contains the predefined fault tree shapes. The BUILD submenu is entered by picking "BUILD" in the primary menu using the center button on the mouse.

To enter into the BUILD menu from the primary menu,

1) Place the cursor on the word "BUILD" inside the primary menu by moving the mouse.

2) Click the CENTER mouse button (❑▣❑).

The BUILD submenu is shown in Figure 2.
The shapes (symbols of a fault tree) corresponding to the names are shown in Figure 3.

To pick a shape (e.g., BEVENT),

1) Place the cursor on the word "BEVENT" using the mouse.

2) Click the RIGHT mouse button (❑❑▣). (Make sure the button is RELEASED!)

The picked shape is shown on the screen as a GHOST (temporary shape) which can be moved to a desired location and placed there by using the mouse.

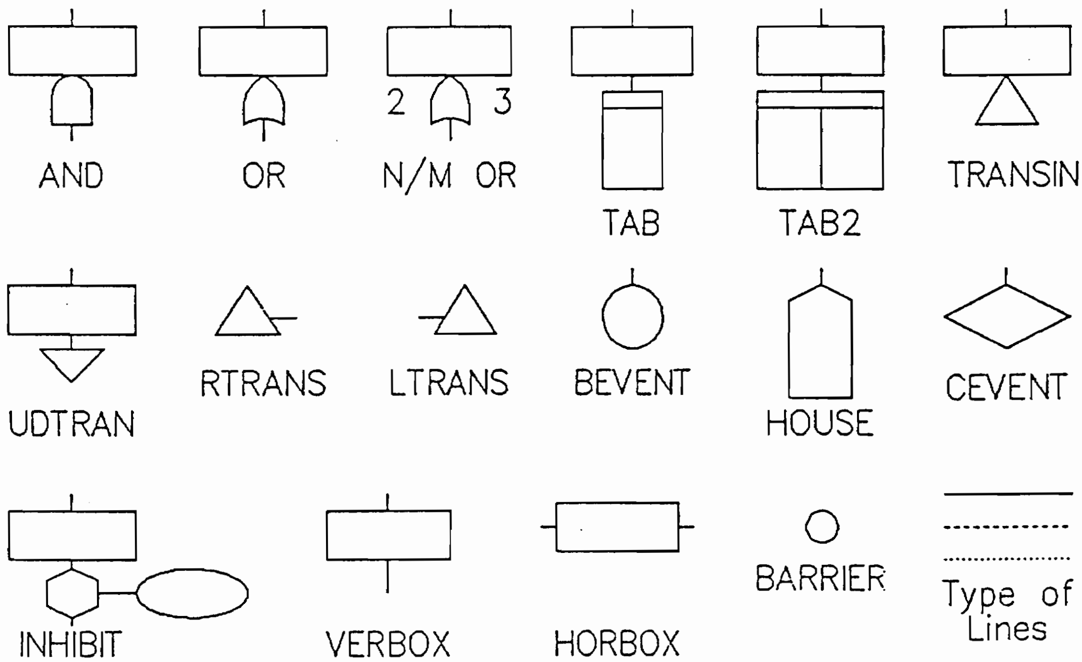


Figure 3. Fault tree symbols

To move the shape (in its ghost form) to a desired location and place it there,

1) WITHOUT touching any button, move (drag) **the GHOST to a desired location by moving the mouse.**

WARNING: Do NOT press and hold the right button down when dragging the ghost to a desired location. You may be tempted to do this if you are used to doing it using other computer hardwares &/or softwares.

2) Click the RIGHT mouse button (❑❑❑).

This click of the right button terminates the command to pick a shape and place it on a desired location. The ghost immediately turns into a permanent shape.

BUILD: Multiple Copies of Shapes

To put MULTIPLE copies of the same shape on the screen (e.g., AND-gate),

1) **Move the cursor to the word "AND" inside the BUILD menu using the mouse.**

2) **Click the RIGHT mouse button (❑❑▣).**

3) **Move the GHOST to a desired location by moving the mouse (WITHOUT clicking any button).**

4) **Click the CENTER mouse button (❑▣❑).**

5) **Repeat steps 3 and 4 for additional copies.**

6) **To terminate this multiple copy option, move the LAST GHOST to a desired location by moving the mouse (WITHOUT clicking any mouse button). Then, click the RIGHT mouse button (❑❑▣).**

OR

Click the LEFT mouse button (▣❑❑) to eliminate the LAST GHOST.

