



3 1293 01421 6406

This is to certify that the

dissertation entitled


Usability Feedback in Education Software Prototypes:
A Contrast of Users and Experts

presented by

Pericles Varella Gomes

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Philosophy


Major professor

Dr. Patrick Dickson

Date April 12, 1996

LIBRARY
Michigan State
University

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.

[illegible]

USABILITY FEEDBACK IN EDUCATION SOFTWARE PROTOTYPES:
A CONTRAST OF USERS AND EXPERTS

By

Pericles Varella Gomes

A DISSERTATION

Submitted to
Michigan State University in
partial satisfaction of the requirements
for the degree of

DOCTOR IN PHILOSOPHY

Department of Counseling, Educational Psychology,
and Special Education

1996

ABSTRACT
USABILITY FEEDBACK IN EDUCATION SOFTWARE
PROTOTYPES:
A CONTRAST OF USERS AND EXPERTS

By
Pericles Varella Gomes

This study compares usability feedback from users and hypermedia designers when evaluating computer-based instruction prototypes. It provides information for defining cost-effective evaluation strategies and methods, and for specifying valid instruments and tools. Usability instruments such as QUIS 5.5b - University of Maryland are combined with think-aloud evaluation techniques to collect feedback from 16 target users (engineering students from the U.S. , China, Korea, India and Pakistan) and 5 educational hypermedia designers. The designers also evaluated data collected from users, which included quantitative, qualitative reports and multimedia files. In the quantitative side, descriptive statistics, non-parametric and cluster techniques were applied to the answers. User groups and designers were compared, as well as some more general trends, when all subjects were combined. Gender comparisons were also studied. Critical incidents multimedia files were produced for each subject, with screen and audio grabs of problems encountered; navigational maps were generated for each subject; written comments about the prototype were collected; and a descriptive list of errors was generated, comparing types of errors encountered.

Designers reported that qualitative instruments in general were more useful. Designers were more critical both about the interface aspects and pedagogical dimensions and significantly found more errors. American users were more efficient in finding errors. In terms of rating the software, Indian users were more forgiving; and the American group was the most critical. Females were systematically more positive about the prototype. Designers were more efficient than users when executing the usability evaluation, but could not completely replace users (some errors were found only by users). Designers were better in the double task of trying to learn and critique a new interface and learn about the content at the same time. The variability of feedback within users and within designers was found high. Methodological considerations for further work include the relative usefulness of combining quantitative and qualitative methods; the issue of when to use designers as opposed to target users, and the importance of gathering information from different ethnic user groups when developing software for an international audience. One major conclusion regarding the ratings of instruments by experts was that the best instruments were the ones that produced contextualized data, both in the quantitative and qualitative aspects (such as the multimedia files, list of problems, and demographic data).

Acknowledgments

This research was highly dependent on the volunteer help and involvement of many people. Thanks go to the students and the hypermedia designers at Michigan State University, who were willing to spend time for my research. I also thank Mr. Alaciel Franklin de Almeida and his colleagues at TELEBRAS, in Brasilia, Brazil, for providing their software and their support and for giving me the opportunity to collaborate with them in this study.

I thank Dr. Patrick Dickson and Dr. Carrie Heeter, members of my committee, for allowing and helping me to choose such a rewarding topic of research and for their help and support throughout this project. My parents in Brazil, and my family in Michigan all deserve thanks for their support, encouragement, and prayers. I especially thank my wife Luciana for her support and patience throughout this research.

I would also like to thank Dr. Leighton Price and Dr. Cindy Nichols for their help and suggestions on the quantitative side of this study. Cindy, I am looking forward to play Mozart with you again!

Table of Contents

List of Tables.....	viii
List of Figures.....	ix
 Chapter 1 Introduction.....	 1
1.1 Prototypes.....	3
1.2 Instruments	4
1.3 Research Questions.....	6
 Chapter 2 Related Research	 8
2.1 Prototyping.....	8
2.2 Usability Questionnaires	10
2.3 Critical Incident	11
2.4 Classification of Usability Factors	13
2.5 User Interface Evaluation by Experts	15
2.6 Number of Subjects	18
2.7 Gender and Ethnic Differences	20
 Chapter 3 Evaluation Setting	 22
3.1 The Prototype Tested	23
3.2 Instructional Objectives	22
3.3 The Interface.....	27
3.4 Description of the Prototype: Sequential Structure	29
3.5 The Computer Environment	31
3.6 Description of the Physical Space	31
 Chapter 4 Description of the Evaluation	 33
4.1 Evaluation Overview	33
4.2 Subjects.....	35
4.2.1 Users.....	35
4.2.2 Experts.....	36
4.3 Procedures	37
4.3.1 Orientation.....	38
4.3.2 The Evaluation Session	39
4.3.3 Data Recording	40
4.3.4 Multimedia Files	41
4.3.5 Questionnaires	42
4.3.6 Meta Analysis.....	43
4.4 Evaluation Chronology	47

Chapter 5 Analysis	48
5.1 Choices for Analysis	48
5.1.1 Different Approaches on the Same Problem	48
5.1.2 Additional Considerations	49
5.2 Statistical Choices	50
5.2.1 Use of Non-parametric Methods	50
5.2.2 Use of Cluster Analysis	51
5.3 Results.....	52
5.3.1 QUIS Comparison of Users & Experts	52
5.3.2 QUIS Comparison Between Gender	58
5.3.3 Reeves & Harmon: Comparison of Users & Experts	63
5.3.4 Reeves & Harmon: Comparisons Between Gender ..	67
5.3.5 List of Problems	68
5.3.6 Results of Cluster Analysis	72
5.3.7 Results of the Meta Evaluation by Experts	74
5.4 Qualitative Analysis - Multimedia Files	75
5.5 ChapterSummary.....	76
 Chapter 6 Discussion	 78
6.1 Cultural Identity of Participants and Observers	78
6.2 Differences Among Ethnic Groups, and Experts	79
6.3 Multimedia Files.....	80
6.4 Analysis of Content, Pedagogy & Interface	81
6.5 Use of Questionnaires in Interface Evaluations	82
6.6 Problems Verbalized Versus Errors Observed	84
6.7 Qualitative and Quantitative Instruments	85
6.8 Statistical Tools in the Evaluation Methodology	86
6.9 Number and Nature of Problems Encountered	87
 Chapter 7 Conclusions	 89
7.1 Differences in Usability: Users and Experts	89
7.2 Differences in Usability: Ethnic Groups	90
7.3 Differences in Usability: Gender	91
7.4 Value of Qualitative Tools	91
7.5 Evaluation of Methodology	93
7.5.1 Videotaping.....	93
7.5.2 The Interaction of Observer and Subjects	93
7.5.3 Number of Subjects	94
7.5.4 Questionnaires.....	94
7.5.5 The List of Problems	95

7.5.6 Navigational Maps.....	95
7.5.7 Use of Statistical Tools	96
7.6 Recommendations.....	96
7.7 Future Research	98
7.7.1 Enhance the Methodology	98
7.7.2 Qualitative Emphasis	99
7.7.3 Quantitative Emphasis	99
7.7.4 The Inclusion of Personality	100
7.7.5 Comparison of different Kinds of Observers	100
7.7.6 Use of Navigational Maps	100
7.7.7 Use of Questionnaires.....	101
 Bibliography	 102
 Appendix A: Consent form	 109
 Appendix B: Preliminary Questionnaires	 110
 Appendix C: Description of Evaluation	 112
 Appendix D: Pre-Requisites	 113
 Appendix E: Questionnaire Reeves and Harmon	 114
 Appendix F: Questionnaire QUIS	 116
 Appendix G: Questionnaire for Meta-Evaluation	 118
 Appendix H: QUIS Comments	 119
 Appendix I: Navigational Maps	 131
 Appendix J: List of Problems	 152
 Appendix K: Report of Usage	 154
 Appendix L: Variables included in Minitab	 155

LIST OF TABLES

Table 4.1	Ethnic Groups and their components	35
Table 4.2	Description of Experts: Qualifications and Jobs.....	37
Table 5.1	QUIS Comparisons between Ethnic Groups & Experts ...	53
Table 5.8	QUIS answers-Gender Comparison	59
Table 5.15	Reeves & Harmon: Comparisons of Users and Experts...	64
Table 5.19	Reeves & Harmon: Gender Comparison.....	67
Table 5.21	List of Problems (in order of coding by researcher)....	70
Table 5.23	Mean number of problems found by user and experts- Categorization of types of problems	71
Table 5.26	Cluster Analysis by Participants and Groups	73

LIST OF FIGURES

Fig. 3.1-Instructional Objectives of Prototype	23
Fig. 3.2-Typical Screen of First Part of Prototype	24
Fig. 3.3-Typical Screen of Second Part of Prototype	25
Fig. 3.4-Example of an Exercise	26
Fig. 3.5-Example of a Summary Screen	26
Fig. 3.6-Look and Feel of the Interface	27
Fig. 3.7-Overview of the Prototype	30
Fig. 3.8-Layout of the Observation Room	32
Fig. 4.1-Instruments and Procedures used in the study	34
Fig. 5.2-QUIS-Comparison of Users & Experts	54
Fig. 5.3-QUIS-Overall items-Ethnic Group Users & Experts.....	54
Fig. 5.4-QUIS-Screen items-Ethnic Group Users & Experts	55
Fig. 5.5-QUIS-Terminology items-Ethnic Group Users & Experts ..	56
Fig. 5.6-QUIS-Learning items-Ethnic Group Users & Experts	57
Fig. 5.7-QUIS-System items-Ethnic Group Users & Experts	57
Fig. 5.9-QUIS-Gender Comparison.....	60
Fig. 5.10-QUIS-Gender-Overall Aspects.....	60
Fig. 5.11-QUIS-Gender-Screen Aspects.....	61
Fig. 5.12-QUIS-Gender-Terminology Aspects.....	62
Fig. 5.13-QUIS-Gender-Learning Aspects.....	62
Fig. 5.14-QUIS-Gender-System Aspects.....	63
Fig. 5.16-Reeves&Harmon:Comparison of Users & Experts.....	64
Fig. 5.17-Reeves&Harmon:Users & Experts-Learning Dimensions	66
Fig. 5.18-Reeves&Harmon:Users & Experts-Interface Dimensions	66
Fig. 5.20-Reeves&Harmon:Gender Comparison.....	68
Fig. 5.22-Mean number of problems found by subject groups.....	69
Fig. 5.24-Mean number of problems found by user and experts - Categorization of types of problems.....	71
Fig. 5.25-Clustering of Subjects by Hierarchical Tree.....	72
Fig. 5.27-Instruments Ratings by Experts	74
Fig. 6.1-Ratings of Instruments by Experts-The context factor.....	86

Chapter 1

Introduction

What is wrong with interfaces? One problem with design is that it tends to be done by people who have off-the-top-of-their-heads ideas and beliefs about imaginary beasts they call “the users”.

Donald Norman

Evaluation may occur at many points in the development of a software application. Within the instructional development context, different kinds of evaluations are available, depending on the aspect to be focused, such as the effectiveness of the program, its impact, use of resources, and ways for improvement. Each of these aspects entails a different facet of the evaluation process.

The item “How to improve a program” involves a facet of evaluation called formative evaluation. The overall purpose of formative evaluation is to provide information to guide decisions about enhancing and interactive multimedia program at various stages of its development.

This dissertation focuses on usability evaluation during the early stages of interface design of instructional software. The importance of practical human-computer interface testing in educational multimedia is clear, yet many of the available models appear to be inappropriate or non useful to designers, particularly when dealing with early prototyping. Part of the problem is due to

the fact that requirements and subsequent specifications evolve throughout the development period [Briggs and Briggs, 1990].

Another reason why interface usability testing is not widely adopted is the lack of understanding of its importance or meaning from project managers. One of the critical factors in improving the acceptance of usability testing is to provide concrete data that would convince managers that usability evaluation is worth executing [Nielsen, 1993].

Usability testing can be costly and time-consuming if sophisticated experimental methods are used such as those prescribed in the “usability engineering” approach [Whiteside, Bennett and Holtzblatt 1988]. Such sophisticated methods require the skills of a human factors specialist and access to usability laboratories. They provide a large quantity of high-quality data but many designers see them as intimidating in their complexity [Bellotti, 1988]. They may also distance the designer from the user rather than bring the user and designer close together. Nielsen has argued that it is possible to get a reasonable level of feedback without using such costly methods. His approach of “discount usability engineering” aims to strike a balance between the quality of feedback obtained and the cost of obtaining it [Nielsen, 1989].

Developing interactive multimedia is a creative, demanding, multifaceted task. One of the most important components of any interactive multimedia project is the interface, but there are no established rules available, perhaps because it is an art that is not

easily learned or described [Laurel 1990]. One of the few rules that appears to be accepted is the one that tells the designer about knowing the target user [Shneiderman 1987]. Yet interfaces are typically created by professionals who have far more contact with computers than the intended target users [Nielsen 1993]. This dissertation seeks to provide educational multimedia designers with the information needed to improve interface evaluations in early stages of design.

1.1 Prototypes

A prototype, by definition, is a working model of the conceptual design. Multimedia designers use prototyping techniques to try out ideas about interface design, among other things. Working within constraints of time and budget, prototyping involves the production of early working versions of future application system and experimenting with them.

Early Prototyping provides a communication basis for discussions among all the groups involved in the development process, especially between users and designers [Benimoff and Whitten, 1989; Diaper, 1990]. It also provides an approach to software development based on experiment and experience.

The adoption of prototyping has grown out of the realization that 1) Frequently requirements do not become apparent until a system is in use; 2) Specifications cannot be completed until the construction process begins; and 3) Developers need to understand the cognitive processes of target users in early stages of design.

1.2 Instruments

There are a variety of methods that can be used to collect data to determine the usability of a system. The most accepted ones are verbal reports from users (think aloud), objective measures of users' performance (either by observation or logging), users' responses to questionnaires, expert reviews, and critical incident techniques [Miller and Jeffries, 1992].

Verbal reports from subjects, or think aloud techniques, consist of asking the participant to reflect audibly on what he or she is doing or wants to do while using the software. Occasionally, the evaluator may intervene to ask for clarification of user comments or to provide help if the program is an early prototype. Most often, a videotape of the monologue is accompanied by unstructured observation of user activity [Roske-Hofstrand, 1989]. The purpose of the think aloud technique is to obtain real time information from users about their processing of the program while they are using it.

Objective measures of users' performance can be done by observation or by automatic computer automatic logging. This technique has the main purpose of collecting measurements in order to compare then with similar measures from users of a different system, or to evaluate them with respect to usability goals. This kind of test usually involves an specific task to be executed by the user.

Questionnaires are most often done in such a way that users either choose from a list of multiple choice answers or mark a

number indicating the strength of their agreement of disagreement with a statement. Questionnaires can be used repeatedly in different usability tests, thus allowing for cross-product comparisons. Respondents usually answer questions after completing the program, but occasionally questionnaires are presented during the use of the application [Flagg, 1990].

Expert Reviews, often called “face validity” consists of showing the prototype to a group of interface designer specialists and asking them to evaluate the interface for usability. The designers should then conduct a detailed analysis of the interface. Often one component of this kind of evaluation involves proceeding step-by-step through task scenarios. Different types of experts can provide different perspectives on the critical aspects of the program [Reeves, 1993].

Critical Incident techniques consist of collecting information on interface problems when they occur and uses open-ended questions to obtain information on missing or non functional features of the software interface. This technique gives the user the chance to react to software by explaining the problem at the time of its occurrence. It also allows designers to collect information on satisfactory features of the interface. Critical incidences are occasions when the system is particularly poor or surprisingly good, and knowing the detailed circumstances of such incidents can often help to avoid worst-case incidents in the final product [Nielsen 1993].

1.3 Research Questions

This dissertation attempts to create a model for usability testing in educational software prototypes, by evaluating an existing prototype of computer-based instruction and applying existing usability testing tools combined with additional complementary methods of data collection. This study includes both target users and educational hypermedia designers specialists. Five main research questions are defined below:

#1) What are the differences in usability feedback between users and hypermedia designers?

This is the main question to be answered in this study. Is it enough to have designers' feedback in order to measure and improve the usability of interfaces? Is user feedback, instead, sufficient to verify the usability of interfaces? What are the differences and similarities between users and designers when evaluating an interface?

#2) What instruments do hypermedia designers value the most when evaluating both quantitative and qualitative data from the users feedback?

If usability evaluation is performed to help hypermedia designers improve the interface of a prototype, which feedback categories are most valuable to facilitate this process? This question should help designers and managers when faced with educational software evaluation planning and execution. Do designers place greater attention on quantitative data (like

questionnaires and demographic information) or qualitative oriented data (like multimedia files and users written comments)?

#3) What are the differences in usability feedback across different ethnic user groups?

This question relates to the issue of cultural and ethnic differences among target users. The prototype in question was developed for an international audience of engineering students and professionals. Do different ethnic groups present significant differences in terms of usability evaluation feedback? 3 main ethnic groups are compared in this study: Chinese/Koreans, Indian/Pakistanis and North Americans engineering students.

#4) What are the differences in usability feedback between males and females?

This question addresses gender differences among target users. Are there significant differences in responses, errors detected and attitudes when gender is taken in consideration? This question could generate exploratory answers to this important aspect of software evaluation.

Chapter 2

Related Research

Over the past 10 years, research in usability evaluation of computer interfaces has been carried out in three main areas: human factors, computer science and cognitive psychology. This research has produced understanding of how to evaluate interfaces in several different contexts. This chapter surveys previous work on usability evaluation that focus mainly on early prototypes.

2.1 Prototyping

This section surveys past research that focused on the utilization of prototyping in software design.

One primary reason for prototyping the user interface is to use the user interface prototype to collect feedback from prospective users (Benimoff and Whitten, 1989; Diaper, 1990). A user interface prototype can be demonstrated to user to elicit their feedback about the functionality of the system and about the interface design. User interface prototypes can be created so that end users can actually use the prototype as they would the final system. Data on the usability of the design (time to complete a task, number and type of errors made, and so on) can be collected before the actual system has been built. Prototypes that are incomplete or that don't match the final specifications can still be used for the collection of user feedback and expert evaluation.

Another reason for prototyping the user interface is that it gives the designer an opportunity to try out various alternative designs. Competing designs can be prototyped and then either tested with prospective users or evaluated by experts (Benimoff and Whitten, 1989). User interface prototyping helps to ensure consistency in user interface design. When the user interface for a new computer system works in a way that is already familiar to users from their experiences with other computer systems, user find much easier to learn to use the new system (Polson, 1988). Through careful evaluation of prototypes, designers are able to catch inconsistencies before they become a part of the system code.

Prototyping reduces cycle time and reduces project costs. Tavolato and Vincena (1984) report findings that 76% of development effort is directed toward late-stage activities such as correcting errors that exist in code and adapting the software to meet new requirements. In addition, they found that about half of the errors that are discovered late in development can be traced to failures in the requirement phase. User's requirements are better understood and are communicated to the software developers more efficiently through rapid prototyping. This, in turn, reduces the number of errors in the code and the number of new requirements introduced later in the product cycle. Thus, the relatively small cost of investing in prototyping in the beginning of the development process can result in large savings at the end of the development process.

Prototyping encourages iteration, expansion of ideas, and risk analysis that are characteristic of new software development models (Boehm, 1988). In a study conducted by Boehm (1984), it was determined that the use of a prototyping approach resulted in 45% less development time than an approach that relied on specifying the design only through requirements and specification documents. These data point to significant cycle-time improvements when prototyping is used.

2.2 Usability Questionnaires

This section surveys research that focused on the utilization of questionnaires for usability issues and research done with QUIS - Questionnaire for User Interface Satisfaction (developed by Chin, Diehl, and Norman, 1987, 1988).

Specific questionnaires for evaluating computer systems and interface designs have been developed. LaLomia and Sidowski (1990) reviewed some of the questionnaires under two general classes: user satisfaction with computer systems, and computer literacy and aptitude. They report five questionnaires which have been developed to address user satisfaction. Each questionnaire addresses slightly different aspects of usability and different kinds of equipment. Because of these different aspects, it is difficult to compare directly different questionnaires.

The most frequently used questionnaire for usability testing evaluation is QUIS - Questionnaire for User Interface Satisfaction - [Chin, Diehl, and Norman, 1988]. Quis has 27 items using a 9-point

Likert scale. According to LaLomia, the items test overall reactions to software (6 items), evaluation of characters on the screen (4 items), use of terms and information throughout the system (6 items), learning to operate the system (6 items), and system capabilities such as speed (5 items). The reliability of the test was found to be 0.94. Validity was tested by how well the items discriminated between PC systems that were liked and disliked. In all cases, the means were higher for the liked systems than for the disliked systems, thereby providing evidence for the validity of the questionnaire.

2.3 Critical Incident

This section surveys research that focused on the issue of using critical incident for usability purposes.

First used by Fitts and Jones (1947) to analyze pilot error, and more recently by Cooper (1982) to investigate errors made by anesthetists, this technique consists of the study of 'critical incidents' to identify common features or elements in order to classify those incidents. In both studies (Fitts and Jones, 1947; Cooper, 1982), critical incidents are defined as human error or equipment failure that did or could have had unsatisfactory results. Cooper (1982) first categorized incidents by their outcome: favorable, unfavorable, neutral, or other. The incidents in each of the outcome categories were then classified further by their causal relationships. Dzida (1978) feels that critical incident technique is a feasible method to obtain user evaluation of

human-computer interfaces and to translate that evaluation into design requirements.