BUILD: Line

To "pop-up" (open) the primary menu,

1) **Move the cursor to a desired location on the screen by moving the mouse.** (Position of the cursor is where the top left corner of the primary menu will appear)

2) **Click** (press & release immediately) **the CENTER mouse button** (❑▣❑).

The primary menu should appear as shown in Figure 1.

The BUILD submenu is entered by picking "BUILD" in the primary menu using the CENTER mouse button.

To enter into the BUILD menu from the primary menu,

1) **Place the cursor on the word "BUILD" inside the primary menu by moving the mouse.**

2) **Click the CENTER mouse button** (❑▣❑).

The BUILD submenu is shown in Figure 2.
The shapes (symbols of a fault tree) corresponding to the names are shown in Figure 3.

The LINE command in the BUILD menu evokes the drawing of line segments. The line command is initiated by picking "LINE" using the right button on the mouse.

To pick and draw a line,

1) **Place the cursor on the word "LINE" using the mouse.**

2) **Click the RIGHT mouse button** (□□▣). ("Enter points for line" will appear on the lower left corner of the screen.)

3) **Move the cursor to your desired starting point location of the line by moving the mouse (Remember, without touching any mouse button!).**

4) **Click (Remember, press and release immediately!) the RIGHT mouse button** (□□▣).

5) **Move the cursor to the desired end point location of the line by moving the mouse.** (You can move the cursor to anywhere on the screen to form a line as well as elongate or shorten a line).

6) **Click the CENTER mouse button** (□▣□). (This terminates the line segment; however, you are still in the LINE command. "Enter points for line" still appears on the bottom left corner of the screen.)

7) **Click the LEFT mouse button** (▣□□) **to get out of the LINE command mode.**

Or

If you want to continue to draw additional lines, **Repeat steps #3 through #6 immediately after step #6** since you are still in the LINE command mode.

TEXT: Write Text

To "pop-up" (open) the primary menu.

1) **Move the cursor to a desired location on the screen by moving the mouse.** (Position of the cursor is where the top left corner of the primary menu will appear)

2) **Click** (press & release immediately) **the CENTER mouse button** (❑▣❑).

The primary menu should appear as shown in Figure 1.

The TEXT submenu is entered by picking "TEXT" in the primary menu using the CENTER mouse button.

To enter into the TEXT submenu from the primary menu.

1) **Place the cursor on the word "TEXT" inside the primary menu by moving the mouse.**

2) **Click the CENTER mouse button** (❑▣❑).

The TEXT submenu is shown in Figure 2.

The WRITE TEXT command in the TEXT submenu initiates entry text.

To write text,

1) Place the cursor on the word "WRITE TEXT" using the mouse.

2) Click the RIGHT mouse button (❑❑▣). ("*Pick location for description*" will appear on the bottom left corner of the screen.)

3) Move the cursor to your desired location for the center of the first line of text. (Text can be placed anywhere on the screen. Text will be centered and the first line will be positioned above the center point of the cursor. The text will always be centered. It can NOT be left justified or right justified.)

4) Click the RIGHT mouse button (❑❑▣). ("*Enter label (Terminate with CR)>*" will be shown on the bottom left corner of the screen. CR stands for "Carriage Return.")

5) Type in the first line of the text using the keyboard.

6) Press and release the ENTER button on the keyboard. (This first text line will appear as centered about the specified location on the screen. "*Next label (Terminate with CR)>*" will be shown on the bottom left corner of the screen.)

7) Type in the second line of the text using the keyboard.

8) Press and release the ENTER button on the keyboard. (This second text line will appear under the first text line and centered as well. "*Next label (Terminate with CR)>*" will still be shown on the bottom left corner of the screen.)

9) If you do NOT want to add more text, **Press and release the ENTER button on the keyboard again.** (This concludes the entry of text.)

Or

Repeat steps #7 to #8 to type in as many lines as you desire. Then, press and release the ENTER button on the keyboard once more to conclude the entry of text.

EDIT: Copy

To "pop-up" (open) the primary menu.

1) Move the cursor to a desired location on the screen by moving the mouse. (Position of the cursor is where the top left corner of the primary menu will appear)

2) Click (press & release immediately) **the CENTER mouse button** (❑▣❑).

The primary menu should appear as shown in Figure 1.

The EDIT submenu is entered by picking "EDIT" in the primary menu using the CENTER mouse button.

To enter into the EDIT submenu from the primary menu.

1) Place the cursor on the word "EDIT" inside the primary menu by moving the mouse.

2) Click the CENTER mouse button (❑▣❑).

The EDIT submenu is shown in Figure 2.

To copy a section of the drawing,

1) **Place the cursor on the word "COPY" using the mouse.**

2) **Click the RIGHT mouse button** (□□▣). ("Pick region to be copied" will appear on the lower left corner of the screen.)

3) Select the objects (or a region of the fault tree) to be copied by drawing a selection box completely around them. To do this, first, decide (imagine) in your mind where this box will be placed. Next, **move the cursor to the top left corner of your imagined box.**

4) **Click the RIGHT mouse button** (□□▣). (Click it gently and quickly.)

5) **Move the cursor to the bottom right corner of your box using your mouse (Remember, do not touch any button!).** (The box will form in continuously varying sizes as you move your cursor to the desired location of the bottom right corner of the box. All Items (symbols, lines, and texts) to be copied MUST lie completely within this box. Any item that is only partially inside the box will NOT be affected.)

6) **Click the RIGHT mouse button** (□□▣). (The box will disappear and "*Pick a reference point*" will be shown on the bottom left corner of the screen.)

7) Pick a point on any part of the region (e.g., symbol(s), line(s), or a branch of symbol(s) and line(s)) that was previously boxed-in. This is the reference point. To do this, **move the cursor to a desired point on the region that was previously boxed-in.** (The best point is the top most point on the region or items to be copied, e.g. vertical stem on the top of an OR-gate).

8) **Click the RIGHT mouse button** (□□▣). ("*Pick new location*" will be shown on the bottom left corner of the screen.)

9) Pick a point of the new location where the copy will be placed. This point is where the reference point of the previously boxed-in region will be copied onto. To do this, **move the cursor to this new point using the mouse.**

10) **Click the RIGHT mouse button** (□□▣). (All parts of the chosen region will be copied to the new location with the reference point lying on the new point.)

EDIT: Move

To "pop-up" (open) the primary menu.

1) **Move the cursor to a desired location on the screen by moving the mouse.** (Position of the cursor is where the top left corner of the primary menu will appear)

2) **Click** (press & release immediately) **the CENTER mouse button** (❑▣❑).

The primary menu should appear as shown in Figure 1.

The EDIT submenu is entered by picking "EDIT" in the primary menu using the CENTER mouse button.

To enter into the EDIT submenu from the primary menu.

1) **Place the cursor on the word "EDIT" inside the primary menu by moving the mouse.**

2) **Click the CENTER mouse button** (❑▣❑).

The EDIT submenu is shown in Figure 2.

To move a section of the drawing,

1) **Place the cursor on the word "MOVE" using the mouse.**

2) **Click the RIGHT mouse button** (❑❑❑). ("Pick region to be moved" will appear on the lower left corner of the screen.)

3) Select the objects (or a region of the fault tree) to be moved by drawing a selection box completely around them. To do this, first, decide (imagine) in your mind where this box will be placed. Next, **move the cursor to the top left corner of your imagined box.**

4) **Click the RIGHT mouse button** (❑❑❑). (Click it gently and quickly.)

5) **Move the cursor to the bottom right corner of your box using your mouse (Remember, do not touch any button!).** (The box will form in continuously varying sizes as you move your cursor to the desired location of the bottom right corner of the box. All items to be moved MUST lie completely within this box. Any item that is only partially inside the box will NOT be affected.)

6) **Click the RIGHT mouse button** (❑❑❑). (The box will disappear and "*Pick a reference point*" will be shown on the bottom left corner of the screen.)

7) Pick a point on any part of the region (e.g., symbol(s), line(s), or a branch of symbol(s) or line(s)) that was previously boxed-in. This is the reference point. To do this, **move the cursor to that desired point on the region that was previously boxed-in.** (The best point is the top most point on the region or items to be moved, e.g. vertical stem on the top of an OR-gate).

8) **Click the RIGHT mouse button** (❑❑❑). ("*Pick new location*" will be shown on the bottom left corner of the screen.)

9) Pick a point of the new location where the chosen region or items will move to. This point is where the reference point of the boxed-in region will be moved to. To do this, **move the cursor to this new point using the mouse.**

10) **Click the RIGHT mouse button** (❑❑❑). (All parts of the chosen region will be moved to the new location with the reference point lying on the new point.)