Galdo, Williges, Williges and Wixon (1987) conducted a study about critical incident evaluation tool for software documentation. In this evaluation, subjects were asked to perform a benchmark task consisting of 19 sub tasks and to use the associated software documentation. Both hard copy and on-line documentation were available. After subjects completed each sub task they were asked to use an on-line questionnaire to report critical incidents encountered in using the hard copy and on-line documentation. The critical incidents were sorted into four categories: on-line documentation failure incidents, on-line documentation success incidents, hard-copy documentation failure incidents and hard-copy documentation success incidents.

The incidents in each failure category were reviewed to identify common documentation features or elements that caused problems. The same process was repeated for incidents categorized as successful to determine satisfactory features of the documentation. The problems were arranged in descending order from most critical to least critical by frequency of critical incidents associated with each problem. Ties in frequency were broken by an average severity index. Average severity was calculated by averaging the incident severity ratings supplied by users at the time each incident was reported.

A list of documentation problems and satisfactory features was presented to the software design team to guide the redesign

process. This evaluation helped validate the critical incident technique as a method for providing software designers with end-user data for revision of software and documentation.

2.4 Classification of Usability Factors

This section surveys research that focused on the issue of classification of usability factors.

Usability Factors have been divided into five main attributes [Nielsen, 1993]:

- **Learnability**: Program should be easy to learn (the user can rapidly start getting some work done).
- **Efficiency**: Program should be efficient to use (once learning is completed, high level of productivity is possible).
- **Memorability**: Program should be easy to remember (user is able to return to system without having be trained again).
- **Errors**: Program should have a low error rate (users make few errors when using the system, or easy to undo errors).
- **Satisfaction**: Program should be pleasant to use (users are satisfied when using it).

Reeves and Harmon (1994) describe two complementary multi-dimensional approaches to evaluating interactive multimedia programs for education and training. The first approach is based upon a set of fourteen pedagogical dimensions such as “experiential value” and “learner control”.

The second approach is based upon a set of ten user interface dimensions such as “easy to use” and “screen design”. They have applied the pedagogical and user-interface dimensions to the evaluation of two interactive multimedia programs, the Jasper Woodbury Problem Solving Series developed by the Cognition and Technology Group at Vanderbilt University, and the Columbus Encounter: Discover and Beyond “Ultimedia” program developed by the IBM Corporation.

Their recommendations, in light of their admittedly preliminary investigations into the value of these dimensions, is to subject the dimensions to rigorous expert review by leaders in the design and application of interactive multimedia in both education and training. They also suggest that, since there is evidence of the qualitative validity of the dimensions, quantitative scales should be integrated into each dimension, e.g., a ten point rating system. They hesitated to add this quantitative aspect to the dimensions to avoid fear that reviewers might get too distracted by the numerical values to concentrate on qualitative ratings of the dimensions themselves. They also recommended that the validated dimensions be applied within a wide variety of education and training contexts to provide evidence for their utility. A last recommendation was that research should be initiated into the relationships among ratings of the pedagogical and user interface dimensions of applications and actual data regarding the instructional effectiveness and impact of these programs.

2.5 Evaluation by Experts

This section surveys research that focused on issues related to usability evaluations by expert reviews.

Usability evaluation methods by experts have been the focus of research during the past 5 years. The objective of this kind of usability evaluation is to contribute to the design of usable software for end users. These evaluations provide a way of quickly inspect and find problems in software prototypes without having to include target users, at least in the early stages of development. There are differences in how expert inspections are conducted depending on the characteristics of the experts, and on the objective of the evaluation itself.

Pollier (1992) studied the activities of human factors specialists charged with evaluating a human-computer interface. Subjects were 4 experienced ergonomists specializing in information systems. Subjects were asked to think aloud and to consult with the experimenter while evaluating the human-computer interface in a multimedia communication system. The resulting verbalizations, videotapes of subjects activities, and subjects written and graphic productions were analyzed to determine the number and type of ergonomic issues taken into consideration and the strategies used in performing the evaluation. Individual differences in these variables were analyzed.

Reeves and Harmon (1993) reported on the application of their user interface and pedagogical dimensions for evaluation purposes by experienced developers. They have conducted preliminary analyses with faculty and graduate students in an instructional technology graduate program. They suggest further research of these dimensions involving experienced personnel in other education and training contexts.

Usability inspection methods, based on informed intuitions about interface design quality, hold promise of providing faster, more cost-effective ways to generate usability evaluations, compared to empirical user evaluation methods. Examples of inspection methods include heuristic evaluation [Nielsen and Molich, 1990], usability walkthroughs [Bias, 1991; Karat and Bennett, 1991a, 1991b], cognitive walkthroughs [Lewis, Polson, Wharton and Reiman, 1990], and applications of guidelines in walkthroughs [Jeffries, Miller, Wharton, and Uyeda, 1991]. These methods have been used in development for some time in one form or another.

Desurvire and Bradford described the use of multiple methods in development projects, to assess real-world applicability, to compare effectiveness of methods, and also to explore how different methods might complement each other [Desurvire, Kondziela and Atwood, 1992]. Desurvire study illustrated that inspections can be used across a wide variety of software interfaces.

In relation to methods of making more and better quality predictions, and streamlining usability evaluations, Monk noted that sensitivity to potential problems probably is driven largely by experts' own problems, the observation of others having similar problems, and expert's skill at reflecting on and generalizing these personal experiences. Nielsen noted the similarity between usability inspection and having problems as a participant in an empirical user test. Research might be directed at finding ways to help experts acquire this experience [Wright and Monk, 1991] and generalizing it for purposes of making predictions and judgments.

Another area of research to be explored is the improvement of the way data from expert evaluations is analyzed, and how results may be more effectively used by the larger development cycle [Mack and Nielsen, 1993]. Design and evaluation should be tightly linked, and this relationship needs to be understood and supported. It would be worthwhile to explore the possibility of developing on-line tools for cumulating and organizing information based on inspection data, and for applying it to new design problems.

Research has confirmed that expert evaluation is an efficient usability inspection method [Jeffries 1991]. However, expert evaluation methods were developed to be used in circumstances where user testing is impractical, and if it is used to the exclusion of user testing, it could mean the loss of one of the most valuable tools for interface evaluation [Jeffries and Desurvire 1992].

2.6 Number of Subjects

This section surveys research that focused on the issue of number of subjects for usability evaluation of interfaces.

Results from several studies indicated that any single expert evaluator would miss most of the usability problems of an interface. Several studies [Molich and Nielsen 1990; Nielsen and Molich 1990; Nielsen 1992; Nielsen 1994] indicated that single evaluators found on average only 35% of usability problems. The results also indicated that since different evaluators tend to find different problems, it is possible to achieve better performance by aggregating the evaluations from several experts. The exact number of evaluators to include should depend on context-specific cost-benefit analysis.

Nielsen [1990] reported an experiment that was designed to measure the percentage of usability problems computer scientists would find using the thinkaloud technique. In this study, 20 groups of minimally trained experimenters independently conducted usability tests of a paint program. Their task was to find as many of the usability problems Nielsen had defined a priori as “major usability problems” as they could. Each evaluator ran an average of 2.8 subjects per evaluation. The results showed that computer scientists were able to apply the thinkaloud method effectively to evaluate user interfaces with a minimum of training and that even methodologically primitive experiments could succeed in finding many usability problems. This experiment was replicated [Nielsen, 1992] with similar results.

Virzi [1992] performed a series of three experiments to extend the exploratory work done by Nielsen. In these experiments he examined the rate at which usability problems were identified as a function of the number of users run in a single usability evaluation when the evaluation was conducted by experts. In all of the three studies, approximately 80% of the usability problems identified would have been found after only five subjects. He concluded that important usability problems are more likely to be found with fewer subjects than are less important problems, and that a practitioner who chooses to run a small number of users will identify most of the major usability problems and some proportion of the less important problems. Experts were able to reach consensus regarding the relative severity of problems without benefit of frequency data. He concludes that usability experts can assess the severity of a problem without explicit knowledge of how frequent the error is likely to be.

Jeffries and Desurvire (1992) point out that different methods have strengths and weaknesses and that the best evaluation of a user-interface comes from applying multiple evaluation techniques combined. The various techniques have differing constraints on their applicability and on the resources required to apply them effectively. User testing and expert evaluation require access to expert evaluators and users; in the case of expert evaluation, a group of them is required. They suggest that when access to multiple experts are available, doing

both expert evaluations and usability testing with users is the best strategy.

2.7 Gender Differences

Many researchers in education and psychology have found that gender accounts for differences in both attitudes toward computers and performance [Premkumar, Ramamurthy and King, 1993; Anderson, 1987]. While Igabaria found that gender had a significant effect on attitudes toward computers [Igabaria, 1990], Parasuraman and Igabaria found no dissimilarity in attitudes between males and females [Parasuraman and Igabaria, 1989]. Cronan et al. found that gender was a major factor influencing the performance of students taking an introductory computer information system course [Cronan, Embry and White, 1989].

According to Dambrot et al. [1985] among first year undergraduates, male students were more likely to have attended and completed computer-related courses and to have knowledge of a computer language. Similarly, in their study of first year university science students, Clarke and Chamber [1989] found that men were significantly more likely to report previous computer experience over a range of applications. Discussion at the Gender and Science and Technology Conference [1990] revealed that a declining female participation in computer studies appeared to be an international trend.

Past research also suggests that, relative to traditional teaching, use of Computer Assisted Instruction (CAI) can give rise

to gender inequities in student achievements [Siann et al, 1990; Sutton, 1991].

Cross-cultural studies also shows that there are gender differences in attitudes towards computers. In a study of Canadian and Chinese high school students' attitudes toward computers, Collins and Willians [1987] found that in both cultures boys in general were significantly more positive than girls in their attitudes toward computers and showed higher self-confidence about working with computers. However, Chinese students displayed fewer gender or age differences, the one exception being the opinions of students concerning the competence of women with regard to science and technology. Females from both countries endorse the idea that women have as much ability as men with respect to science and technology, whereas males were significantly more skeptical.

While it seems to be an agreement among most researchers on the presence of significant sex differences in using and learning about computers, there is less agreement on the causes of this gender differentiation [Shashaani, 1992].

In conclusion, there is little research done that focus specifically on interface usability and gender. Currently, there is no evidence that one gender is more effective than the other when conducting evaluation of educational software.

Chapter 3

Evaluation Setting

This chapter describes the prototype tested, its instructional objectives, its interface, the prototype sequencing, the physical space and the computer environment in which the prototype was evaluated. The educational prototype tested was about teletraffic concepts.

3.1 The prototype tested

The prototype tested in this study is a computer based training (CBT) unit in its early stages of development. This prototype was developed in 1993 by TELEBRAS, the Brazilian Telecommunications Company, at their central training facility in Brasilia. It was developed with Assymetrix Multimedia Toolbook by a team of instructional designers, content experts and programmers.

At that time, TELEBRAS was conducting negotiations with the International Telecommunications Union (ITU), the International Teletraffic Congress (ITC) and the Economic European Community (EEC) to develop a series of telecommunication training modules to be used as professional training materials by telephone companies of participating countries of ITU, which includes telecommunications professionals from Asia, the Americas, Africa and Europe.

The courseware to be developed, using this prototype as the interface model, would be based on a combination of pedagogical strategies: tutorials, simulations, problem-solving, hypertext information retrieval structures at appropriate points, the learner control approach, and student performance records with a

multimedia interface (audio and graphics). The course is intended to be self-contained, in the sense that all information necessary for instruction is included in the software program.

3.2 Instructional Objectives

The prototype consists of a lesson about teletraffic routing. The instructional objectives of this lesson, routing and equivalent graphics, are displayed in Figure 3.1.

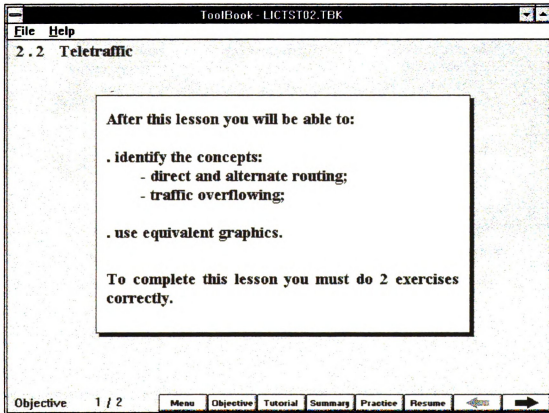


Figure 3.1: Instructional Objectives of the Prototype

The lesson is subdivided into two main pedagogical sections: the first section covers the definition and calculation of alternative routing in telephone traffic. In this section, the student learns

definitions, concepts and how to calculate alternative routing using simulations, examples and graphics. The unit is composed of a total of 13 screens. Figure 3.2 shows a typical screen of this section.

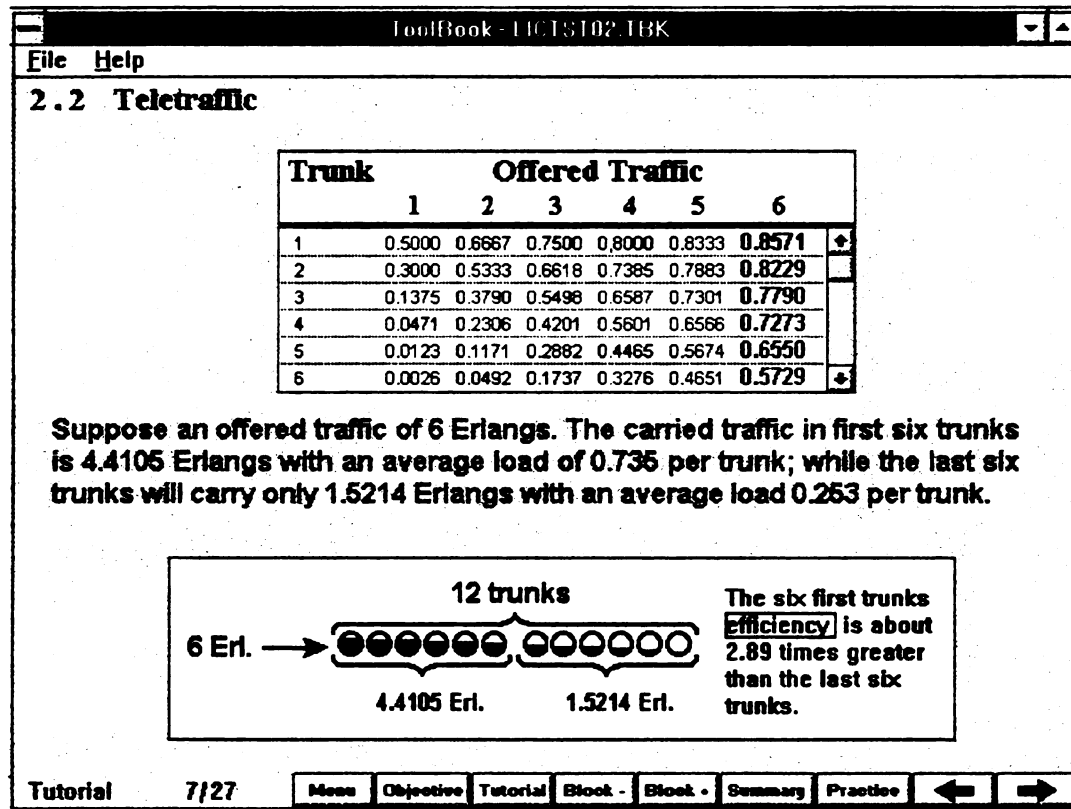


Figure 3.2: Typical Screen of the First Part of the Prototype

In the second section, the student learns how to build equivalent diagrams of alternative routings covered in the first section. In this section, the student is presented with examples and drills, as well as definitions of equivalent diagrams. It is composed of 14 screens. Figure 3.3 represents a typical screen of this section.

The prototype contains a practice section that includes all the exercises from both of the main sections. In this practice section, the student can verify if he or she already knows the material presented in the tutorials. This practice is provided to the student as

a way of quickly testing the material covered in the tutorial without having to run the entire application. Figure 3.4 presents an example of the kinds of exercises included in this prototype.

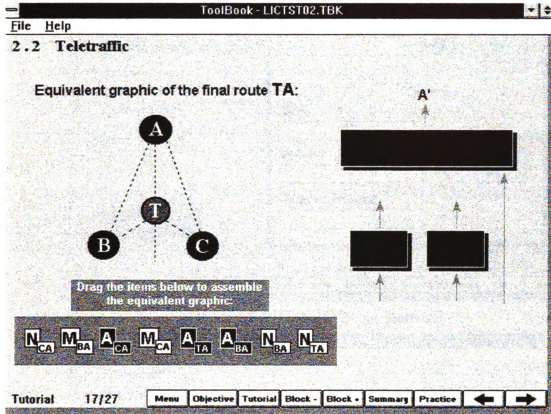


Figure 3.3: Typical Screen of the Second Part of the Prototype

The program also contains a summary section which presents the content of the lesson in a compact form (total of 5 screens). This summary is provided to the student as a way of quickly reviewing the material covered in the tutorial without having to run the entire application. Figure 3.5 exemplifies the screens contained in the summary.

The instructional objectives of this prototype follow strict recommendations made by the International Telecommunication Union, which provided the content expertise to TELEBRAS. In terms

ToolBook - LICTST02.TBK

File Help

2.2 Teletraffic

Calculate the overflow Variance (V) and the Mean (M) in the route of 14 trunks with a poisson traffic of 9.00 Erl.

Variance (V)

Mean (M)

Input

Number of Trunks

Offered Traffic

Output

Mean

Variance ☒

Peakness Factor

Practice 1 / 5 Menu Objective Tutorial Summary Practice Resume

Figure 3.4: Example of an Exercise

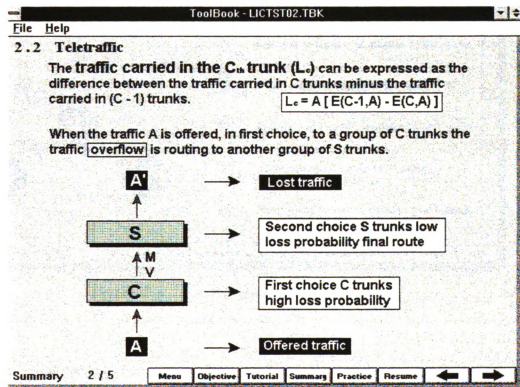


Figure 3.5: Example of a Summary Screen

of the pedagogical format, the lesson prototype was developed using traditional instructional design methodology.

3.3 The Interface

The prototype lesson was developed using Assymetrix Multimedia Toolbook, which runs on Windows 3.1 as its operational system. In terms of its graphic interface style, the prototype presents a look and feel that is shown in Figure 3.6.

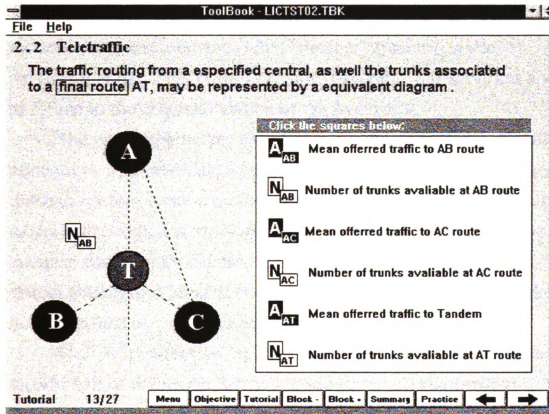


Figure 3.6: Look and Feel of the Prototype

In general, the screens have a navigational menu bar on the bottom, with the following options: Tutorial, Practice, Summary, "Block +", "Block -", left arrow, right arrow. The screens also contain a

top menu bar with “File” and “Help” options. In the lower left corner of every screen there is a location indicator (“tutorial 4/7”, for example). In the upper left corner, the name of the lesson is present on all screens. The central portion of the screen is dedicated to content, and is different every screen.

The left and right arrows are for moving backward and forward in the program, although at times they perform other functions (such as playing audio narration). The block buttons are intended for jumping to the next or previous sections in the lesson. The practice and summary buttons serve as pointers to these sections. At times, a button called “resume” appears on the navigational bar. This button is context-specific, but serves as a way to return to the original location of the hyperlink.

The prototype makes use of multiple windows, such as tables, calculators, and calculation programs. These windows are accessed through buttons located on the central part of the screen, according to the instructional flow. Most of the simulations are accessed by opening additional windows. The lesson makes extensive use of simple animations. Most of the animations represent basic traffic flow by means of color cycling or blinking graphics and letters.

The use of direct manipulation activities is more intense in the second half of the lesson, when students can build equivalent diagrams by clicking and dragging boxes with letters and names. The program utilizes several different input alternatives of this kind of manipulation. There are multiple choice exercises in the prototype, in which the student clicks on boxes to answer the quiz questions. The calculation exercises require the students to scroll tables, make use of the Microsoft calculator, and use paper and pencil.

In terms of audio usage, the use of voice is restricted to one long narration at screen 3 of the tutorial and as audio feedback for exercises, such as "incorrect", "try again" or "correct". The beep of the computer is configured as a piano chord, which is a sound bite that comes with Microsoft Windows 3.1. The use of color throughout the prototype varies. Blue, yellow and red are used depending on the context. Hyperlinks are indicated by a black transparent rectangle around the text to be clicked.

3.4 Description of the Prototype: Sequential Structure

A description of the sequential structure of the prototype is presented here.

The sequence of the program, if visited in a linear fashion, consists of 40 screens, beginning with the objective (2 screens), and followed by the first half of the tutorial, screens 1 through 9. At screen 9, the student is faced with a calculation exercise. If the student successfully answers this exercise, he or she can progress to the second half of the program, screens 13 through 22. If the student fails this exercise, the program shows the procedure and answer of the problem and then presents a new exercise to the student, which is essentially the same problem with different values. If the student fails again, the application sends him or her back to the beginning of the tutorial.