EDIT: Delete

To "pop-up" (open) the primary menu.

1) **Move the cursor to a desired location on the screen by moving the mouse.** (Position of the cursor is where the top left corner of the primary menu will appear)

2) **Click** (press & release immediately) **the CENTER mouse button (❑⊟❑).**

The primary menu should appear as shown in Figure 1.

The EDIT submenu is entered by picking "EDIT" in the primary menu using the CENTER mouse button.

To enter into the EDIT submenu from the primary menu.

1) **Place the cursor on the word "EDIT" inside the primary menu by moving the mouse.**

2) **Click the CENTER mouse button** (❑⊟❑).

The EDIT submenu is shown in Figure 2.

To delete a section of the drawing,

1) **Place the cursor on the word "DELETE" using the mouse.**

2) **Click the RIGHT mouse button** (□□▣). ("Pick region to be deleted" will appear on the lower left corner of the screen.)

3) Select the objects (or a region of the fault tree) to be deleted by drawing a selection box completely around them. To do this, first, decide (imagine) in your mind where this box will be placed. Next, **move the cursor to the top left corner of your imagined box.**

4) **Click the RIGHT mouse button** (□□▣). (Click it gently and quickly.)

5) **Move the cursor to the bottom right corner of your box using your mouse (Remember, do not touch any button!).** The box will form in continuously varying sizes as you move your cursor to the desired location of the bottom right corner of the box. All Items (symbols, lines, and texts) to be deleted MUST lie completely within this box. Any item that is only partially inside the box will NOT be affected.)

6) **Click the RIGHT mouse button** (□□▣). (All items completely inside the box will be deleted.)

APPENDIX F: EXPERIMENTAL INSTRUCTIONS (For Phase 2 Study)

Thank you for participating in this research. The purpose of this experiment is to evaluate a user documentation of a fault tree computer software, Integrated Reliability Risk Assessment Software (IRRAS). You have been asked to participate because of your basic knowledge of fault tree concepts. You will be expected to participate for one two-hour session.

As a participant, you will be asked to perform an 8-part benchmark task on a computer using the IRRAS fault tree software and its user documentation. These 8 subtasks include basic operation of the pop-up menus using the Logitec mouse, choosing and placing fault tree symbols, moving a diagram, copying a diagram, deleting a diagram, writing texts, editing texts, and printing a finished fault tree. The IRRAS user documentation will serve as the sole source of instruction for you to perform these subtasks. However, if you absolutely need additional help, please call on the experimenter.

You will need to refer to a specific and relevant section of the user documentation to perform each of the 8 subtasks. You will be given specific instructions to follow for each of the 8 subtasks. As soon as you start to read the given section of the document, you will be timed until the subtask is completely and correctly performed as shown on the screen or printed on a plotter paper. If you make an error, please correct the error. After the completion of each of the 8 subtasks, you will be asked to respond to a questionnaire containing 9 questions.

After you finish filling out the questionnaire for your 8th subtask, you will be asked some questions regarding the use of the IRRAS documentation. Thank you again for your participation.

APPENDIX G:

READING PASSAGES for the READING ABILITY TESTING

The atmosphere contains water vapor, but there is a limit to how much water can be evaporated into a given volume of air, just as there is a limit to how much sugar can be dissolved in one cupful of coffee. More sugar can be dissolved in hot coffee than in cold. A given volume of air can hold more water vapor at a higher temperature than at a lower temperature. The air is said to be *saturated* when it holds as much water vapor as it can at that temperature. At 20° C a cubic meter of air can hold about 17 gm of water vapor; at 30° C it can hold about 30 gm. Usually the atmosphere is not saturated. *Relative humidity* (expressed in per cent) is the ratio of the mass of water vapor actually present in a given volume of air to the mass which would be present in it if it were saturated. For example, if a cubic meter of air at 20° C contains 12 gm of water vapor, the relative humidity is $\frac{12 \text{ gm}}{17 \text{ gm}} \times 100 = 71\%$. Hygrometers are instruments for measuring relative humidity. Readings on wet and dry bulb thermometers can be compared with the aid of a chart from which one can then read off the relative humidity. The basic principle of this is that evaporation is a cooling process. The rate of evaporation from the wet-bulb thermometer will be high when the relative humidity is low, and therefore on such a day the wet-bulb thermometer will read considerably below the dry-bulb one. There is no simple formula for converting this temperature difference to relative humidity, and therefore a chart is used.