In the second part of the program, the student is presented with the concept of an equivalent diagram. At screen 15 and 16, an equivalent diagram is to be constructed by the student. Here, if the student fails, the computer allows him or her to progress. At the end of this second half, the student is asked to answer a multiple choice

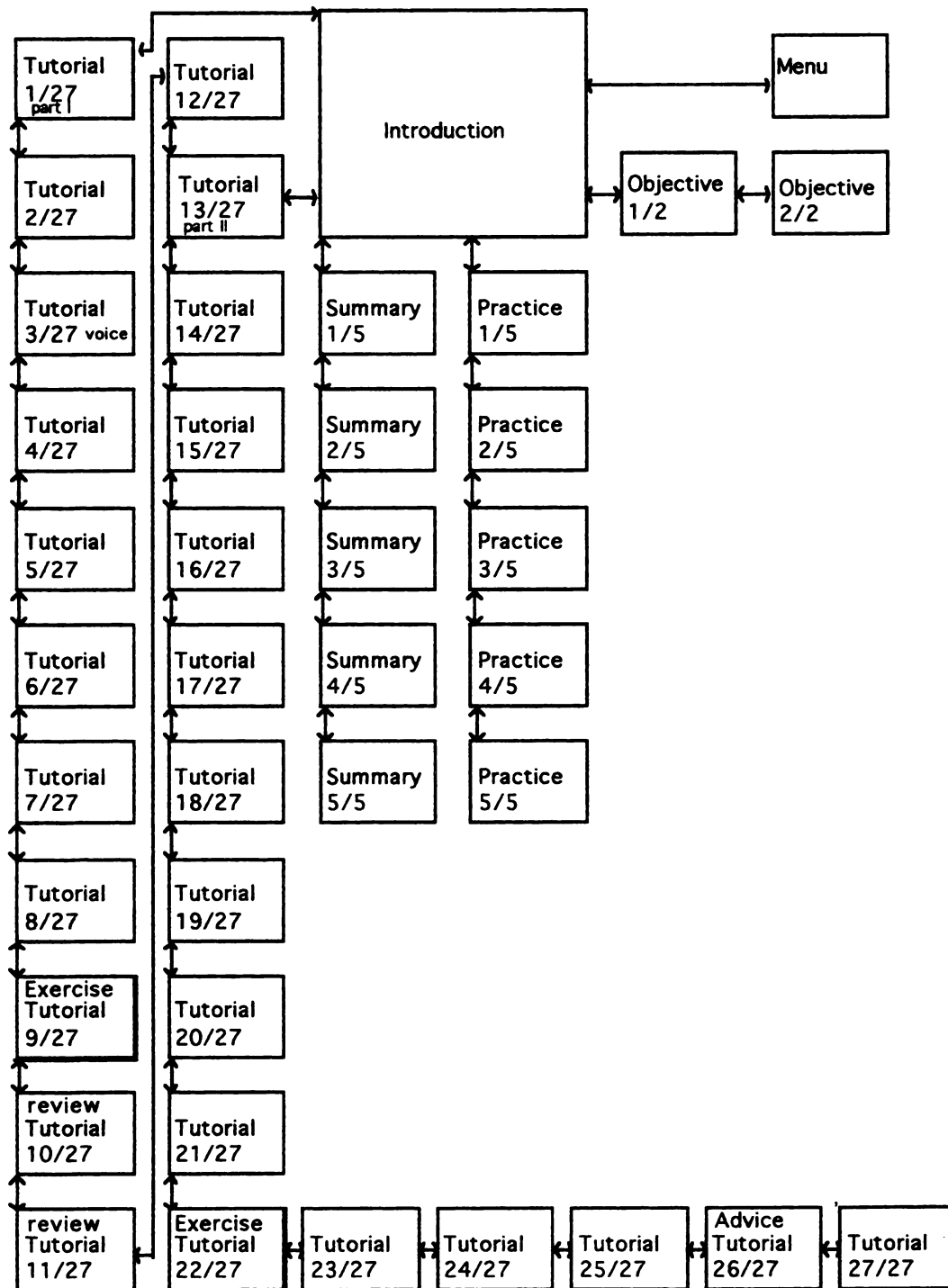


Figure 3.7: Overview of the Prototype

quiz to verify if he or she has learned the content of the prototype. At any time, through the use of the navigational bar, students can move back, forward, jump to the beginning and end of sections, or go to the practice or summary sections.

Figure 3.7 represents the entire prototype, indicating the position of exercises and the main sections of the application. This graphic was created by the researcher and serves as a navigational and visualization instrument for the meta-evaluation.

3.5 The Computer Environment

The computer software and hardware utilized in this study consisted of a laptop PC compatible brand "Pro Star". This laptop was configured with an Intel microprocessor 486 DX4, running at 100 megahertz. It had 12 megabytes of random access memory (RAM) and 800 megabytes of disk space. Although the laptop had a trackball, the researcher preferred to use a 3 button Dexamouse as the means of input for the participants. The screen was a liquid Crystal Display (LCD) of 10.5 inches, passive. This laptop had 8 bit sound capability built-in, with an internal speaker. No extra speakers were necessary.

In terms of software configuration, the computer ran Microsoft Windows 3.1. The prototype, which needed Assymetrix Toolbook to run, could be launched from the Windows Program Manager.

3.6 Description of the Physical Space

The research was conducted in an office of 10 by 15 feet. This office contained a wide window to the exterior, and a door to the

corridor. The office had its furniture layout prepared for the study. A small computer table with chairs was located near the entrance of the room, on which the laptop and auxilliary monitor sat, and where the subjects were seated. A second table (a desk) was available for the observer. The room office lights were kept dim to avoid glare on the computer screens. A tripod with an 8mm camera was set up for video and audio recording of the monitor and subject, via a remote microphone attached to the lapel of the subject. Figure 3.8 represents the layout of the room.

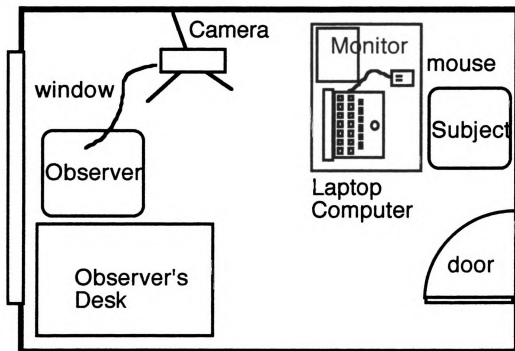


Figure 3.8: Layout of the Room

Chapter 4

Description of the Evaluation

This usability study was composed of two parts: 1) the evaluation of the prototype described in the previous chapter, which was tested both with users and experts, and 2) a meta-evaluation of the instruments and results of the first part, made by experts. For this meta analysis, data collected from users was presented to the experts, who would then rate the different instruments and tools.

This chapter describes the evaluation design, the subjects who participated in the study (target users and educational multimedia experts), the instruments and procedures of data collection and the summarization process.

4.1 Evaluation Overview

This evaluation made use of a combination of qualitative and quantitative data gathering techniques. Videotaping and "think-aloud" techniques were used during the sessions. Two questionnaires were utilized to collect the quantitative data: QUIS (Questionnaire for User Interface Satisfaction) version 5.5, developed at the University of Maryland's Human Computer Interface Laboratory, and a new questionnaire developed specifically for this evaluation, which was based on Reeves' and Harmon's guidelines [Reeves and Harmon,1994] for evaluating interactive multimedia for education and training. Figure 4.1 displays the instruments and procedures utilized in the first part of the study.

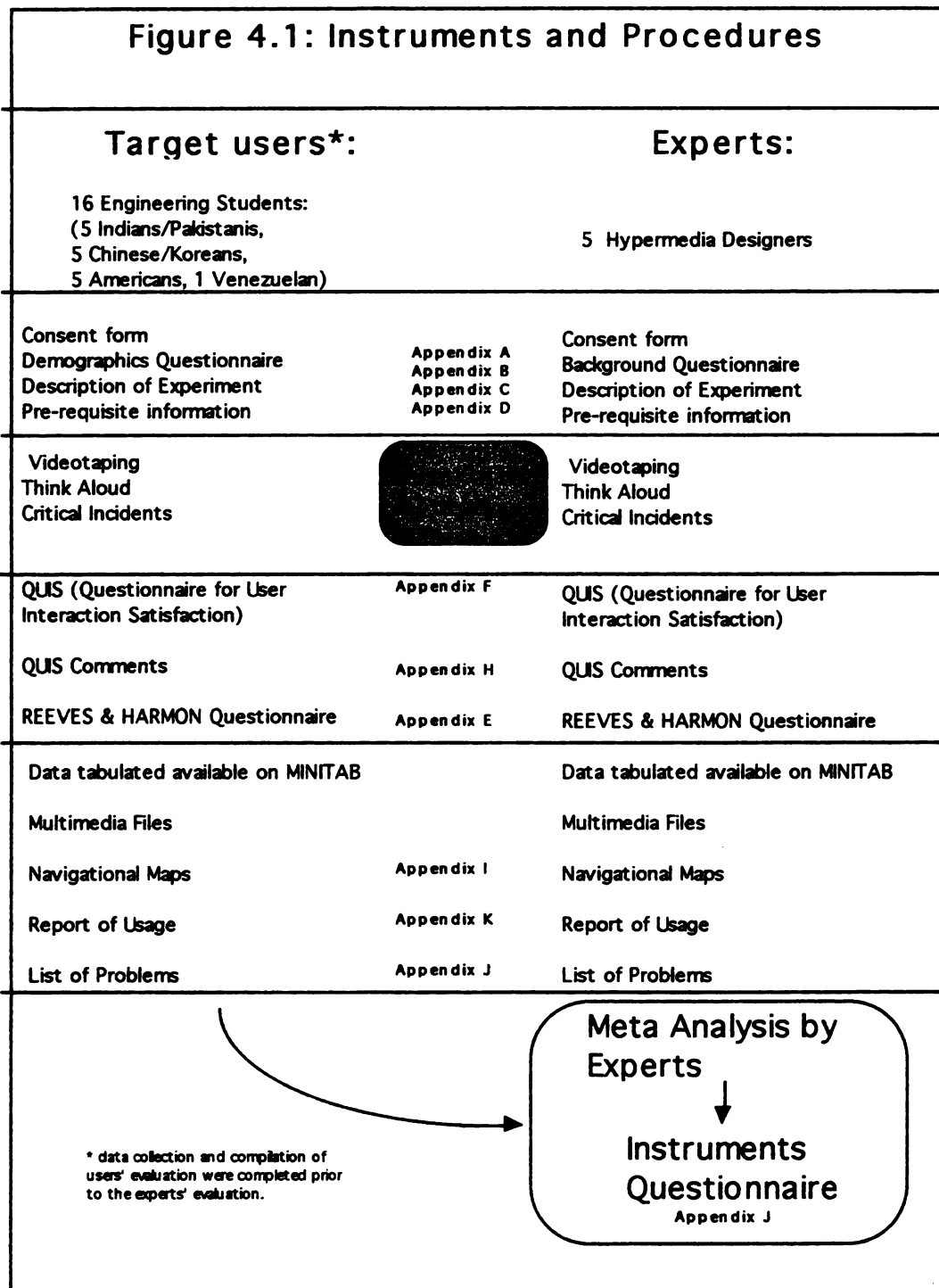


Figure 4.1: Instruments and Procedures used in the study.

In the second part of the study, the meta-analysis, which took place immediately after the experts finished evaluating the prototype, the instruments, procedures and data collected during the users' evaluations were examined and rated by the experts.

4.2 Subjects

The subjects included in this fit into two groups: 16 engineering students, were defined as potential users for the prototype being tested; and 5 educational multimedia designers were defined as experts. They are described in more detail below.

4.2.1 Users

Table 4.1 lists the potential user categories. Potential users were chosen to form three main groups with similar cultural background. The Latin group was excluded for lack of enough subjects available. All users included in this study were current students recruited from the School of Engineering at Michigan State University. There was a mix of undergraduate and graduate students.

Ethnic Group:	Components:
1) Indian, Pakistanis	3 Indians, 2 Pakistanis
2) Chinese, Koreans	3 Chinese, 2 Koreans
3) Americans	5 Americans
4) Latin Americans	1 Venezuelan

Table 4.1: Ethnic Groups and their components

The students were recruited by electronic mail and by the use of flyers posted around the Engineering Building. A compensation of

ten dollars was offered in exchange for their participation in the study.

The students recruited for the evaluation had the following characteristics: ages ranging from 19 to 30 years, with a mean of 21 years; 87.5% of the users were familiar with Windows; 75% were male; 59% were from the field of Electrical Engineering; 59% knew some kind of programming; 50% owned their own microcomputer. Their self-perception of knowledge in telecommunications presented a mean of 3.3 (Likert scale with a range of 1 for minimum and 9 for maximum); their enthusiasm in using CBT had a mean of 7.5 (Likert scale with a range of 1 for minimum and 9 for maximum); their previous use of CBT presented a mean of 3.7. The international students (nine total) presented a mean of 592 in the TOEFL, Test of English as a Foreign Language; and their amount of time in the USA presented a mean of 2 years with a range of 0.1 to 4 years.

Potential users were scheduled in advance to participate in the study. Their participation in the evaluation took approximately two hours. Subjects participated one at a time.

4.2.2 Experts

Table 4.2 lists the experts that participated in this study. They were recruited from the Michigan State University community of multimedia designers. They were invited to participate in the evaluation via letter, and no compensation was paid. All five experts were people with whom this researcher had worked before, and felt comfortable that they would be willing and capable of handling the task of making the evaluation and the meta-analysis.

The experts that participated in the meta-analysis had the following characteristics: ages ranging from 24 to 46, with a mean of 35.6 years; 80% were males; all were Americans; previous participation in evaluations similar to this study presented a mean of 4, with a range of 0 to 9 times.

1) PhD. in Educational Technology; Instructional Designer and Professor
2) BS and MS in Physics, MS and PhD. in Computer Science; Hypermedia Designer and Project Manager
3) BA in Telecommunications; MA in Educational Systems Development; Interface Designer and programmer
4) BA in English; MA in Telecommunications; Interface Designer; Project Manager; Hypermedia Designer
5) BS in Astrophysics; MS in Aerospace Engineering; Ph.D. in Educational Technology; Hypermedia Designer and Programmer

Table 4.2: Description of Experts: Qualifications and Job Titles

Experts were scheduled in advance to participate in the study. Their participation in the evaluation took approximately three hours, including the meta-analysis. Experts participated one at a time.

4.3 Procedures

The procedures and instruments utilized in this study are described below. Detailed information is provided for each component of the study.

4.3.1 Orientation

Before the actual evaluation began, each subject was given a brief orientation of the session. During this time, subjects could ask questions as long as they would not interfere with the evaluation itself. The orientation contained the following topics:

1) Informed Consent:

Informed consent was required by the Committee for the Protection of Human Subjects at Michigan State University. This form briefly described the purpose of the research, stated that subjects would be videotaped, and emphasized that subjects were not to feel any coercion to participate in the study. This form, which is included in Appendix A, was signed by all subjects, including the experts.

2) Completion of Preliminary Questionnaire

Subjects were given a preliminary questionnaire covering demographics, prior experience with computers and some attitudinal questions. This questionnaire is included in Appendix B. Once they finished this questionnaire, they were paid with cash for the participation in the study (students only).

3) Background of Evaluation

Subjects were then given a brief description of the purpose of the evaluation and the work they would be doing. This description can be found in Appendix C. Most relevant was the information that the subjects were not being tested, but that the software prototype was the focus of the study.

4) Pre-requisites List

A printed list of teletraffic pre-requisites was presented and explained to subjects prior to the beginning of the evaluation. This list contained relevant information for subjects that were not familiar with the terminology, facts and concepts related to teletraffic engineering. A version of this list can be found in Appendix D.

5) Introduction to the computer system

Subjects were then given a brief introduction to the computer and prototype they were to use. They were asked if they were left or right handed and told to use the left button of the mouse. They were also asked to attach the lapel microphone and given an explanation that the second monitor was facing the opposite direction, for videotaping. They were shown how to wiggle the cursor on the screen every time they wanted to explain something to the observer.

Following the completion of the orientation, the evaluation was begun immediately, and the time limit of one hour was set.

4.3.2 The Evaluation Session

Once the user got started, the observer would maintain verbal contact with the participant, in order to get them to "think-aloud" during the session. This process of reminding the subject to speak would vary, depending on the personality of the participant. At critical points, such as trying to learn to navigate around the program, the observer would ask what specific problems the participant was facing. This critical-incident technique allowed the verification of specific problems by most of the subjects.

In the specific case of the exercise on screen 9, the observer tried to see how far the subjects could get. If the subject was spending too much time without making progress, the observer would ask the participant to verbalize the strategy of solution being attempted and to continue with the evaluation.

The time to explore the prototype was set at a limit of one hour. It would be unrealistic to compile more than one hour of videotape for each subject, and it would be difficult for each subject to allocate more than 3 hours for this evaluation. This time limit was obtained during pilot testing and proved to be adequate for completing the task by most of the participants.

4.3.3 Data Recording

The observer recorded (on paper) key timing and success information during the evaluation. Further notes were made indicating comments made by the subjects, specific observations by the observer, and details of problems or particularly interesting occurrences.

Video recordings were made of each subject, beginning with the introduction of the prototype in the orientation section, and continuing through the follow-up questionnaires. The videotape included the subject's voice, the observer's voice, the computer screen, and the computer sounds (narrations, beeps, mouse clicks and keyboard strokes). The videotapes served as verification of problems encountered and to allow further analysis of subject and experimenter behavior. The videotapes were also the raw material for the production of the multimedia files, which are described in detail below.

4.3.4 Production of Multimedia Files

The creation of multimedia was dictated by the clear need of a tool or instrument that could give random and fast access to the problems and critical incidents detected. The multimedia files were created by following a pre-determined sequence. After each interview, and for every user, the observer would process the videotape by executing a series of steps which are described below.

The videocamera was connected to a multimedia capable computer which had audio digitizing software installed. This setup was used to grab the critical incidences and relevant comments for each user. These audio files were grabbed at 8 bits and 11 Khz of sampling rate, in order to keep files small. This process would typically require the observer to watch the tape in small segments, rewind the tape and start grabbing the audio portion, when relevant information was encountered. Each audio grab was saved with a name that indicated which user and which screen the comment or critical incident was from, as well as a numbering protocol, so that the audio segments could be assembled in cronological order later.

At the end of each digitizing section, A Macromedia Director file was created, when screen shots and audio segments were combined in cronological order, and basic interaction was incorporated so that the files could be manipulated quickly and easily by the researcher. The processing time required to produce a multimedia file for each user would vary, depending on the amount of "thinkaloud" talking and the number of problems encountered, but typically it would take around 2 to 3 hours for each user.

The criteria for including audio segments was based on the severity of the problem and the richness of the comment. This, of course, would vary if another observer would process the data, but there was a basic criteria to include as much data as possible, keeping in mind the usefulness of the comments. This procedure should be improved, specially if more than one observer would be involved in the process of producing the multimedia files.

It is important to mention that during the same time a list of problems was being generated, which included not only problems verbalized by the subjects but also the problems perceived by the researcher. This was accomplished by filling a simple form for each problem, which were compiled later into a comprehensive list of problems, which is described later. Problems were classified among 4 main categories: Interface, Instructional, English-Gramatical, and Programming. At the same time, with the use of another computer, the navigational maps were generated, which allowed a significant reduction of time spent watching videotapes, always a potential problem in this kind of evaluation.

4.3.5 Questionnaires

Once the time limit of one hour expired (immediately after the session) the subjects were asked to answer the questionnaires. The first questionnaire to be filled was the Reeves & Harmon one (see Appendix E). This instrument was presented to the participants in printed form. This questionnaire took a few minutes to be fill out.

Participants then were asked to answer the QUIS questionnaire, which was presented on-line on the same computer utilized for the evaluation (see Appendix F). This instrument took around 20 to 30

minutes to complete. The videotaping was not interrupted at any time during the session.

Once the participants completed the questionnaires, they were asked if they had any comments, suggestions or questions about the whole process. Many users indicated interest in the study and gave interesting comments.

4.3.6 Meta Analysis

The experts, once they had finished answering the QUIS questionnaire, were given a verbal explanation about the meta-analysis to be conducted. In this short explanation, they were told about the need for a classification, evaluation and prioritization of instruments and procedures and were asked to become, for the meta-analysis' sake, "managers" of the prototype's development project they just finished evaluating.

The experts were then asked to examine carefully all ten instruments, the procedures, and the data collected during the potential users' evaluation (described below). Most of this information was presented in printed form in a packet, with the exception of the multimedia files and the statistical data tabulated in Minitab, which had to be presented on the computer.

All 10 instruments and procedures evaluated by the experts in the meta-analysis are described below:

1) Preliminary Questionnaire

The preliminary questionnaire consisted of background information and demographic questions about the participants. This document is presented in Appendix G. This instrument

was presented to the participants in printed form and consisted of a one-sided page. The questionnaire included attitudinal items, computer usage background and general information about the participants.

2) *Reeves and Harmon Questionnaire*

This instrument was included in the evaluation with the objective of incorporating instructional and pedagogical dimensions in addition to interface dimensions. It was composed of two printed pages, containing 10 items on each page. Two main dimensions were defined: Pedagogical and Interface dimensions (see Appendix E).

3) *QUIS Questionnaire*

The QUIS Questionnaire (Questionnaire for User Interface Satisfaction) consists of a total of 69 items, subdivided into 5 categories (overall, screen, terminology, learning, and system). This instrument was included in the study with the objective of incorporating prior research of a well established usability instrument (see Appendix F).

4) *Users Comments From QUIS Questionnaire*

The information presented in this item is a subset of QUIS. All written comments were combined and presented as a separate category. Appendix H presents the comments generated during the users' evaluations. These comments were typewritten at the end of each of the five categories of QUIS (overall, screen, terminology, learning, and system).

5) *Multimedia Files of users*

Multimedia files are qualitative computer documents that were created for each subject. They contained screen shots and audio bites of the problems detected by each user. The problems were extracted from the videotapes, and arranged in chronological order. These multimedia files were created with Macromedia Director 4.0, which has the capability of integrating graphics and audio in an interactive way. These files allowed experts to quickly examine the incidences, taking advantage of the random access nature of digital media.

6) *Navigational Maps*

Navigational-visualization maps were included in this study with the objective of providing a way of verifying the navigation and frequency of screen visits for each participant. In this study, the navigational maps were generated during the video compilations. An example of a navigational map is available in Appendix I. These maps include indications in numerical order of screens visited, visualized in a sequential diagram of screens. The observer's comments were included on these maps, in order to contextualize the navigational strategy of each user.

7) *Ethnic Groups Results*

The objective of including ethnic groups in this provided a diverse range of usability perspectives from international users. Descriptive statistics were presented showing differences among the three different ethnic groups. The multimedia files,

and QUIS comments files also indicated the national origin of the participants.

8) *List of Problems*

A list of problems was generated with the objective of providing an efficient way of reporting and summarizing the problems detected by all subjects, including experts. With this list, which can be seen in Appendix J, experts could quickly grasp the incidence, the location and the description of each problem. This list was generated by the researcher in the order of coding, when watching the videotapes. This list categorizes the problems encountered into four groups: Interface, Instructional, English and Programming problems.

9) *Report of Usage*

This item was created with the objective of providing the experts with a report of usage, which included the amount of time spent on the prototype, the number of screens visited, and the level of response to the exercises and quizzes for each participant tested. Descriptive statistics were presented to the experts (Appendix K).

10) *Statistical Data Tabulated in Minitab*

All the data generated in the evaluation was tabulated and presented to the experts using the statistical software Minitab. A total of 114 variables were generated, including data of all the instruments described above. A printout of this information was given to the experts for reference, consultation and manipulation (Appendix L).

During the meta-analysis, the observer was available for questions and clarifications of any kind. Once the experts finished the analysis of the instruments, they were asked to answer a ten-item Likert scale questionnaire about the instruments and procedures. This questionnaire is available in Appendix G. The meta-analysis was also recorded on videotape for further clarification.

4.4 Evaluation Chronology

Initial work on this evaluation began in October of 1994. Related research literature was studied, and potential prototypes and instruments were considered. Originally, this evaluation was intended to include 4 or 5 ethnic groups. Because of the difficulty of obtaining the subjects, however, the study was narrowed to include three main ethnic groups. Two subjects participated in a pilot study by December of 1994. The pilot study led to considerable changes in the study, particularly for the sake of the users. The Reeves & Harmon Questionnaire was shortened, as well as QUIS Questionnaire. A list of pre requisites was incorporated to the information given prior to each session in order to allow non-telecommunication engineering students to be able to participate in the evaluation.

The first subject was observed in March 1995, after which no more changes were made to the evaluation design nor to the instruments or the prototype. All users participated by June 1995. During the month of July, the data collected from the users' evaluations were compiled, summarized and prepared to be shown to the experts. Experts participated in the evaluation in August and September of 1995.

Chapter 5

Analysis

This chapter presents and analyzes the quantitative and qualitative data collected during this study. It is organized as follows: first, the analysis methodologies are surveyed and a rationale is given for the methodology used. Second, the information collected is described and analyzed. Last, a summary of the analysis is presented.

5.1 Choices for Analysis

A primary decision that must be made in any observational research is how to analyze the data that are collected. This decision is largely determined by the general knowledge of the field of study, the specific knowledge of the problem domain, and the perspective of the researcher.

5.1.1. Different Approaches to the Same Problem

The current stage of human-computer interface research depends on the perspective of the researcher, since it involves a cross-section of the fields of computer science, education, cognitive science, and statistics.

From the computer science standpoint, the discipline of usability engineering has the objective of designing better interfaces. Professionals that take this approach are very practical in terms of finding out how to make applications more usable. Research of this nature studies the interaction of users and computers with the objective of finding statistically significant results.

From the cognitive and educational standpoint, human-computer interaction is located between a descriptive and an explanatory standpoint. Human thinking is not understood well enough to predict it entirely. Therefore, research of this nature tries to expand the knowledge of the interaction between humans and machines. Case studies of small a number of users are ideal in this scenario [Bell, 1992].

Statistics can improve the researcher's capacity to generalize the results. Since differences are observed and quantified, statistical methods can be used to verify whether these results are significant. When questions can be asked in advanced with enough detail a study could include these questions.

This study attempts to incorporate a combination of the above approaches. The goal was to gain a better understanding of the methodologies for testing educational interfaces, combined with problem detection and attitudes. In order to accomplish this, it combines detailed observation of subjects found in a case study with the specific comparison of different types of users and specialists, such as in human factors studies.

5.1.2. Additional Considerations

In addition to the concerns mentioned above, other factors influenced the statistical approach taken in the study, the most important one being the limited availability of participants. Since a significant amount of the data collected were categorical, and other data were not normally distributed, parametric analyses could not be utilized. This was a critical decision because this researcher wanted

to follow a trend observed in previous studies [Nielsen and Landauer 1993].

This researcher wanted to include several categories of users because of the issue of international use of interfaces, and because previous research indicates that particular types of interfaces are better suited for particular types of users [Nielsen 1990]. This researcher also wanted to use several different instruments in this evaluation because this would allow for the comparison of results and opinions about their relative usefulness. Consequently, the study was designed to include qualitative and quantitative instruments which should help to better define future evaluations of this kind.

5.2 Statistical Choices

5.2.1 *Use of non-parametric methods*

The use of Kruskal-Wallis and Mann-Whitney U tests were chosen because of the small number of participants in this research, the non-normal distribution of its variables, the unequal variances of the sample groups, and the scale inconsistencies over the range of measurements. Another reason was the great variability of responses obtained in the questionnaires.

A more liberal alpha level was adopted for testing the hypotheses in order to improve the probability of detecting differences between the groups. The researcher considered that the risk of committing type II error in this case was not serious. The computer software utilized for this statistical procedure was StatSoft [1991].

5.2.2 Use of Cluster Analysis

The researcher decided to use cluster analysis for exploring the data generated by the questionnaires and for grouping users. In human-computer interaction research, subjects are analyzed in small numbers, and measures usually are characterized by fairly large amounts of error variance, as well as being of indeterminate underlying distribution. This limits the use of factor analytic techniques [Kirakovsky and Corbett 1990]. Much more useful are the cluster analysis techniques.

By definition, the cluster analysis is based on one or more similarities coefficients, or distance measures [Aldenderfer & Blashfield, 1984; Morris, Blashfield, & Satz, 1981]. The cluster analysis in the present study employed the k-means clustering method. The computer application used for this procedure was StatSoft [1991]. Computationally, the k-means clustering technique minimizes variability within clusters and maximizes variability between clusters. The program tries to move cases in and out of groups (clusters) to get the most significant results.

In this kind of cluster technique, the researcher specifies, in advance, the number of clusters and the computer clusters the cases accordingly. In the present study, the researcher tried different numbers of clusters, to see if the participants would cluster in more distinct groups. Two and four cluster analyses were tried, using a maximum of 10 iterations (rotations). The best solution was found in between two and five iterations for the majority of the analyses. Missing data values were substituted by means. In a sense, cluster analysis finds the most practical and efficient solution possible.

5.3 Results

5.3.1. *QUIS Comparisons between Ethnic Groups, Users and Experts*

Figure 5.1 presents the computations of the means and ranges for users, experts and ethnic groups, as well as non-parametric median test results among ethnic users and among all users and experts. Results which obtained statistical significances are indicated accordingly. The Kruskal-Wallis test was executed among all four groups; the Mann-Whitney test was performed among all user group and experts. For each main category of variables, means were computed for each group, as well as overall means for each group. An examination of Table 5.1 indicates the presence of differences between ethnic groups, and differences between experts and users. A pattern was detected when the values of each sub-category were compared. Figures 5.2 through 5.7 show graphically these differences.

Figure 5.2 compares users and experts and displays at the same time all questions of QUIS. This graphic clearly shows a trend -- experts were, in general, more critical than users when evaluating this prototype. It also shows a wider range among experts' answers.

Figure 5.3 focuses on the Overall aspects of the interface. This graphic shows all three ethnic groups, their combinations and the experts' responses. The item "Adequate Power" was the only item that did not obtain any statistical significance, when comparing users and experts.

Group:	Indian		Chinese		Americans		All Users		Experts		Kruskal (Ethnic Groups)	Mann-Whitney (Users & Experts)
	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean		
Overall User Reactions:		6.3		5.9		5.0		5.6		3.9		
Wonderful	7 to 8	7.2	6 to 9	7.0	4 to 6	5.2	4 to 9	6.3	2 to 7	4.2	$p < 0.01$	$p < 0.01$
Satisfying	3 to 8	5.2	3 to 7	5.4	2 to 6	4.0	2 to 8	4.7	2 to 6	3.2	$p < 0.2$	$p < 0.1$
Stimulating	7 to 9	8.2	5 to 7	5.6	3 to 7	4.8	3 to 9	6.1	1 to 8	4.0	$p < 0.05$	$p < 0.2$
Easy To Use	2 to 8	5.4	4 to 8	6.0	2 to 8	5.0	2 to 8	5.4	1 to 7	3.2		$p < 0.1$
Adequate Power	5 to 7	6.4	3 to 7	5.6	4 to 9	6.8	3 to 9	6.2	2 to 8	5.4		
Flexible	4 to 8	5.4	5 to 7	6.0	1 to 7	4.0	1 to 8	5.0	2 to 7	3.6	$p < 0.2$	$p < 0.2$
Screen Aspects:		7.4		6.2		6.8		6.8		5.3		
Easy to Read	8 to 9	8.8	4 to 8	7.0	5 to 9	8.0	4 to 9	8.0	4 to 9	7.6	$p < 0.2$	
Sharp Characters	7 to 9	8.5	5 to 9	6.8	7 to 9	8.4	5 to 9	7.9	6 to 9	6.8	$p < 0.2$	
Very Legible	7 to 9	8.6	4 to 9	7.0	9 to 9	9.0	4 to 9	8.3	7 to 9	8.0	$p < 0.1$	
Hints on screen helpful	8 to 9	8.6	5 to 8	6.8	7 to 9	8.2	5 to 9	7.8	2 to 9	6.5	$p < 0.2$	
Use of Reverse Video	5 to 8	6.6	3 to 9	7.0	8 to 9	8.7	3 to 9	7.2				
Use of Binking	6 to 9	8.2	5 to 8	6.3	5 to 9	7.5	5 to 9	7.5				
Screen Layouts	6 to 9	7.6	3 to 8	6.2	3 to 8	6.8	3 to 9	6.9	2 to 9	4.8		$p < 0.1$
Amount of info on Screen	6 to 9	7.2	2 to 7	5.0	1 to 9	5.0	1 to 9	5.9	2 to 9	6.0		
Arrangement of info on Screen	7 to 9	8.0	3 to 7	5.8	4 to 8	6.4	3 to 9	6.6	1 to 9	4.8	$p < 0.2$	
Sequence of Screens	3 to 9	6.2	4 to 8	6.0	2 to 9	4.8	2 to 9	5.5	1 to 7	4.2		
Next Screen in a Sequence	1 to 9	5.4	2 to 8	5.0	2 to 9	4.6	1 to 9	4.9	1 to 7	4.2		
Going back to previous screen	3 to 9	6.0	3 to 8	6.6	4 to 9	6.6	3 to 9	6.1	2 to 4	3.0	$p < 0.1$	$p < 0.05$
Beginning, Middle, End of Tasks	4 to 9	6.8	4 to 8	5.6	3 to 7	4.8	3 to 9	5.4	1 to 4	2.2	$p < 0.01$	$p < 0.01$
Terminology & System Info		7.0		6.0		6.2		6.3		5.0		
Use of Terms Throughout System	6 to 9	7.8	5 to 8	6.2	8 to 9	8.4	5 to 9	7.5	6 to 9	7.0	$p < 0.1$	
Task Terms	7 to 9	8.2	4 to 8	6.2	6 to 9	8.0	4 to 9	7.4	5 to 9	6.6	$p < 0.1$	
Computer Terms	7 to 9	8.5	6 to 7	6.7	7 to 9	8.0	6 to 9	7.8	5 to 9	6.8	$p < 0.2$	
Terminology Related to Your Work	5 to 9	4.2	2 to 7	5.8	1 to 7	4.2	1 to 9	4.9	1 to 9	5.5		
Computer Terminology Used	5 to 9	7.6	4 to 8	6.3	8 to 9	8.5	4 to 9	7.5	5 to 9	7.3		
Terms on the Screen	4 to 8	6.8	3 to 7	5.6	4 to 9	7.0	3 to 9	6.6	2 to 7	3.6	$p < 0.1$	$p < 0.05$
Messages that appear on screen	2 to 9	6.8	1 to 7	5.2	5 to 9	7.2	1 to 9	6.4	1 to 8	5.0		$p < 0.2$
Position of instructions on screen	7 to 9	8.0	2 to 8	6.0	3 to 9	5.8	2 to 9	6.4	3 to 8	5.4		
Messages that appear on screen	6 to 9	7.6	3 to 8	5.8	3 to 8	6.0	3 to 9	6.3	2 to 7	4.8	$p < 0.2$	$p < 0.2$
Instructions for commands	6 to 9	7.6	2 to 9	6.6	3 to 8	6.2	2 to 9	6.5	2 to 7	4.0	$p < 0.2$	$p < 0.1$
Instructions for correcting errors	5 to 9	6.8	1 to 8	6.0	3 to 8	4.5	1 to 9	5.6	1 to 8	3.4		$p < 0.2$
Intuits what it is doing	3 to 9	5.8	2 to 8	5.4	2 to 6	3.8	2 to 9	4.9	2 to 8	4.4		
Predictable Results	4 to 9	6.8	2 to 8	6.2	3 to 8	4.6	2 to 9	5.7	2 to 8	4.2		$p < 0.2$
User control feedback amount	3 to 7	5.5	3 to 7	5.6	2 to 7	4.0	2 to 7	4.7	1 to 4	2.4	$p < 0.05$	$p < 0.05$
Learning Aspects		6.7		5.7		5.9		5.9		4.7		
Learn to operate the system	6 to 9	7.8	6 to 7	6.8	5 to 9	6.4	5 to 9	6.7	2 to 9	5.4		
Getting Start	3 to 9	6.4	1 to 8	5.6	2 to 9	5.8	1 to 9	5.7	2 to 9	6.0		
Learn advanced features	3 to 8	5.6	3 to 7	5.3	1 to 8	3.8	1 to 8	4.9	2 to 8	4.5		
Time to learn to use	5 to 7	6.2	2 to 8	5.6	3 to 9	5.6	2 to 9	5.6	2 to 8	4.3		
Explore features by trial and error	1 to 9	6.2	3 to 9	6.2	3 to 9	5.6	1 to 9	5.8	3 to 7	5.3		$p < 0.1$
Exploration of Features	6 to 9	8.2	4 to 8	6.4	2 to 9	4.8	2 to 9	6.1	2 to 6	3.8	$p < 0.05$	$p < 0.1$
Discovering new features	5 to 9	6.6	5 to 8	6.6	4 to 8	6.4	4 to 9	6.2	2 to 7	4.0		
Remember names & commands	1 to 9	5.6	1 to 8	5.6	4 to 9	6.2	1 to 9	5.9	4 to 8	5.3		
Remember rules about command	3 to 9	5.8	3 to 8	6.0	4 to 8	6.0	3 to 9	6.1	3 to 8	5.3		
Tasks are straight-forward	5 to 8	6.8	1 to 8	5.0	3 to 8	5.4	1 to 8	5.6	3 to 7	5.2		
Number of steps per task	6 to 9	8.0	5 to 9	7.0	4 to 8	6.4	4 to 9	7.0	5 to 9	6.2		
Steps of a sequence are logical	6 to 9	8.2	2 to 8	5.6	5 to 9	7.4	2 to 9	6.8	4 to 8	5.8	$p < 0.1$	
Completion of a sequence of step	7 to 9	8.2	3 to 7	5.6	1 to 8	5.0	1 to 9	6.0	1 to 8	5.0	$p < 0.05$	
Help messages on screen	5 to 9	6.0	2 to 8	6.0	5 to 9	7.3	2 to 9	6.2	2 to 9	5.0		
Accessing help messages	5 to 8	6.3	1 to 7	4.6	2 to 9	6.0	1 to 9	5.8	1 to 7	2.8	$p < 0.2$	$p < 0.05$
Content of Help Messages	5 to 8	6.3	1 to 5	4.2	6 to 7	6.7	1 to 9	5.7	2 to 8	4.0	$p < 0.2$	
Amount of Help	4 to 8	5.7	1 to 5	4.2	4 to 7	5.7	1 to 8	4.9	1 to 6	2.4		$p < 0.1$
System Capabilities		7.6		6.5		6.7		6.9		6.7		
System Speed	5 to 9	8.0	5 to 9	7.0	9 to 9	9.0	5 to 9	8.0	6 to 9	8.2	$p < 0.1$	
Response time	8 to 9	8.6	6 to 9	7.6	9 to 9	8.8	6 to 9	8.3	8 to 9	8.2		
Rate of info displayed	6 to 9	7.8	3 to 8	5.8	8 to 9	8.8	3 to 9	7.6	6 to 9	7.8	$p < 0.1$	
How reliable is the system	5 to 9	8.0	4 to 8	6.4	3 to 9	7.6	3 to 9	7.3	4 to 9	7.8		
Operations are dependable	6 to 9	8.3	2 to 8	5.6	3 to 9	7.6	2 to 9	7.1	5 to 9	7.8	$p < 0.2$	
Falkies occur seldom	6 to 9	8.0	3 to 9	7.0	5 to 9	8.0	3 to 9	7.6	9 to 9	9.0	$p < 0.2$	$p < 0.2$
System warns problems	1 to 7	3.7	1 to 9	4.6	1 to 1	1.0	1 to 9	3.6	5 to 7	5.7		
System tends to be quiet	9 to 9	9.0	2 to 9	6.8	3 to 9	7.6	2 to 9	7.8	8 to 9	8.8	$p < 0.1$	
Mechanical devices are quiet	8 to 9	8.6	1 to 9	7.2	9 to 9	9.0	1 to 9	8.0	8 to 9	8.8		
Computer sounds are pleasant	8 to 9	8.8	5 to 9	7.8	3 to 9	5.2	3 to 9	6.9	9 to 9	9.0	$p < 0.1$	
Correcting typos is easy	8 to 9	7.0	3 to 9	6.6	1 to 9	4.6	1 to 9	6.1	2 to 9	5.0		
Ability to undo operations	8 to 9	8.5	3 to 9	6.0	5 to 9	7.6	3 to 9	7.3	9 to 9	9.0		
Needs of both new & experts	5 to 9	7.5	2 to 9	5.3	2 to 9	5.3	2 to 9	5.7	2 to 8	5.4		
Novices can start using easily	3 to 9	5.6	3 to 8	6.6	2 to 6	4.4	8 to 9	5.4	1 to 7	4.0		
Experts can use features easily	3 to 8	5.2	2 to 8	6.3	2 to 8	5.2	2 to 8	5.7	1 to 9	3.8		
Grand mean	7 to 9	8.5	7 to 9	8.0	6 to 9	7.5	6 to 9	8.0	1 to 4	3.0	$p < 0.1$	$p < 0.05$
		7.1		6.1		6.2		6.4		5.3		

Table 5.1: QUIS Comparisons between Ethnic Groups and Experts.

3

2

;

6

5

3

2

!