If unsaturated air is cooled, its relative humidity goes up. If the temperature of the air drops sufficiently, saturation is reached and excess moisture precipitates out. The *dew point* is the temperature to which the air must be cooled so that it will be saturated and condensation will just form.

90. When the readings on the wet bulb thermometer and the dry bulb thermometer are similar, we may assume that
    1. we have relative humidity
    2. the air is saturated
    3. the thermometers are inefficient
    4. the temperature is going to rise
    5. the temperature is about to fall
91. When the air is saturated, it is likely to be
    1. raining
    2. windy
    3. clear
    4. getting warmer
    5. relative
92. The dew point is most often reached
    1. early in the morning
    2. at noon
    3. late in the afternoon
    4. at dusk
    5. after midnight
93. A chart is used to determine the relative humidity after using a wet and dry bulb thermometer because
    1. it comes with the instrument
    2. the mathematics involved is complicated
    3. there is no need to duplicate the work
    4. people do not know how to handle per cent
    5. cool air makes the relative humidity rise
94. When the temperature of the air rises above the dew point,
    1. dew will form
    2. it will rain
    3. the relative humidity exceeds 100%
    4. evaporation is likely to take place
    5. the hygrometer will become inaccurate

We now know that what constitutes practically all of matter is empty space; relatively enormous voids in which revolve with lightning velocity infinitesimal particles so utterly small that they have never been seen or photographed. The existence of these particles has been demonstrated by mathematical physicists and their operations determined by ingenious laboratory experiments. It was not until 1911 that experiments by Sir Ernest Rutherford revealed the architecture of the mysterious atom. Moseley, Bohr, Fermi, Millikan, Compton, Urey, and others have also worked on the problem. Matter is composed of molecules whose average diameter is about $1/125$ millionth of an inch. Molecules are composed of atoms so small that about five million could be placed in a row on the period at the end of this sentence. Long thought to be the ultimate, indivisible constitutent of matter, the atom has been found to consist roughly of a proton, the positive electrical element in the atomic nucleus, surrounded by electrons, the negative electric elements swirling about the proton.

101. The center of the atom, according to this passage,
   1. contains one electron
   2. was seen as early as 1911
   3. has not yet been seen by the naked eye
   4. is about the size of a period
   5. might be photographed under microscopes

102. The paragraph indicates that the atom
   1. is the smallest particle
   2. is very little larger than a molecule
   3. is composed of several particles
   4. has been seen
   5. is empty space

103. Scientists agree that molecules are
   1. voids
   2. the most mysterious particles
   3. not divisible
   4. not basically composed of electric elements
   5. huge compared with electrons

Why does the Foundation concentrate its support on basic rather than applied research? Basic research is the very heart of science, and its cumulative product is the capital of scientific progress, a capital that must be constantly increased as the demands upon it rise. The goal of basic research is understanding, for its own sake. Understanding of the structure of the atom or the nerve cell, the explosion of a spiral nebula or the distribution of cosmic dust, the causes of earthquakes and droughts, or of man as a behaving creature and of the social forces that are created whenever two or more human beings come into contact with one another—the scope is staggering, but the commitment to truth is the same. If the commitment were to a particular result, conflicting evidence might be overlooked or, with the best will in the world, simply not appreciated. Moreover, the practical applications of basic research frequently cannot be anticipated. When Roentgen, the physicist, discovered X-rays, he had no idea of their usefulness to medicine.

Applied research, undertaken to solve specific practical problems, has an immediate attractiveness because the results can be seen and enjoyed. For practical reasons, the sums spent on applied research in any country always far exceed those for basic research, and the proportions are more unequal in the less developed countries. Leaving aside the funds devoted to research by industry—which is naturally far more concerned with applied aspects because these increase profits quickly—the funds the U.S. Government allots to basic research currently amount to about 7 percent of its over-all research and development funds. Unless adequate safeguards are provided, applied research invariably tends to drive out basic. Then, as Dr. Waterman has pointed out, "Developments will inevitably be undertaken prematurely, career incentives will gravitate strongly toward applied science, and the opportunities for making major scientific discoveries will be lost. Unfortunately, pressures to emphasize new developments, without corresponding emphasis upon pure science . . . tend to degrade the quality of the nation's technology in the long run, rather than to improve it."