1

1

d.

gra

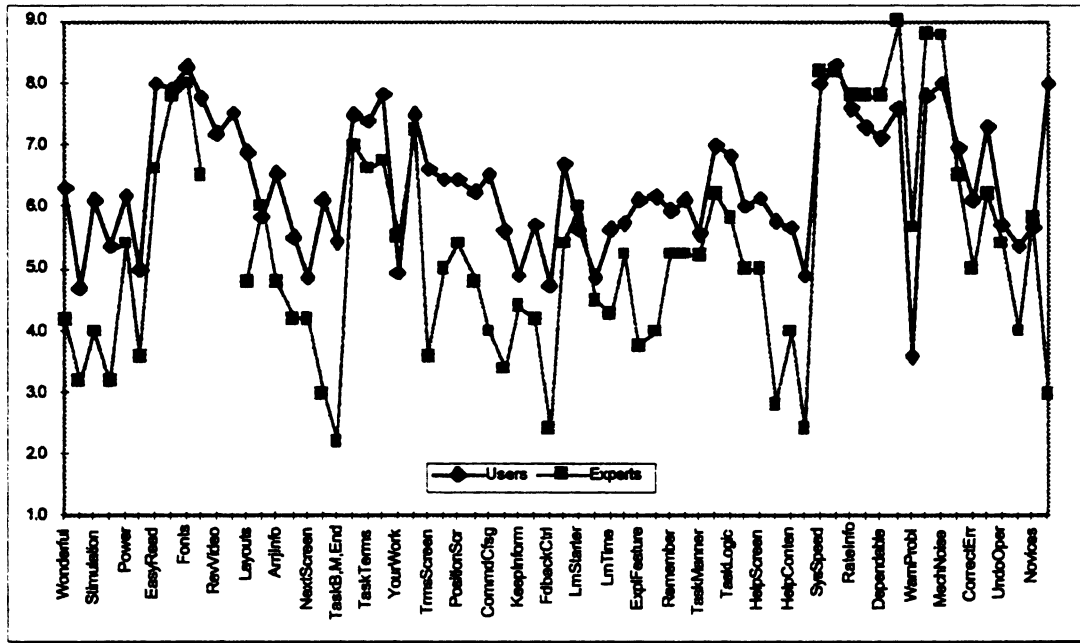


Figure 5.2: QUIS answers - Comparison of users and experts

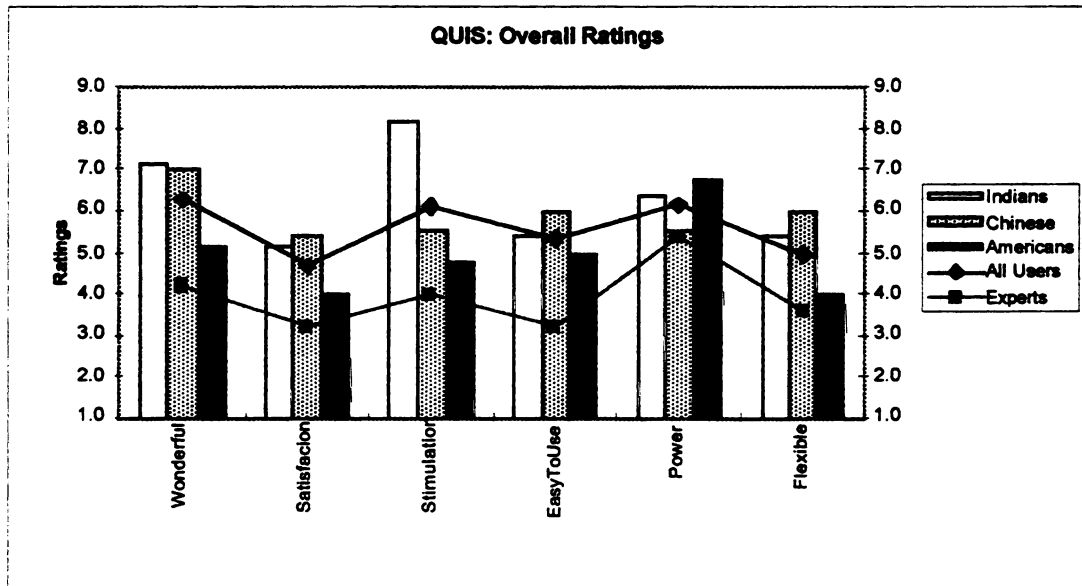


Figure 5.3: QUIS answers - Overall items - Ethnic group users and experts

Figure 5.4 focuses on the screen aspects of the interface. This graphic shows all three ethnic groups, their combination and the

experts' responses. One observation is the lack of responses on two items by the experts group: "Reverse Video" and "Blinking", which indicates a more careful interpretation of the questions by them. The use of reverse video and blinking was not available in the prototype tested. "Going back to previous screen" and "Beginning, Middle and End of Tasks" both presented differences that were statistically significant. This could represent that experts were more efficient than users in finding navigational problems in the prototype.

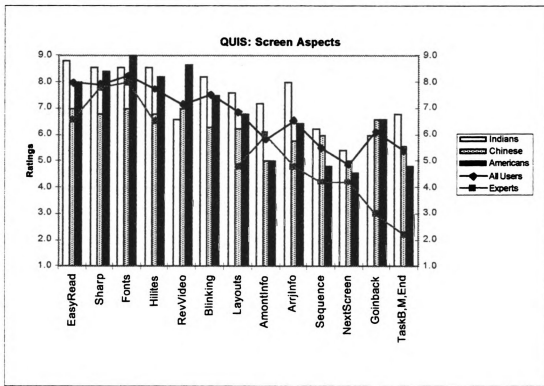


Figure 5.4: QUIS answers - Screen items - Ethnic group users and experts

Figure 5.5 focuses on the Terminology aspects of the interface. This graphic shows all three ethnic groups, their combinations and the experts' responses. Several items presented significant differences: "Terms on Screen" and "User Control Feedback" obtained higher probability levels. It is interesting to note that experts were

more critical than users regarding terminology. This could be due to the fact that none of the experts were familiar with the content.

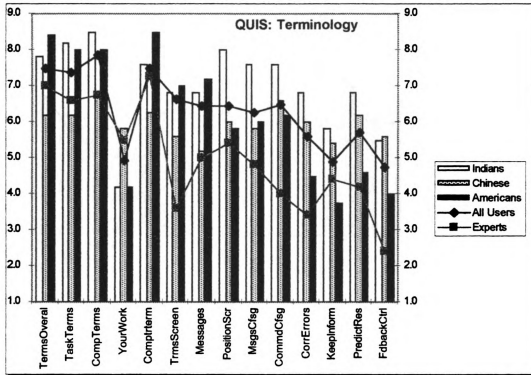


Figure 5.5: QUIS answers - Terminology items - Ethnic group users and experts

Figure 5.6 focuses on the Learning aspects of the interface. This graphic shows all three ethnic groups, their combinations and the experts responses. “Accessing help messages” was statistically significant.

Figure 5.7 focuses on the System aspects of the interface. This graphic shows all three ethnic groups, their combination and the experts’ responses. In this category, experts were more positive in relation to other categories. One item that obtained high significance statistically was “Experts can use features easily” - users rate this item high (8) and experts low (3). This could be explained by the

possibility that experts had a lower expectation of the efficiency of the prototype in relation to users.

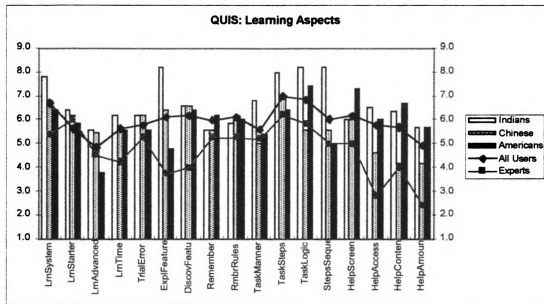


Figure 5.6: QUIS answers - Learning items - Ethnic group users and experts

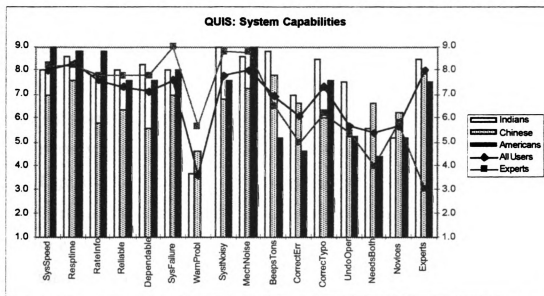


Figure 5.7: QUIS answers - System items - Ethnic group users and experts

An important trend detected in this section was that, among the ethnic groups, the Indians were more positive about the

prototype in all categories. The Chinese were more critical than the American users, with the exception of the overall aspects category.

5.3.2 *QUIS Comparisons between Gender*

Table 5.8 presents the computations of means and range for males and females, as well as non-parametric median test (Mann-Whitney) results. Results which obtained statistical significance are indicated accordingly. For each main category of variables, means were computed for each gender, as well as overall means for each gender. An examination of Table 5.8 indicates the presence of differences between gender. A pattern was detected, then the values for each category were compared. Figures 5.9 through 5.14 show graphically these differences. Figure 5.9 compares gender and displays at the same time all questions of QUIS. This graphic shows a trend detected in this section: males were slightly more critical than females when evaluating this prototype.

Figure 5.10 focuses on the Overall aspects of the prototype. With the exception of the item “Flexible”, all items were graded lower by males. In terms of statistical significance, only “Wonderful” was the only item significant (although at a level of .2).

Figure 5.11 focuses on the Screen aspects of the prototype. The item “Screen Layouts” was statistically significant at $p < .05$. The sequence of screens was significant at $p < .2$. In this category, the differences of ratings between males and females were less evident, although existent.

	Male		Female		Mann-Whitney
	Range	Mean	Range	Mean	
Overall User Reactions:		5.55		6.25	
Wonderful	4 to 7	6.18	5 to 9	7.25	p <0.2
Satisfying	2 to 7	4.54	3 to 8	5.75	
Stimulating	3 to 9	5.82	7 to 8	7.25	
Easy To Use	2 to 8	5.28	5 to 7	6.00	
Adequate Power	3 to 9	6.19	6 to 7	6.50	
Flexible	1 to 8	5.27	3 to 6	4.75	
Screen Aspects:		6.68		7.24	
Easy to Read	4 to 9	7.73	8 to 9	8.50	
Sharp Characters	5 to 9	7.90	7 to 9	8.00	
Very Legible	4 to 9	8.27	7 to 9	8.00	
Hit/ites on screen helpful	5 to 9	7.73	8 to 9	8.25	
Use of Reverse Video	3 to 9	7.37	5 to 9	7.00	
Use of Blinking	5 to 9	7.44	6 to 9	7.67	p <0.05
Screen Layouts	3 to 8	6.36	8 to 9	8.25	
Amount of info on Screen	1 to 9	5.73	2 to 8	5.75	
Arrangement of info on Screen	3 to 9	6.54	6 to 8	7.25	
Sequence of Screens	2 to 9	5.19	4 to 8	7.00	p <0.2
Next Screen in a Sequence	2 to 9	4.45	1 to 9	6.50	
Going back to previous screen	2 to 9	6.64	3 to 8	5.75	
Beginning, Middle, End of Tasks	1 to 9	5.55	4 to 8	6.25	
Terminology & System Info		6.21		6.88	
Use of Terms Throughout System	5 to 9	7.36	6 to 9	7.75	
Task Terms	4 to 9	7.37	6 to 9	7.75	
Computer Terms	6 to 9	7.87	7 to 9	7.67	
Terminology Relate to Your Work	1 to 8	4.64	1 to 9	5.00	
Computer Terminology Used	4 to 9	7.67	5 to 9	7.00	
Terms on the Screen	3 to 9	6.00	6 to 9	7.50	
Messages that appear on screen	1 to 9	6.45	2 to 9	6.25	
Position of instructions on screen	2 to 9	6.19	7 to 9	7.75	
Messages that appear on screen	3 to 8	6.18	6 to 9	7.25	
Instructions for commands	2 to 9	6.55	6 to 9	7.50	
Instructions for correcting errors	1 to 9	5.73	5 to 8	6.33	
Informs what it is doing	2 to 9	4.90	3 to 8	5.50	
Predictable Results	2 to 9	5.18	7 to 8	7.75	p <0.05
User control feedback amount	1 to 7	4.90	3 to 7	5.25	
Learning Aspects		6.04		6.19	
Learn to operate the system	2 to 9	6.12	5 to 9	7.20	
Getting Start	1 to 9	7.10	2 to 7	6.75	
Learn advanced features	1 to 8	6.18	4 to 6	5.25	
Time to learn to use	2 to 9	6.00	3 to 7	5.25	
Explore features by trial and error	1 to 9	5.53	4 to 8	6.00	
Exploration of Features	1 to 9	6.10	4 to 8	5.75	
Discovering new Features	1 to 9	6.80	5 to 6	5.75	
Remember names & commands	1 to 9	6.00	3 to 6	5.00	
Remember rules about command	3 to 9	6.36	3 to 6	4.75	p <0.2
Tasks are straight-forward	1 to 8	5.45	5 to 8	6.50	
Number of steps per task	4 to 9	7.00	6 to 9	7.50	
Steps of a sequence are logical	2 to 9	6.90	6 to 9	7.50	
Completion of a sequence of step	1 to 9	5.73	7 to 9	7.75	p <0.2
Help messages on screen	2 to 9	6.37	5 to 8	6.25	
Accessing help messages	1 to 9	5.00	5 to 8	6.75	
Content of Help Messages	1 to 9	5.37	5 to 7	5.67	
Amount of Help	1 to 8	4.75	5 to 7	5.66	
System Capabilities		6.89		7.06	
System Speed	5 to 9	8.12	5 to 9	7.50	
Response time	6 to 9	8.45	7 to 9	8.00	
Rate of info displayed	3 to 9	7.70	6 to 9	7.25	
How reliable is the system	3 to 9	7.40	5 to 9	7.00	
Operations are dependable	2 to 9	6.90	6 to 9	7.50	
Failures occur seldom	3 to 9	7.10	9 to 9	9.00	p <0.1
System warns problems	1 to 7	3.28	1 to 9	4.33	
System tends to be quiet	2 to 9	7.36	9 to 9	9.00	p <0.2
Mechanical devices are quiet	1 to 9	7.89	8 to 9	8.67	
Computer sounds are pleasant	2 to 9	7.00	6 to 9	8.00	
Correct your mistakes is easy	3 to 9	6.19	1 to 9	5.75	
Correcting pos is simple	3 to 9	7.19	7 to 8	7.67	
Ability to undo operations	2 to 9	6.30	2 to 7	4.50	
Needs of both new & experts	3 to 9	5.63	4 to 6	5.25	
Novices can start using easily	2 to 8	5.40	3 to 7	5.75	
Experts can use features easily	7 to 9	8.40	6 to 9	7.75	
Grand Mean		6.28		6.72	

Table 5.8: QUIS answers - Gender comparison

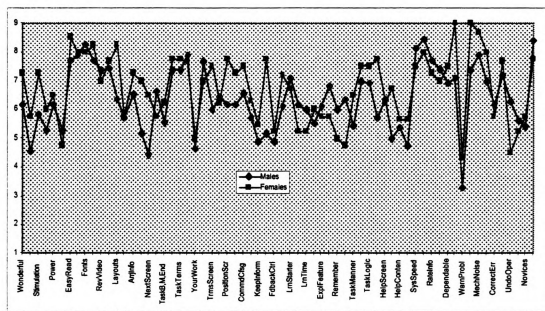


Figure 5.9: QUIS answers - Gender comparison

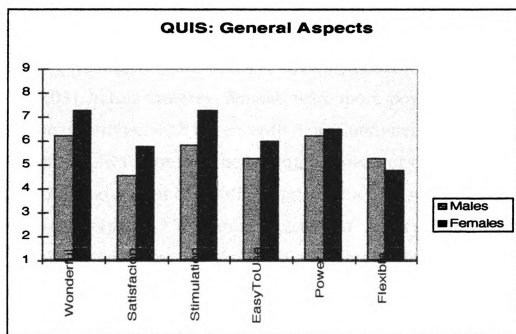


Figure 5.10: QUIS answers - Gender - Overall aspects

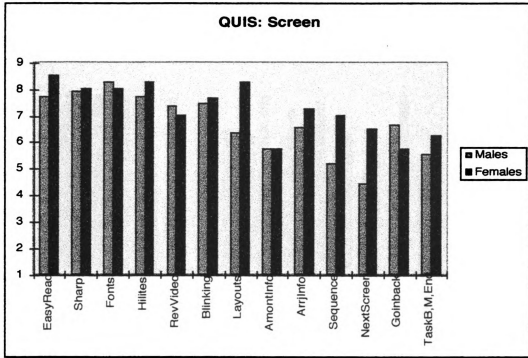


Figure 5.11: QUIS answers - Gender -Screen aspects

Figure 5.12 focuses on the Terminology aspects of the prototype. “Predictable Results” was the only statistically significant item ($p < .05$). In this category, females were more positive when rating the prototype, with the exception of “Computer Terms”.

Figure 5.13 focuses on the Learning aspects of the prototype. This section had a balanced ratings and significance was obtained in only two items ($p < .2$): “Remember Rules” and “Steps of a Sequence”.

Figure 5.14 focuses on the System aspects of the prototype. This section had a relative balance in ratings, and significance was obtained in two items: “Failures Occur Seldom” ($p < .1$) and “System tends to be quiet” ($p < .2$)

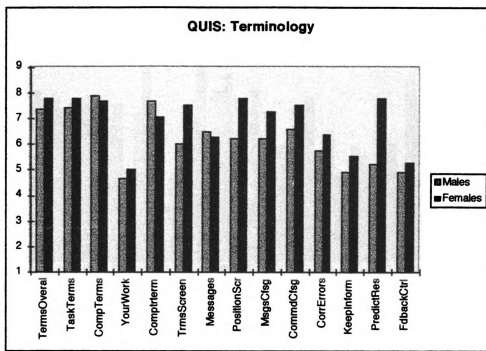


Figure 5.12: QUIS answers - Gender -Terminology aspects

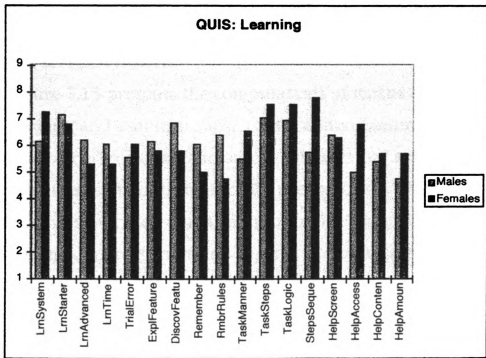


Figure 5.13: QUIS answers - Gender -Learning aspects

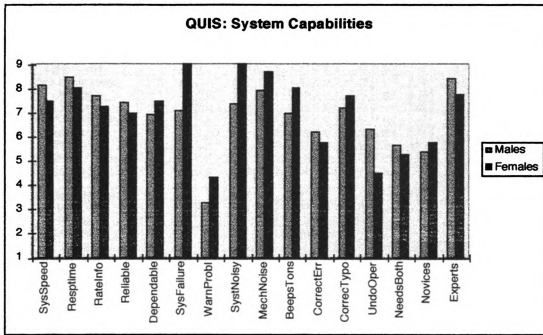


Figure 5.14: QUIS answers - Gender -System aspects

5.3.3 Reeves & Harmon Questionnaire: Comparisons between Ethnic Groups, Users and Experts

Figure 5.15 presents the computations of means and range for users, experts and ethnic groups, as well as non-parametric test results among ethnic users and among all users and experts. Results which obtained statistical significance are indicated. The Kruskal-Wallis test was executed among all groups; the Mann-Whitney test was performed among users and experts. For both categories of variables, means were computed for each group. An examination of Table 5.15 indicates, again, the presence of differences between ethnic groups, and differences between experts and users.

	Indians		Chinese		Americans		All Users		Experts		Kruskal	Mann-Whitney
	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean	(Ethnic Groups)	(Users & Experts)
Learning Dimensions:		3.66		3.20		2.94		3.20		2.75		
Goal Orientation	3 to 5	4.00	2 to 4	3.20	1 to 4	2.80	1 to 5	3.19	2 to 5	3.60		
Experiential Value	3 to 5	4.00	2 to 4	3.00	3 to 4	3.20	2 to 5	3.44	1 to 4	2.40	$p < 0.1$	$p < 0.1$
Use of Error as Learning	4 to 5	4.25	2 to 4	2.80	1 to 4	2.80	1 to 5	3.20	1 to 4	2.40	$p < 0.2$	
Motivation	3 to 5	3.60	3 to 4	3.60	2 to 4	2.80	2 to 5	3.25	1 to 4	2.60	$p < 0.2$	$p < 0.1$
Structure	2 to 5	3.80	2 to 5	3.40	2 to 3	2.60	1 to 5	3.13	2 to 4	3.00		
Acomm. Individual Differ.	1 to 4	2.20	2 to 4	3.40	2 to 4	3.00	1 to 4	2.81	1 to 2	1.40	$p < 0.05$	$p < 0.05$
Learner Control	2 to 4	3.40	2 to 4	3.60	3 to 4	3.40	1 to 4	3.31	1 to 4	2.40		$p < 0.2$
User Activity	3 to 4	3.75	2 to 4	2.80	3 to 5	3.80	2 to 5	3.40	3 to 4	3.60		
Pedagogy of Objectives	3 to 4	3.60	2 to 5	3.00	1 to 4	2.75	1 to 5	3.20	3 to 4	3.80		
Cognitive Psychology	4 to 4	4.00	3 to 4	3.20	2 to 3	2.25	2 to 4	3.08	1 to 3	2.25	$p < 0.05$	$p < 0.2$
Interface Dimensions:		4.09		3.28		3.07		3.47		2.66		
Easy of Use	4 to 5	4.20	2 to 4	3.60	2 to 5	3.20	2 to 5	3.56	1 to 4	3.00		
Navigation	3 to 5	4.20	3 to 4	3.40	2 to 4	3.00	1 to 5	3.38	1 to 4	2.40	$p < 0.1$	$p < 0.2$
Cognitive Load	3 to 5	4.30	2 to 4	3.00	3 to 5	3.60	2 to 5	3.57	1 to 5	3.40		
Mapping	2 to 5	3.60	2 to 3	2.80	1 to 3	2.60	1 to 5	3.00	1 to 3	1.80	$p < 0.1$	$p < 0.05$
Screen Design	4 to 5	4.40	1 to 5	3.80	2 to 4	3.20	1 to 5	3.81	1 to 5	2.80		$p < 0.2$
User Control	2 to 5	3.60	2 to 4	3.20	2 to 4	2.80	2 to 5	3.25	2 to 4	2.80		
Information Presentation	3 to 4	3.80	2 to 5	3.00	1 to 4	2.80	1 to 5	3.31	2 to 5	3.20		
Media Integration	4 to 5	4.40	3 to 4	3.40	2 to 4	3.40	2 to 5	3.69	1 to 5	2.20	$p < 0.1$	$p < 0.05$
Aesthetics	4 to 5	4.40	2 to 4	3.00	3 to 4	3.25	2 to 5	3.67	1 to 5	2.40	$p < 0.05$	$p < 0.1$
Overall Functionality	3 to 5	4.00	3 to 4	3.60	2 to 3	2.80	2 to 5	3.44	1 to 4	2.60	$p < 0.05$	$p < 0.2$
Grand mean		3.88		3.24		3.00		3.33		2.70		

Table 5.15: Reeves & Harmon answers - Comparison of Users and Experts

A pattern was detected when the values of each dimension were compared. Figures 5.16 through 5.18 show graphically these differences.

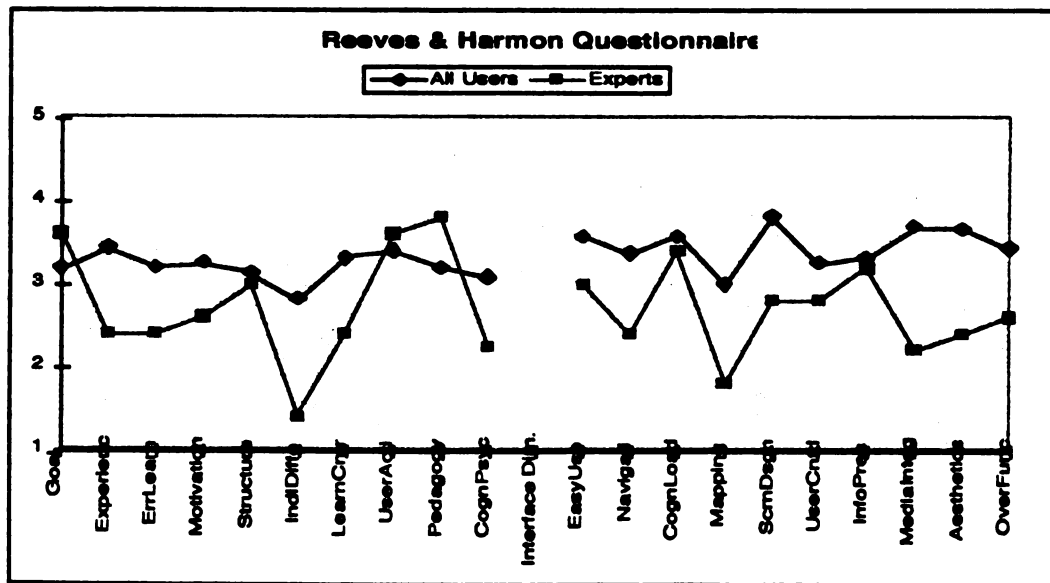


Figure 5.16: Reeves & Harmon answers - Comparison of Users and Experts

Figure 5.16 compares users and experts and displays all items of the Questionnaire. This graphic confirms the trend found with QUIS: experts were in, general, more critical than users when evaluating this prototype. It also confirms the trend of more variability in range among the experts. Two items presented significant differences: Individual Differences and Media Integration. The first item matched with results from QUIS (Experts graded the individual differences lower). The second item presents new a perspective. Experts considered the media integration poorer than users. A plausible explanation would be their previous knowledge of other instructional multimedia applications.

Figure 5.17 focuses on the Learning aspects of the prototype. This graphic shows all three ethnic groups, their combinations and the experts' responses. The item "Accommodation of Individual Differences" was significant at $p < .05$, "Experiential Value" and "Motivation" were significant at $p < .1$, "Learning Control" and "Cognitive Psychology" were significant at $p < .2$.

Figure 5.18 focuses on the Interface aspects of the prototype. This graphic shows all three ethnic groups, their combination and the experts' responses. The items "Mapping" and "Media Integration" were significant at $p < .05$; "Aesthetics" was significant at $p < .1$, and "Navigation", "Screen Design" and "Overall Functionality" were significant at $p < .2$.

The trend observed with QUIS indicating that the Indian group to be more positive about the prototype was confirmed here. However, The Americans were the most critical users here, contradicting the trend found with QUIS.

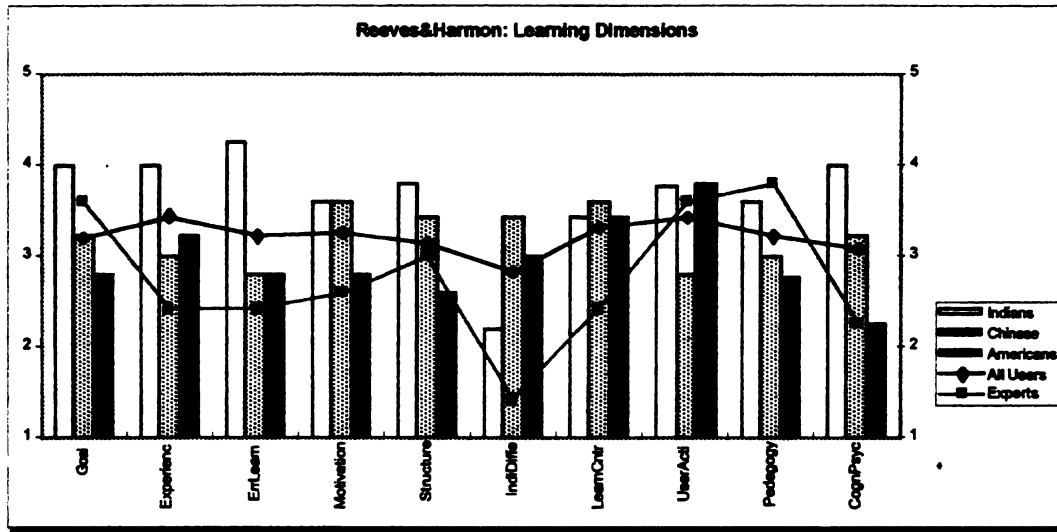


Figure 5.17: Reeves & Harmon answers - Comparison of Users and Experts - Learning Dimensions

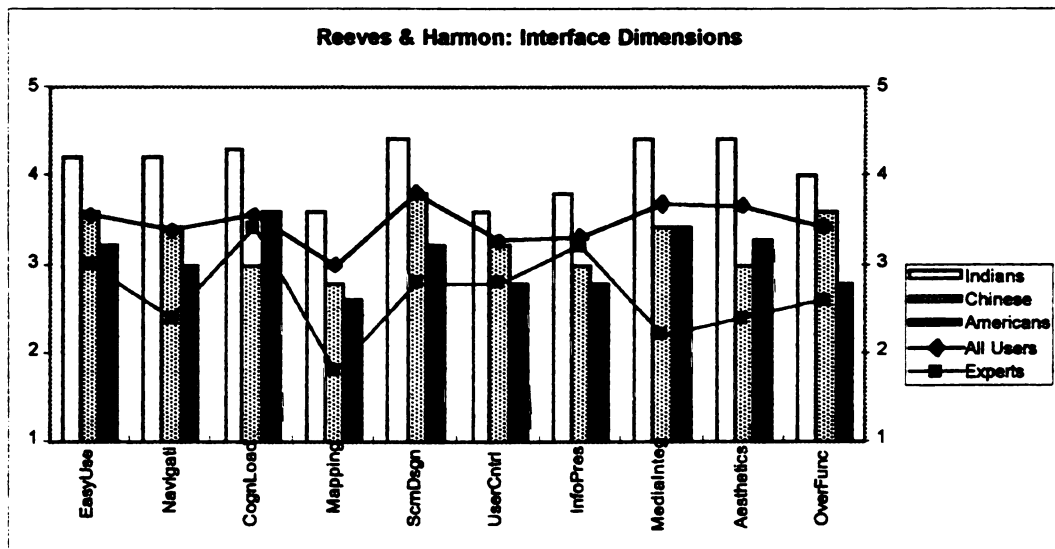


Figure 5.18: Reeves & Harmon answers - Comparison of Users and Experts - Interface Dimensions

5.3.4. Reeves & Harmon: Comparisons between Gender

Table 5.19 presents the computations of means and range for males and females, as well as non-parametric test results (the Mann-Whitney test was performed). Results which obtained statistical significance are indicated. For both dimensions of variables (learning and interface) means were computed for each gender.

	Male		Female		Mann-Whitney
	Range	Mean	Range	Mean	
Learning Dimensions:		3.04		3.65	
Goal Orientation	1 to 5	3.18	2 to 5	3.75	
Experiential Value	2 to 4	3.18	3 to 5	4.00	$p < 0.1$
Use of Error as Learning	1 to 5	3.20	1 to 4	3.25	
Motivation	2 to 4	3.00	3 to 5	4.00	$p < 0.1$
Structure	1 to 5	3.00	2 to 5	3.75	
Acomm. Individual Differ.	1 to 4	2.18	1 to 4	3.00	
Learner Control	1 to 4	3.54	2 to 4	3.25	
User Activity	2 to 5	3.40	3 to 4	3.50	
Pedagogy of Objectives	1 to 4	2.81	4 to 5	4.33	$p < 0.05$
Cognitive Psychology	2 to 4	2.89	3 to 4	3.67	$p < 0.2$
Interface Dimensions:		3.41		3.68	
Easy of Use	2 to 5	3.54	3 to 5	4.00	
Navigation	1 to 5	3.36	3 to 5	4.00	
Cognitive Load	2 to 5	3.60	3 to 4	3.33	
Mapping	1 to 4	3.10	3 to 5	3.50	
Screen Design	1 to 5	3.73	2 to 5	4.00	
User Control	2 to 5	3.27	2 to 4	3.00	
Information Presentation	2 to 5	3.00	1 to 5	3.50	
Media Integration	3 to 5	3.82	2 to 4	3.50	
Aesthetics	2 to 5	3.45	3 to 5	4.00	
Overall Functionality	2 to 4	3.27	3 to 5	4.00	$p < 0.2$
Grand mean		3.23		3.67	

Table 5.19: Reeves & Harmon answers - Gender Comparison

An examination of Table 5.19 confirms the presence of a trend in differences between males and females. This pattern was detected when the values of each dimension were compared, and this trend was more evident on the learning dimensions. On the interface side, the means differences between males and females were smaller. “Pedagogy of Objectives” was significant at $p < .05$; “Experiential Value”

and “Motivation” were significant at $p < .1$; “Cognitive Psychology” and “Overall Functionality” were significant at $p < .2$. Figure 5.20 represents graphically these results.

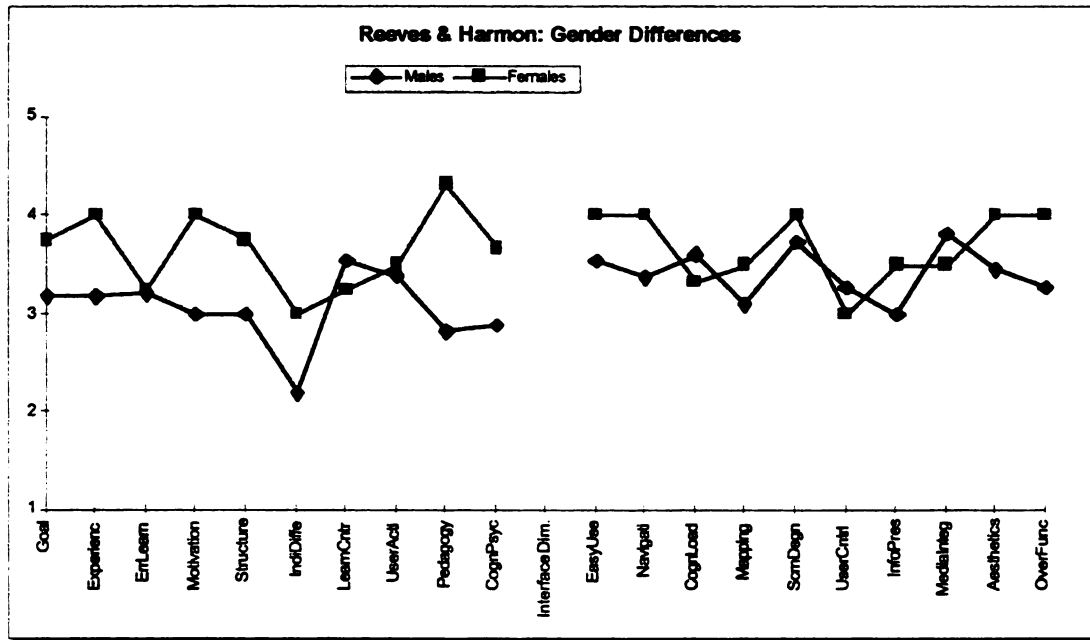


Figure 5.20: Reeves & Harmon answers - Gender Comparison

5.3.5. List of Problems: Comparisons between Ethnic Groups, Users and Experts, and types of problems

The analysis of the list of problems encountered by the participants (Table 5.21) allowed a quantification and classification of problems. An examination of this list indicates four types of problems: Interface problems, Instructional problems, English problems, and Programming problems. This categorization was helpful when trying to identify patterns in terms of which group detected which kinds of problems more frequently. Experts (5 participants) encountered 92 problems, out of a total of 114. In other

words, experts found 51 problems that users did not find, as opposed to 22 problems found only by the users (a total of 15 users).

The mean number of problems encountered by each group: experts 28.4; Americans 12.8; males 10.8; all users combined 9.98; Indians 9.4; females 8.25; and chinese 7.75. Figure 5.22 shows these results graphically. The difference between these groups was statistically significant at $p < .01$.

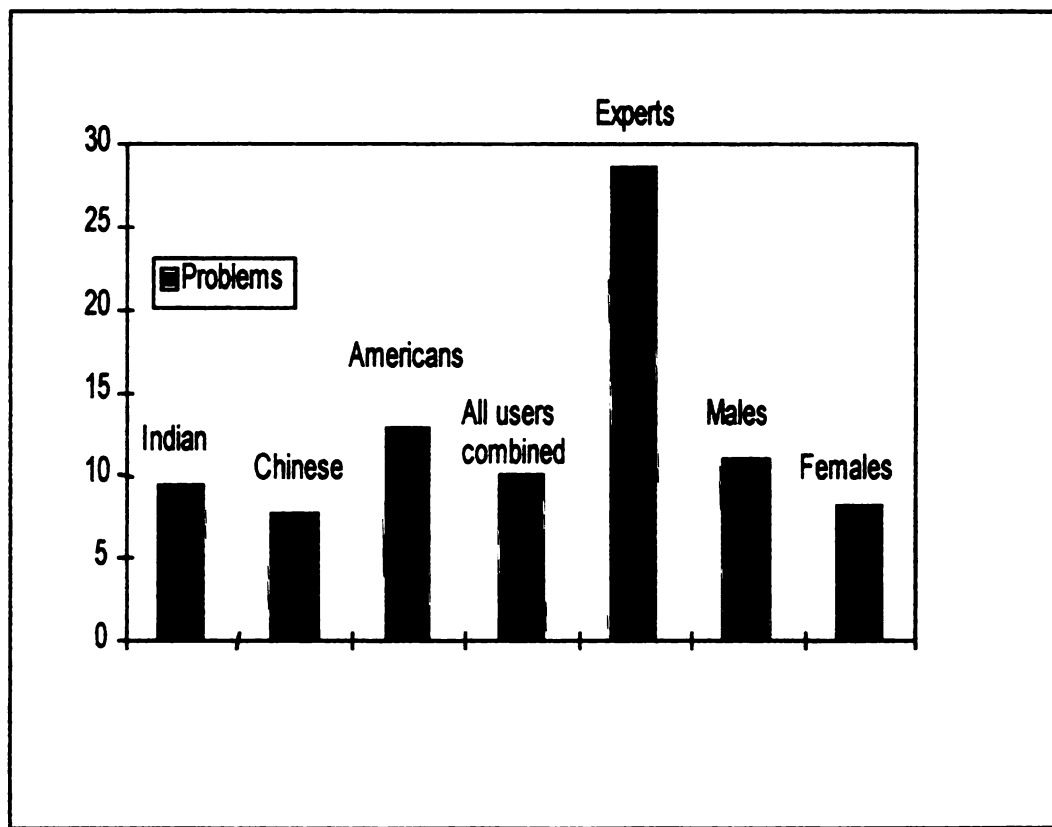


Figure 5.22: Mean number of problems found by subject groups

Location of Problem:	Type of Problem:	Description of Problem:	Users (1 to 16)	Experts (17 to 21)	total	total (%)
<p>Location is defined by the Screen location of each problem found</p> <p>example: Tutorial 2/27 (means that at the second screen of the tutorial a problem was found)</p>	<p>Four kinds of Problems were used in this column:</p> <p>Instructional</p> <p>Interface</p> <p>English</p> <p>Programming</p>	<p>In this space, a brief description each problem encountered is reported</p> <p>example: "Lack of better way for moving around"</p>	<p>The "X"s placed in this area indicate the incidence for each problem.</p> <p>If counted horizontally, it will give the total incidence of one specific problem across all the subjects in the study.</p> <p>If counted vertically, it will give the total number of problems a subject found.</p>	<p>→</p>		
		Totals		↓		

Table 5.21: Explanation of the List of Problems.

The Components are explained above. The actual list is split into two pages for microfilm purposes and is available at Appendix J.

(The original list was one page)

Problems were included in order of coding by the reseacher.

Clearly, the experts were much more efficient in detecting problems. An examination of Table 5.23 and Figure 5.24 allows a closer verification of the kinds of problems detected by users and experts. In this case, it is important to notice that the number of users was three times larger than the number of experts. Among the types of problems, it seems that experts were particularly efficient in encountering interface problems.

Type of problem	Users	Experts	Total by type
Instructional	19	28	32
Interface	31	50	59
English	10	11	18
Programming	3	3	5
Total by group	63	92	114
Obs: 16 users and 5 experts participated in this study.			

Table 5.23: Mean number of problems found by user and experts -
Categorization of types of problems

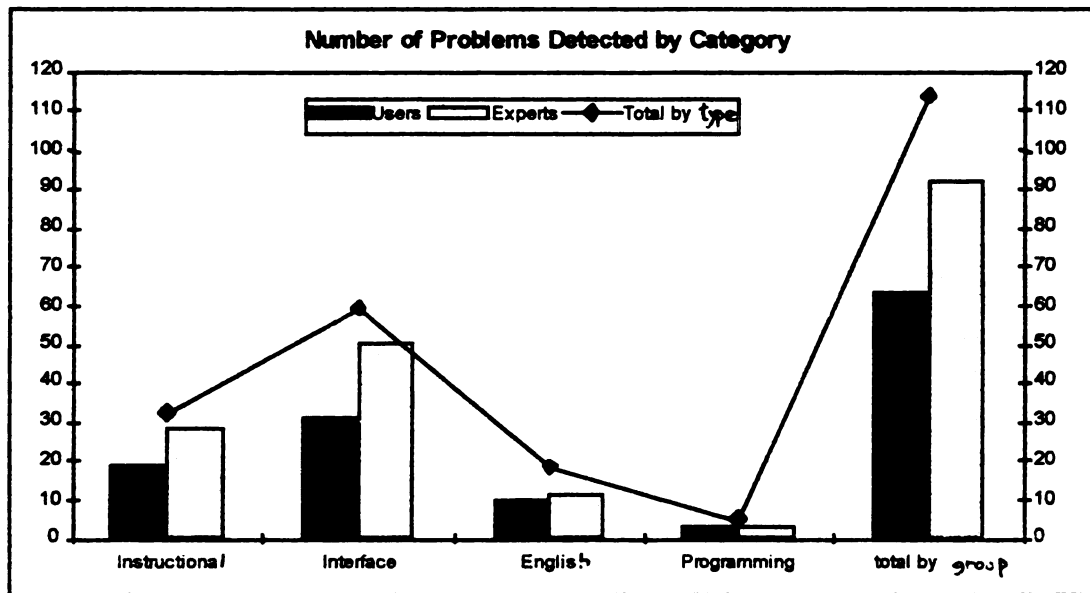


Figure 5.24: Mean number of problems found by user and experts -
Categorization of types of problems

5.3.6. Results of Cluster Analyses

The process of clustering the participants was performed using two methods. In the first method (Joining), a hierarchical tree was generated, which is represented in Figure 5.25. Missing data values were substituted by means of variables.

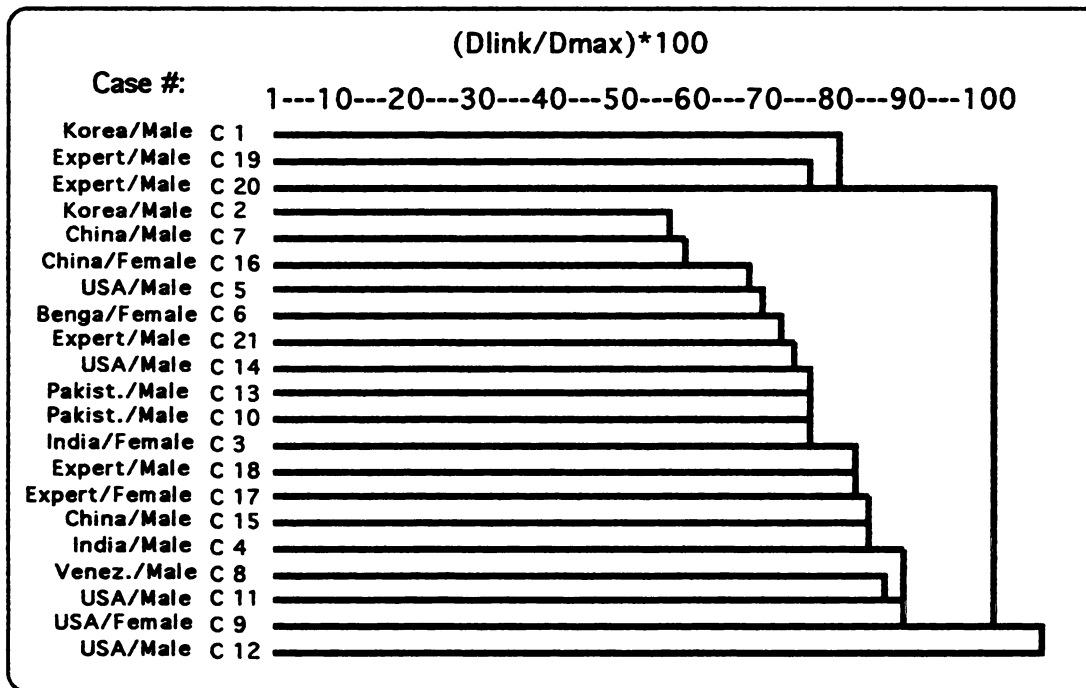


Figure 5.25: Clustering of subjects by Hierarchical Tree

An examination of Figure 5.25 indicates two main clusters. Participants 1, 19, 20, 9 and 12 (one Korean and two experts) form the first cluster. At the bottom of the tree, participants 8, 11, 9 and 12 (all Americans) could be interpreted as a second cluster. A clear distinction between users and experts, or between ethnic groups, with the exception of the American group, could not be detected.