104. The title below that best expresses the ideas of this passage is
   1. Foundation Funds
   2. Roentgen's Ignorance of X-rays
   3. The Attractiveness of Applied Research
   4. The Importance of Basic Research
   5. Basic Research vs. Applied Research

105. Industry is primarily interested in applied research because it
   1. provides better understanding
   2. is frowned upon by the Foundation
   3. offers immediate profit
   4. drives out basic research
   5. solves practical problems

106. Basic research is vital because
   1. it leads to results that can be appreciated
   2. it is driven out by applied research
   3. it provides the basis for scientific progress
   4. its results cannot be anticipated
   5. it tends to degrade the nation's technology

107. The federal government
   1. encourages basic research
   2. devotes more than 90% of its research and development funds to applied research
   3. spends far more on applied research than on military problems
   4. opposes the Foundation's grants to basic research
   5. is not concerned about the nature of the research being done by the scientists it employs.

108. Less developed countries
   1. spend little on research
   2. devote a large portion of their budget to applied research
   3. realize that progress depends on basic research
   4. encourage their career scientists to experiment
   5. devote less than 7% of their scientific budget to basic research

The *range in frequencies* of musical sounds is approximately 20—20,000 cycles per second (cy/sec). Some people can hear higher frequencies than others. Longitudinal waves whose frequencies are higher than those within the audible range are called *ultrasonic* frequencies. Ultrasonic frequencies are used in sonar for such purposes as submarine detection and depth finding. Ultrasonic frequencies are also being tried for sterilizing food since these frequencies kill some bacteria. Sound waves of all frequencies in the audible range travel at the same speed in the same medium. In the audible range, the higher the frequency of the sound the higher is the *pitch*. The term *supersonic* refers to speed greater than sound. An airplane traveling at supersonic speed is moving at a speed greater than the speed of sound in air at that temperature. *Mach 1* means a speed equal to that of sound; *Mach 2* means a speed equal to twice that of sound, etc.

Musical sounds have three basic *characteristics:* pitch, loudness, and quality or timbre. As was indicated above, *pitch* is determined largely by the frequency of the wave reaching the ear. The higher the frequency the higher is the pitch. *Loudness* depends on the amplitude of the wave reaching the ear. For a given frequency, the greater the amplitude of the wave the louder the sound. To discuss quality of sound we need to clarify the concept of overtones. Sounds are produced by vibrating objects; if these objects are given a gentle push, they usually vibrate at one definite frequency producing a pure tone. This is the way a tuning fork is usually used. When objects vibrate freely after a force is momentarily applied, they are said to produce their *natural frequency*. Some objects, like strings and air columns, can vibrate naturally at more than one frequency at a time. The lowest frequency which an object can produce when vibrating freely is known as the object's *fundamental frequency;* other frequencies that the object can produce are known as its *overtones*. The *quality* of a sound depends on the number and relative amplitude of the overtones present in the wave reaching the ear.

143. A soprano would probably have a frequency of
   1. 200 cy/sec
   2. 500 cy/sec
   3. 5000 cy/sec
   4. 10,000 cy/sec
   5. 20,000 cy/sec

144. The timbre of a musical sound is dependent on its
   1. fundamental
   2. amplitude
   3. frequency
   4. overtones
   5. speed

145. Which of the following individuals would likely use terms like Mach 5 or Mach 9?
   1. Jet pilot
   2. Musician
   3. Astronaut
   4. Submarine navigator
   5. Biologist

146. Ultrasonic frequencies are
   1. inaudible
   2. excessively fast
   3. characterized by a great amplitude
   4. death rays
   5. overtones

VITA

Samuel S. Lee was born on September 28, 1964 in Seoul, Korea. He received his Bachelor of Science degree in biochemistry from Virginia Polytechnic Institute & State University in 1987.

He worked as a graduate research assistant from 1989 to 1991 in the Industrial Ergonomics Laboratory of the Virginia Tech Human Factors Center. He also worked as an intern engineer at the Systems Engineering Research Institute of the Korea Institute of Science and Technology during the summer of 1990 in Seoul, Korea. He is an affiliate of both the American Society of Safety Engineers (ASSE) and the Human Factors Society (HFS) and a member of the Korean-American Scientists and Engineers Association.