In the second method (k-means) the number of clusters to the statistical software was pre-determined to be two and four groups.

The researcher wanted to test if the participants would cluster into users and experts, or into ethnic groups and experts. Table 5.26 summarizes the results of this clustering process. Since two questionnaires were utilized in this study (QUIS and Reeves & Harmon) The researcher run separate and combined cluster analyses for each possible combination. It shows participants by groups. An examination of the different results for each combination shows an equivalence of results between Reeves & Harmon , QUIS, and QUIS and Reeves & Harmon combined when two clusters were executed. None of the clustering combinations revealed a clear agglutination of the ethnic groups, gender or experts groups. But a trend can be detected in the way Indian/Pakistanis users cluster together, across different combinations.

	2 Clusters:	4 Clusters:
QUIS	Cluster 1: 1,1,1,1,1,2,2,2,2,3,3,3,4,4,4 Cluster 2: 2,3,3,4,4,*	Cluster 1: 1,1,1,1,2,2,3,4 Cluster 2: 2,3,3,4,4 Cluster 3: 1,2,3,3,* Cluster 4: 2,4,4
Reeves & Harmon	Cluster 1: 1,1,1,1,1,2,2,2,2,3,3,3,4,4,4 Cluster 2: 2,3,3,4,4,*	Cluster 1: 1,2,3,3,4,4,* Cluster 2: 1,2,2,2,3 Cluster 3: 2,3,3,4,4 Cluster 4: 1,1,1,4
QUIS + Reeves & Harmon	Cluster 1: 1,1,1,1,1,2,2,2,2,3,3,3,4,4,4 Cluster 2: 2,3,3,4,4,*	Cluster 1: 1,1,2,3,3,3,4,4 Cluster 2: 1,1,1,2,2,4 Cluster 3: 2,3 Cluster 4: 2,3,4,4,*

1= Indians/Pakistanis, 2=Chinese/Koreans, 3= Americans 4=Experts
 * Venezuelan

Table 5.26: Cluster Analysis by Participants and Groups

5.3.7. Results of the Meta-Evaluation by Experts

Figure 5.27 represents the results of the questionnaire for evaluating the instruments and procedures utilized in this study. This questionnaire was answered only by experts, after they finished the evaluation of the prototype itself. They were asked to evaluate the instruments and the data generated by the users with these instruments. This meta-evaluation was intended to verify which instruments experts would rate higher, in the context of developing educational multimedia. An examination of this figure indicates that the multimedia segments were rated the highest, with all five experts rating nine (maximum value of the scale). The list of problems came in second, with a mean value of 8.2; the Demographics Questionnaire mean was 7.8; the QUIS Written Comments mean was 7.0; Ethnic Groups 6.6; Navigational Maps 6.4; Statistical Data at Minitab 5.8; Report of Usage 5.6; QUIS Questionnaire 4.2, and Reeves&Harmon Questionnaire 4.0.

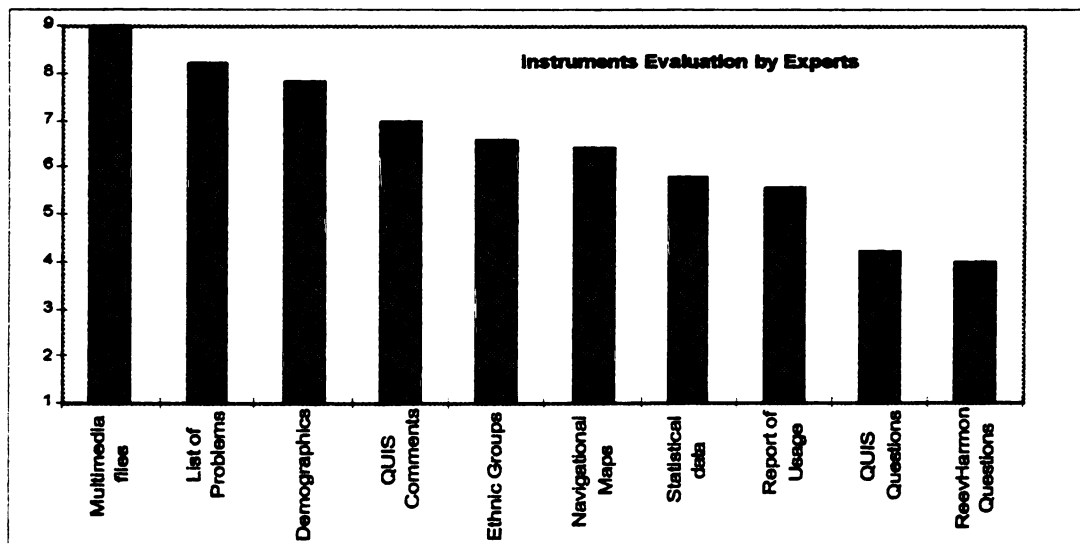


Figure 5.27: Instruments Ratings by Experts

5.4 Qualitative Analysis of Comments from Multimedia Files

A careful examination of the comments collected in the multimedia files generated some preliminary and exploratory results, in terms of the kinds of differences in feedback could exist between males and females, as well as among the 3 ethnic groups included in the study.

In terms of gender differences, the comments showed a trend that females were more focused and detailed when going over the prototype. In general, their comments showed an interest in the instructional aspects of the prototype, but with sufficient understanding of the interface to suggest and indicate valid problems and modifications.

This could mean that females are more attentive and willing of trying to really learn from the prototype. Their comments were generally more frequent and longer. As an example, one of the female subjects was trying so hard to execute the calculation on screen 9 that she ended up finding a bug in the Microsoft Windows Calculator.

In terms of differences between ethnic groups, the comments collected in the multimedia files indicated that the Indian group was more focused on the interface aspects, and generated longer and more rich comments. The Chinese subjects were the most quiet and seemed to be focused mostly on the learning, generating a few really useful comments.

American users seemed to generate a more balanced set of comments, although not as rich as the Indian comments. For the ethnic comparisons, both males and females subjects were included.

The validity of these findings are of relative merit, though, considering that the sample was small and too many confounding variables could be present, such as language difficulty, socioeconomic status, age, the gender of the observer, field of engineering. It is important to note, also, that the fact that one group is less talkative than the others does not necessarily means that they are less useful. The role of the observer is not only to detect the problems verbalized by the subjects, but to detect problems encountered by the subjects. This discrepancy is apparent if one compares the number of verbalized problems of each subject (present in the multimedia files) in relation with the number of detected problems present in the list of problems.

5.5 Chapter Summary

This chapter contained a description of the analysis performed on the data collected during this study and the supplementary meta-evaluation covering instruments and procedures by the experts.

Experts were more efficient in evaluating the prototype. They detected more problems than any other user group. Experts found significantly more interface problems than users.

There were interface problems detected only by users; there were interface problems detected only by experts. This result suggests that a combination of both kinds of participants would be the ideal solution for testing prototypes.

The Multimedia Segments and the List of Problems were the instruments most preferred by the experts. Questionnaires were the least appreciated of the instruments.

Ethnic groups reacted differently in relation to the prototype. The Indian/Pakistani group was the most positive about the prototype. Americans were more critical, according to the Reeves & Harmon Questionnaire; the Chinese were more critical, according to QUIS.

Females were more positive than males when rating the prototype. Males found more problems in the interface.

The process of clustering users according to their responses did not indicate a clear existence of ethnic groups or a clear distinction between users and experts. The high variability of responses in the questionnaires seem to be the cause of this result.

The graphical representations of the questionnaires' answers given by experts were more meaningful for interpretation than the answers given by the users, since there was more contrast between positive and negative aspects of the prototype.

The next chapter contains more details on particular issues discovered during this analysis, as well as a discussion of the methodology utilized in this study.

Chapter 6

Discussion

This chapter discusses particular issues about the data (both quantitative and qualitative), instruments, procedures and the overall methodology used in this study.

The objectives of this study were threefold: a) to examine the differences among experts and users, when evaluating educational prototypes; b) to compare ethnic groups and gender in the process of evaluating educational prototypes, and c) to implement a methodology for evaluating prototypes.

The cognitive diversity of the participants involved in this study challenged the researcher and could have generated more questions than answers. This does not mean that evaluating is unnecessary. Instead, it shows the need for better methodologies and more precise, yet flexible, instruments and procedures.

The verification of differences between groups of users and experts magnifies the issue of the need for collecting more than one perspective when developing and evaluating educational multimedia.

6.1 Cultural identity between participants and observer

One important observation of this study was the realization of the importance of the observer as a crucial element in evaluations of this kind. It is not enough to bring subjects to the laboratory, ask them to try a piece of software, start videotaping and then watch a monitor in a remote room. It is necessary to have the observer

present and interacting with the participants, in order to obtain the maximum amount of quality information.

The quality of the evaluation feedback depends directly on the quality of the relation between participants and the observer at the time of the evaluation. One limitation of the present study was the lack of good communication skills between some of the participants, specially among the Chinese/Korean group. The observer had a good grasp of English (although not his native language) and presented good communications skills, being able to establish a cordial and relaxed rapport with most of the subjects. In some instances, however, not knowing more about the culture and the language of the users proved to be a real barrier. Also, the task of “think-aloud” for subjects in languages other than their native ones was a barrier for some participants.

One possible solution would be to let observers ask questions in English, and allow the comments and answers to be given in the subject’s native language, which could be translated later during the compilation of the videotapes. Ideally speaking, being able to have an observer familiar with the culture and language of the subjects would be the preferred solution.

6.2 Differences among ethnic groups, and experts

Besides the quantitative results reported in the previous chapter, it is relevant to comment on the quality of verbal comments generated by each cultural group. Some users, independent of their cultural background, were shy. This was true even for experts. Differences in subjects’ personalities could become an important

issue to be included in studies of this nature. The researcher was not capable of incorporating this dimension in the present study, but obtaining some variables in this direction (perhaps by way of the demographics questionnaire) is recommended, in terms of exploratory research.

Experts were clearly more comfortable with the evaluation sessions than most of the users. This observation has limited value, though, when one takes into consideration the fact that all experts were colleagues of the researcher, and they all had English as their native language.

The comparison of the depth and quality of comments between different groups indicates that the experts' comments were more complete and useful. Both Indians and Americans, depending on the personality of the user, had very interesting and useful comments. The Chinese/Korean group was the least usable in terms of the quality of their comments.

6.3 Multimedia Files

The multimedia files were incorporated in this study with the intent of providing the experts and designers a qualitative tool that could allow fast and efficient access to problem feedback, in the context of their occurrence. The high acceptance of this instrument among the experts indicates its relevance to future evaluations.

Although technologically possible, digital video technology is not cost-effective when incorporated in studies of this nature. The utilization of audio and screen shots was used as an alternative to digital video, and proved to be adequate for this specific prototype

evaluation. One of the experts suggested the use of recordable videodisk instead of multimedia files, but it was this researcher's intent to see if the available digital multimedia technology would be able to provide reasonable quality, with an affordable cost.

This instrument extends the usefulness of videotaping as a technique for usability testing [Brun-Cottan and Wall 1995]. Videotaping captures and demonstrates to designers user-relevant methods of finding, addressing, and resolving interface problems. Multimedia files take this process one step further by providing random access to the information.

6.4 Simultaneous Analysis of Content, Pedagogy & Interface

The detailed observation of users and experts detected a critical issue, in terms of evaluating educational prototypes. The ideal evaluation participant should be able to handle three different tasks at the same time: a) Evaluate the interface, b) Learn the content, and c) Evaluate the pedagogy. It turns out that very few participants could in fact handle this complex task smoothly. Even among some of the experts, this task overwhelming to handle. This problem is more severe when there is a time limit imposed on participants, which is often the case.

This issue was not too evident to the researcher during the recording of the evaluation sessions, most likely due to the intensity of the interaction between the observer and the participants. Being able to watch carefully the videotapes later on allowed these aspects to become more apparent. Typically, the participant would start the evaluation, and after a few minutes he or she would focus on one

aspect and ignore the others. This effect was less frequent among experts, who seemed more prepared for multitasking, although at different levels, depending on their backgrounds. Some of the experts were very comfortable, due to their educational background in math or science.

The vast majority of target users tried to concentrate on the content aspect of the program, and those who did not get lost in navigational aspects would progress in exploring the prototype. Some users, however, preferred to start exploring or trying to understand the interface aspects and navigation tools, which caused some cognitive overload in the evaluation process. Experts seemed to be better prepared and seemed to have brought with them some previous cognitive strategies in order to deal with these situations.

6.5 Use of Questionnaires in Evaluations

The utilization of two Questionnaires (QUIS and Reeves & Harmon) in this study generated some thoughts about the use of this kind of instrument in evaluations of educational prototypes. The use of questionnaires in usability evaluations are a widespread procedure in the field of Human Computer Interaction. QUIS is a commercial tool, available from the University of Maryland for a fee (200 dollars for universities and 1,000 dollars for the industry).

QUIS could be answered either in paper form or electronic form. There is evidence that using questionnaires on-line presents advantages [Slaughter, Harper and Norman 1994]. On the other hand, there is evidence that short printed questionnaires (up to two pages) can cover the main aspects of evaluations more efficiently [Lewis

1992]. QUIS was developed with the scope of testing software applications, in general. It is not a specific tool for evaluating educational prototypes.

The Reeves & Harmon Questionnaire was created for this study based on recommendations that educational software presents peculiarities that cannot be detected precisely with more general tools. In this sense, this instrument has the potential to be improved and to fill a gap in terms of instruments available to educational designers.

Both questionnaires were rated low by experts, when compared with other instruments in this study. This result suggests that questionnaires are reasonable tools to provide overall feedback about the prototype, but lack the necessary context for the problems detected. It seems that this kind of instrument would be more adequate for a large number of participants, when quantitative summarization is necessary and more in depth analysis is not feasible.

In terms of length, QUIS seemed to be too long, and at times, redundant. The advantage of filling long questionnaires on-line, such as QUIS, is that the users do not know how long they are. The electronic format utilized did not allow participants to verify the number of “pages”, or how much was left to be filled out, which sometimes can be a problem.

The Reeves & Harmon questionnaire was composed of two printed pages, face to face, with a total of 20 items, divided into two categories, learning and interface. Participants did not seem intimidated by this questionnaire, in terms of length. However, some

of the items presented were not meaningful to users and had to be explained in more detail verbally.

6.6 Problems Verbalized Versus Errors Observed

This issue has important implications at the level of replication of the present study. The issue here is the method utilized for detecting the problems during the evaluation sessions. The problems listed in this study were not only the ones verbally expressed by the participants, but also the ones perceived by the observer. This means that different observers would most likely generate different lists, according to their background, prior usability experience and other subjective aspects. The decision to include all problems was based on the fact that this is the way usability evaluation happens. It is unreasonable to ask an observer, during any evaluation of this kind, to include only problems verbalized by the users. This procedure would be considered counterproductive, to say the least. The ability of the observer to detect problems when observing users is part of the process and needs to be included in the evaluation. One solution for this problem would be to use more than one observer, to balance or attenuate bias.

This point explains why there are more problems in the List of Problems than in the Multimedia Files, for each participant. This discrepancy is due to the fact that in many instances, the observer was able to detect a problem, while the subject was busy cognitively verbalizing the occurrence. In other instances, in fact, the subject was unaware of the occurrence of a problem.

6.7 Combination of Qualitative and Quantitative Tools

The use of several qualitative and quantitative instruments combined in this study had the objective of developing a methodology that could execute a triangulation of useful information about the prototype. The need of techniques and methods that integrate qualitative and quantitative tools is imperative within the field of interface design research, where the substantive issues necessitate the integration of both kinds of methods in order to understand complex research problems and applications.

Some instruments were, by nature, qualitative, such as the multimedia files; other instruments were clearly quantitative, such as the questionnaires and the demographic data collection. Some instruments could be categorized as both qualitative and quantitative, such as the List of Problems. This observation is important when analyzing the results of the meta-analysis.

Why did experts value qualitative instruments most highly in this study (see Figure 6.1)? It seems that these instruments were more capable of providing context for the problems. The lack of context of quantitative tools is due to the very nature of these instruments. The process of summarizing the information collected in this study quantitatively discarded the context of the problems' occurrence in exchange for an estimator that could represent the entire population of the study.

As an example of this process, one can look at the item "Going back" from QUIS. This item showed a low mean, which indicates that, in general, the participants had difficulty going back in the prototype. But what if someone asks the following question:

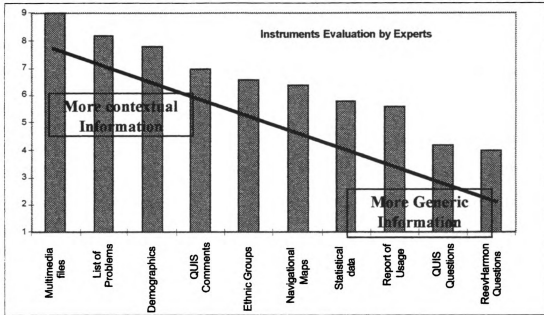


Figure 6.1: Ratings of Instruments by Experts:
The Contextual Aspect .

“Where precisely in the prototype did users encounter difficulties going back?” This person would have to examine the list of problems or the multimedia files to obtain an answer to his/her question. This example serves to demonstrate the usefulness of having both type of data available, when interpreting evaluations of this kind. These results also serves to indicate that, in situations of limited budgets and time, experts would rather have access to qualitative data, when developing evaluation of educational prototypes.

6.8 Use of Statistical Tools in the Evaluation Methodology .

Three types of statistics were used in this study: a) descriptive statistics; b) non-parametric statistics, and c) a cluster analysis. The use of descriptive statistics was straightforward. It was simple to

tabulate the data and obtain descriptive statistics . It was useful to be able to summarize the data into more manageable results, and this process was relatively easy to be executed and interpreted by the experts.

The use of non-parametric statistics to detect differences between groups was a little more demanding, in terms of presenting the results in a usable format. Interpreting the non-parametric results required extra knowledge, research, and a moderate degree of statistical knowledge. An advantage of the use of both descriptive and non-parametric statistical analyses is that they are available in most popular statistical software applications.

The third statistical application, the use of cluster analysis, was more demanding. There are several methods of cluster analysis to choose from; the literature is divergent in some aspects of its use, and interpreting the results of cluster analysis was challenging and laborious. Although it seemed promising to use this kind of statistical analysis, the results were somewhat inconclusive. The development of specific tools and techniques to be used in this context would simplify and broaden its use in human-computer interface design.

6.9 Number and Nature of Problems Encountered

With reference to the number and the nature of problems encountered in this study, the results were more conclusive. The problems were clustered into 4 main categories: Interface, Instructional, English, and Programming. This taxonomy was generated as an attempt to give more depth on the number of problems in relation to each group of participants. This categorization

was complex to execute, and it was of preliminary value, due to the consideration that in some cases, problems could be classified in more than one category.

Ratings about the severity of problems encountered in this study were not implemented, due to the complexity and subjective character of defining this concept [Nielsen, 1994]. One way of measuring severity would be to use the incidence of each problem among the participants available in the List of Problems, and to use these values as an indication of their severity. For example, the lack of control on the audio narration was detected by 60% of the participants. However, if another example is taken (the typo “righ” which also was detected by 60 % of the participants), we can see that this criteria does not provide a consistent method.

Chapter 7

Conclusions

This chapter presents the conclusions that were drawn from this study. In the first section, the conclusions are presented in terms of the specific hypotheses studied. The second section contains a discussion of the methodology used. Finally, directions for future research are described.

This study has attempted to answer a broad range of questions. The results are limited due to the lack of enough subjects to produce statistically significant results. Twenty-one subjects took part in this evaluation. Roughly twice as many subjects may yield a statistical significance. The results, however, do lead to several conclusions.

7.1 Differences in usability feedback among users and experts

It comes as no surprise that experts detected significantly more problems than users. In terms of ratings, experts were more critical in both questionnaires, and presented more variability of answers. This wider variability could simplify the task of interpreting the results of evaluations of this nature. Feedback from experts was more usable in general than the users comments.

The differences found between users and experts can be explained by many factors, including language (English), personality, and background experience with educational software. However, more important than exploring the reasons for these differences is the fact that, despite the differences found in this study, the two groups complement each other, in terms of problem detection. This

becomes apparent if one considers the number of problems detected only by users, a total of 22, in relation to 51 problems encountered only by experts.

Experts demonstrated better strategies in handling the triple task of evaluating interface, pedagogical, and content aspects simultaneously.

7.2 Differences in usability feedback across ethnic groups

The comparison of three distinct ethnic user groups indicated that differences in usability feedback and attitudes exist. The Indian/Pakistanis group of users was consistently more positive and less critical about the prototype tested. The American group presented the highest number of problems detected among all three groups and was the most critical group. The Chinese/Korean group presented an intermediary result, if the number of problems and answers given in questionnaires is considered. In terms of qualitative answers, the Chinese/Korean group presented the least usable results. The American Group was the most usable group, in this regard.

The lack of good interaction between the observer and the Chinese/Korean group in this study limits the validity of the above results. The amount and quality of feedback generated in evaluations of this nature is dependent on the interaction between observers and participants.

Overall, having three ethnic user groups in the study allowed for a wider range of issues and perspectives to be considered that would not have become apparent if only one user group was

targeted. This issue is particularly important when developing educational software for international audiences.

7.3 Gender differences in usability feedback

Comparisons of usability feedback between males and females were indicative of a more positive attitude among women, when testing the prototype. The number of problems detected by males was slightly higher than the number of problems encountered by females. In the qualitative side of the feedback, no apparent differences were detected.

This research question took only users into consideration. It was decided not to include experts in the comparison in order to prevent bias towards males. The small number of experts in the study did not allow a gender comparison among experts (16 users and 5 experts).

One aspect to consider in this research question that could limit the external validity of these results is the fact that the observer was a male. This aspect might have introduced a bias in the interaction between the observer and the participants in either direction, depending on the personality of the observer. This aspect is similar to the issue of cultural identity presented earlier.

7.4 Value of Multimedia Files as Qualitative Tools

There was a consensus among all experts in this research that the multimedia files were the best of all instruments utilized in this evaluation. This result was surprising, considering that these files consisted of audio and screen shots, since digital video was not viable in this study.

The instrument “multimedia file” used in this study was an experimental idea of this researcher that proved to be highly effective and yet relatively simple and cheap to be implemented. Its simplicity relies on the use of commercial multimedia software coupled with low cost video equipment, which combined, gives the social scientist and the interface designer a really powerful way of collecting and examining critical incidents and problem when evaluating instructional technology.

The biggest advantage this instrument offers to designers is the ability for them to see users struggling with their software’s problems without having to spend a great amount of time watching videotapes, or having to deal with real users which a lot of designers dislike. It is the intent of this researcher to further develop this instrument and some of these issues in future research opportunities.

Another instrument well received among the experts was the list of problems. This document, which was rated second best, presented a contextual list of problems, with descriptions and locations of problems, as well as incidences.

In contrast, both the QUIS, and Reeves and Harmon questionnaires were rated low by the experts in relation to other instruments. A plausible explanation for weak ratings could be the lack of context these instruments presented. The usefulness of these tools seemed to be limited according to the opinions of the experts.

The other instruments and procedures were rated in-between these two poles. It seems that the higher the contextualization of the instrument, the higher the ratings they received. This trend indicates a preference towards more qualitative instruments by the experts.

7.5 Evaluation Methodology of Educational Prototypes

The development of cost-effective methodologies for evaluating educational prototypes was a central question to be studied in this dissertation. Some conclusions are presented below in this regard.

7.5.1 Videotaping

The importance of videotaping usability evaluations was confirmed: the generation of multimedia files was dependent on the availability of the videotapes; the generation of a list of problems was dependent on the videotapes; a detailed quantification of problems detected was also generated from the videotapes; the process of videotaping also allows replication of the results, as well as an efficient form of archiving user -interaction for future references.

Videotapes also serve as an essential communications medium in situations where it may be difficult to persuade developers and managers that a certain usability problem is in fact a problem. Seeing a real user struggling with the problem convinces managers and developers [Pauch 1991].

7.5.2. *The interaction of observer and subjects*

The importance of the observer being physically present and the quality of the interaction between observer and participants was evident in this study. For most of the participants, the think-aloud process was not easy and having someone willing to help, prompting then through the evaluation, or just having someone to direct the speech was very important.

The use of critical incidence and think-aloud techniques demonstrated to be efficient ways of obtaining feedback from participants of evaluations, in particular with experts.

One disadvantage of the “think-aloud” method was that it did not lend itself very well to most types of performance measurements. On the contrary, its strength was the wealth of qualitative data that could be collected from a fairly small number of users. Also, the subjects’ comments often contained vivid and explicit quotes that could be used to make the results more readable and memorable.

7.5.3. Number of Subjects

The more subjects included in usability evaluations, the more generalizable the results become. However, problem discovery showed diminishing returns as a function of sample size. Observing four to five participants uncovered between 75 and 85 % of the usability problems, which was a trend found in previous studies [Lewis, 1994]. In most situations, it is not cost efficient or viable to evaluate several people and still preserve the richness and depth or thickness of feedback like that obtained in this study. There is a trade off between the statistical significance and the level of context of the results that needs to be taken into consideration. A recommendation would be to have a minimum of one group of four to six users, and one group of four to six experts. The inclusion of more participants should be dictated by the relative availability of people, time, and equipment.

7.5.4 Questionnaires

Questionnaires are recommended, if they can be kept short and if the terminology used can be familiar to the subjects. The use of existing generic questionnaires for usability evaluations seemed to be of limited value, if one takes the experts’ feedback from this study into consideration. Questionnaires, however, if done

appropriately and combined with other qualitative instruments, can be useful in terms of indicating general strengths and weaknesses when testing educational prototypes.

From the usability perspective, however, questionnaires are indirect methods, since they do not study the user interface itself, but only users' opinions about the user interface.

7.5.5. List of Problems

The generation of a detailed list of usability problems is recommended, as this instrument was rated as one of the most valuable tools by the experts in this study. The availability of a list of this nature can simplify the work of instructional designers when developing educational prototypes. Such a list should indicate the location of the problem, its incidence, and a clear description. This list could be connected via hyperlinks to multimedia files for each observer, giving the designers fast and efficient random access to the problems detected [Nielsen, 1994].

A careful analysis of the List of Problems indicated the existence of four main categories of problems:

a) interface problems; b) instructional problems; c) language problems; and d) programming problems. This preliminary taxonomy of problems could be developed further.

7.5.6. Navigational Maps

Experts showed a relative lack of interest in the generated maps of navigational patterns in this study. Experts did not demonstrate interest in analysing the paths taken and the users, and graded this instrument lower than the other qualitative tools.

This might suggest that there is a need to find different and better instruments and methods for studying navigational issues. It

could also mean that navigational mapping of users in instructional software is not as important as many might think. The fact that one user browses or jumps around more than another might be of relative importance to the efficiency.

7.5.7. Use of Statistical Tools

The utilization of statistical tools is of relative usefulness. Descriptive and non-parametric statistics are used more frequently in usability studies, due to the small number of subjects, and due to the relative ease of execution and interpretation of the results.

The reliability of usability studies could be a problem because of the huge differences between subjects' responses. It is not uncommon to find that the best user is 10 to 15 times faster than the slowest user [Egan 1988]. Usability testing fosters situations where designers have to make decisions on the basis of fairly unreliable data, which is still better than making decisions with no data at all.

The use of cluster analysis is more complex and the results are more difficult to be converted into practical solutions for the developers. Very often, clustering methods are not standardized and may be implemented differently. Also, the problem with doing cluster testing is the difficulty of specifying what the null hypothesis should be. Perhaps a better way of determining clusters would be by trying to examine the validity of various solutions to the data, or by carrying out replication studies [Kirakowski and Corbett 1990].

7.6 Recommendations

In terms of instruments and procedures to include in evaluations, it became apparent in this study that the use of video recording is critical. The importance of use and implementation of

tools such as the multimedia files created in this study became apparent. These tools are particularly recommended for projects with many designers involved, or in situations where designers are not part of the evaluation team. The inclusion of someone not directly involved in the development of the prototype as part of the evaluation team is also recommended, unless an experienced usability specialist is part of the design team.

The use of several kinds of participants and several types of instruments is often not viable because of cost and time limitations. What to do in such circumstances will depend on the context of the evaluation. However, some testing is better than no testing at all. A general recommendation would be the inclusion of a minimum of five users and five experts, whenever possible. Being able to have at least two groups of participants would give the designers the chance to conduct some preliminary comparisons, and would avoid premature generalizations regarding the efficiency of the prototype.

The use of statistical procedures coupled with qualitative instruments is recommended when dealing with methodologies for evaluation of educational interfaces. The use of graphical ways of representing statistical results is also recommended as a way to speed up and facilitate the process of interpretation of these results. In most instances, being able to visualize the results quickly could be the decisive factor for determining the use of the data collected.

The generation of a list of problems is also recommended, including percentages of incidences of problems and location of each problem in the prototype. The use of a pre-questionnaire with

demographic information about the participants completes this minimum configuration.

Also, suggestion for future studies would be to combine the qualities of both instruments into one instrument, making it on-line, relatively short and avoiding complicated terminology and redundant items.

Of lower priority comes the use of questionnaires, as well as navigational maps. These items require the acquisition or creation of specific tools for the task (such as graphic programs, statistical packages, and on-line questionnaires such as QUIS). A possible solution is the use of a simplified printed questionnaire.

7.7 Future Research

The most important application of this study is using it as the basis for future research. This study raised many questions that future studies should attempt to answer.

7.7.1. *Enhance the Methodology*

One important avenue for future research is to find better ways to perform studies similar to this one. The results of this study were based on the use of one prototype only. Other prototypes, in different fields of knowledge could lead to distinct results. The use of more than one prototype could also lead to more complete answers. Therefore, variations of the approach utilized in this study should be explored, or new alternatives attempted, in order to obtain a more complete method. Another important aspect of the methodology used here that could be studied, are the possible connections between quantitative and qualitative instruments.

7.7.2. Qualitative Emphasis

For future research, the qualitative instruments used in this study should be analyzed and developed more in depth. The results of the meta-analysis indicated a preference for qualitative kinds of instruments. Some of the research questions to be studied are: a) Why these instruments were so popular among the experts b) What could be done to improve these instruments, and c) How could these instruments become widely available.

Future research should be directed towards understanding better the issue on the level of contextualization of the instruments used in the present study. This line of research could generate some promising indications for future methodologies in evaluations of educational multimedia. For example, a comparison of different kinds of multimedia files for each subject could measure the contextualization aspect. The generation and comparison of different versions of these multimedia files, such as videodisks, digital videos, or audio-screen files (used in this study) are technological possibilities that need to be understood.

7.7.3 Quantitative Emphasis

The need for future research with a quantitative emphasis is clear. Specifically, research should be performed in order to achieve statistically significant results. One viable approach in this regard is to narrow the scope of the study. Studies could be performed for each area of specific interest. For example, one study could explore the number of problems detected, without having to worry about the generation of context-specific results. Another study could explore attitudinal differences between ethnic groups or gender. By narrowing the scope of the study, and using a moderately higher and

more homogeneous number of subjects, the results could produce much greater confidence levels.

More research and development on the use of cluster analysis in evaluations of this nature is necessary. For example, this kind of study could help in the process of classification usability problems, as well as grouping users.

7.7.4. The Inclusion of Personality

Studies of usability in educational software should take into consideration the participants' personalities. Future studies should try to include personality variables as part of the body of information collected for each participant. There is some preliminary evidence in this study that personality plays a major role in the quality and amount of feedback generated.

7.7.5 Comparison of different types of observers

Future studies should try to compare different types of observers, and usability specialists. For example, a comparison of observers that took part in the development of the prototype, against independent observers could generate important results. The cultural background and gender of the observers are topics to be studied in future studies. Studies that could research the use of more than one observer simultaneously as a way of avoiding bias, or for verification of results could also generate important results.

7.7.6 Use of Navigation Maps

More research and development of navigational maps as tools for visualizing the participants' feedback is necessary. The present study attempted to study this issue only superficially. There are many issues that need to be analyzed in more depth. The navigational instrument generated in this study was primitive, but it

should serve as a starting point for future exploration. The use of spatial modeling for visualization of navigational aspects is recommended.

7.7.7 Use of Questionnaires

The low ratings obtained by the questionnaires in the meta-analysis portion of this study indicate the need for more in depth research in this area. The development of more adequate or context-specific questionnaires, as well a more detailed comparison of existing questionnaires could help answer some of the questions raised here. The issue of printed versus on-line questionnaires is also one topic that needs to be explored.

Interesting findings have been made, and they will serve as the basis for more studies to come. Perhaps this study's biggest contribution is to point the way and declare the need for future research that compares different instruments for usability evaluation of instructional software.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Anderson, R. E. (1987). Females Surpass Males in Computer Problem Solving: Findings from Minnesota Computer Literacy Assessment. Journal of Educational Computing Research, 3, 39-51.
- Bell, J. E. (1990) A Case Study of Ad Hoc Query Interfaces to Databases. Doctoral Dissertation, University of California, Berkeley.
- Benimoff, N. I., & Whitten, W. B. I. (1989). Human Factors Approaches to Prototyping and Evaluating User Interfaces. AT&T Tech Journal, 5(68), 44-45.
- Brun-Collan, F., & Wall, P. (1995). Using Video to Represent the User. Communications of the ACM, 38(5), 61-71.
- Chapanis, A. (1991). Evaluating Usability. In B. Shackel & S. J. Richardson (Eds.), Human Factors for Informatics Usability (pp. pp. 359-395). Cambridge: Cambridge University Press.
- Chin, J. P., Diehl, V., & Norman, K. (1988). Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface. In CHI'88: Human Factors in Computing Systems, (pp. pp. 213-218). New York: Association for Computer Machinery.
- Chin, J. P., Norman, K., & Shneiderman (1987). Subjective Evaluation of CF Pascal Programming Tools. In unpublished.
- Clarke, V. A., & Chambers, S. M. (1989). Gender-based Factors in Computing Enrollments and Achievement: Evidence from a Study of Tertiary Students. Journal of Educational Computing Research, 5(4), 409-429.
- Collins, B. A., & Willians, R. L. (1987). Differences in adolescents' attitudes toward computers and selected school subjects. Journal of Educational Research, 8, 17-27.
- Cronan, T. P., Embry, P. R., & White, S. D. (1989). Identifying Factors that Influence Performance of Non-computing Majors in the Business Computer Information Systems Course. Journal of Research on Computing in Education, Summer, 431-443.

- Dambrot, F. H., & Watkins-Malek, M. A. (1985). Correlates of sex differences in attitudes toward and involvement with computers. Journal of Vocational Behavior, 27, 71-86.
- Day, M. C., & Boyce, S. J. (1993). Human Factors in Human-Computer System Design. In M. Yovits (Eds.), Advances in Computers (pp. pp. 381-430). San Diego: Academic Press.
- Diaper, D. (1990). Simulation: a Stepping-Stone Between Requirements and Design. In M. A. Life, C. S. Narborough-Hall, & W. I. Hamilton (Eds.), Simulation and the User Interface (pp. 59-71). London: Taylor and Francis.
- Eberts, R. E. (1994). User Interface Design. Englewood Cliffs, New Jersey: Prentice-Hall.
- Egan, D. E. (1988). Individual Differences in Human-Computer Interaction. In I. Helander (Eds.), Handbook of Human Computer Interaction (pp. 543-568). Amsterdam: North-Holland.
- Flagg, B. N. (1990). Formative Evaluation for Educational Technologies. Hillsdale: Lawrence, Erlbaum Associates.
- Galdo, E. M. d., Williges, R., Williges, B. H., & Wixon, D. R. (1987). A Critical Incident Evaluation Tool For Software Documentation. In L. Mark, J. Warm, & R. Huston (Eds.), Ergonomics and Human Factors (pp. pp. 253-258). N. York: Springer-Verlag.
- Gardner, D. G., Discenza, R., & Dukes, R. L. (1993). The measurement of Computer Attitudes: An Empirical Comparison of Available Scales. Journal of Educational Computing Research, 9(4), pp. 487-607.
- Gould, J. D., Boies, S. J., & Lewis, C. (1991). Making Usable, Useful, Productivity. Enhancing Computer Applications. Communications of the ACM, 34, no. 1, pp. 74-85.
- Gould, J. D., & Lewis, C. (1985). Designing for usability: Key Principles and What Designers Think. Communications of the ACM, 28, no. 3, pp. 300-311.
- Granstam, I. (1990). Contributions GASAT. In . Jonkoping, Sweden: Jonkoping University.

Gray, D. E., & Black, T. R. (1993). Prototyping of Computer-Based Training Materials. Computers in Education, 22, no.3, pp. 251-256.

Green, A. J., & Gilhooly, K. (1990). Individual Differences and Effective Learning Procedures: The Case of Statistical Computing. International Journal of Man-Machine Studies, 33, 97-119.

Harper, B., & Norman, K. (1993). QUIS: The Questionnaire for User Interaction Satisfaction. In College Park, MD: University of Maryland at College Park.

Hazari, S., & Reaves, R. R. (1994). Student Preferences Toward Microcomputer User Interfaces. Computers in Education, 22, no.3, pp. 225-229.

Igabaria, M. (1990). End-user Computing Effectiveness: A Structural Equation Model. Omega, 18(6), 637-652.

Jeffries, R., & Desurvire, H. (1992). Usability Testing versus Heuristic Evaluation: Was there a Contest? SIGCHI Bulletin, 24, No. 4, pp .39-41.

Karat, C.-M., Campbell, R., & Fiegel, T. (1992). Comparison of Empirical Testing and Walkthrough Methods in User Interface Evaluation. In CHI' 92, (pp. pp. 397-404). Monterey, CA: Association for Computer Machinery.

Kirakowski, J., & Corbett, M. (1990). Effective Methodology for the Study of HCI. Stuttgart: North-Holland.

Laurel, B., & Mountford, J. (1990). The Art of Human-Computer Interface Design. San Francisco: Addison-Wesley Publishing.

Lehner, P. E. (1987). Cognitive Factors in User/Expert-System Interaction. Human Factors, 29, no.1, pp. 97-109.

Lewis, C., & Polson, P. G. (1991). Cognitive Walkthroughs: A Method for Theory-Based Evaluation of User Interfaces (Tutorial No. SIGCHI ACM).

Lewis, J. R. (1992). Psychometric Evaluation of the Post-Study System Usability Questionnaire: the PSSUQ. In Human Factors Society 36th Annual Meeting, (pp. 1259-1263).

Lewis, J. R. (1994). Sample sizes for usability studies: additional considerations (Special Issue: Fatigue). Human Factors, 36, No.2, pp. 368-379.

Lewis, S. (1991). Cluster Analysis as a Technique to Guide Interface Design. International Journal of Man-Machine Studies, 35, 251-265.

Mack, R., & Nielsen, J. (1993). Usability Inspection Methods: Report on a Workshop held at CHI'92. SIGCHI Bulletin, 25, No. 1, pp. 28-33.

McGraw, K. (1993). Conducting User Interface Evaluation (chapter 11). In Designing and Evaluating User Interfaces for Knowledge-Based Systems (pp. pp.171-186). N. York: Ellis Horwood.

McGraw, K. L. (1994, November 1994). Knowledge Acquisition and Interface Design. IEEE Software, p. pp. 90-92.

Miller, J. R., & Jeffries, R. (1992, September 1992). Usability Evaluation: Science of Trade-Offs. IEEE Software, p. pp. 97-102.

Nadin, M. (1988). Interface Design and Evaluation - Semiotic Implications. In R. Hartson & D. Hix (Eds.), Advances in Human-Computer Interaction (pp. pp.45-100). Norwood, New Jersey: Ablex Publishing Corporation.

Nielsen, J. (1992). Finding Usability Problems through Heuristic Evaluation. In CHI'92, (pp. pp. 373-380). Monterey, CA: Association for Computer Machinery.

Nielsen, J. (1993a, November, 1993). Is Usability Engineering Really Worth It? IEEE Software, p. pp. 90-93.

Nielsen, J. (1993b, November 1993). Is Usability Engineering Really Worth It? IEEE Software, p. 90-92.

Nielsen, J., and Landauer, T.K. (1993c). A mathematical model of the finding of usability problems. Proceedings ACM INTERCHI'93 conference, 206-213.

Nielsen, J. (1993d). Usability Engineering. Cambridge, MA: AP Professional.

Nielsen, J., & Levy, J. (1994). Measuring Usability: Preference versus Performance. Communications of the ACM, 37, No. 4, pp. 67-75.

Norman, K. (1994). Navigating the Educational Space with Hypercourseware. Hypermedia, 6, no. 1, pp. 35-60.

Parasuraman, S., & Igabaria, M. (1989). An Examination of Gender Differences in the Determinants of Computer Anxiety and Attitudes Towards Microcomputers among Managers. International Journal of Man Machine Studies, 32, 327-340.

Parasuraman, S., & Igbaria, M. (1990). An Examination of Gender Differences in the Determinants of Computer Anxiety and Attitudes Towards Microcomputers Among Managers. International Journal of Man-Machine Studies, 32, 327-340.

Pausch, R. (1991). Virtual Reality on five dollars a day. In ACM CHI'91, (pp. 265-270). New Orleans:

Pollier, A. (1992). Evaluation d'une Interface Par Des Ergonomes: Diagnostics et Stratégies. Travail Humain, March, pp. 71-95.

Premkumar, G., Ramamurthy, K., & King, W. R. (1993). Computer Supported Instruction and Student Characteristics: An Experimental Study. Journal of Educational Computing Research, 9, No. 3, pp.373-396.

Rauterberg, M. (1993). AMME: an Automatic Mental Model Evaluation to Analyse User Behavior Traced in a Finite, Discrete State Space. Ergonomics, 36, no.11, pp. 1369-1380.

Reeves, T. (1993). Evaluating Technology-Based Learning. In G. M. Piskurich (Eds.), The ASTD handbook of Instructional Technology (pp. pp. 15.1-15.31). N. York: McGraw-Hill.

Reeves, T. C. (1991). Ten Commandments for the Evaluation of Interactive Multimedia in Higher Education. Journal of Computing in Higher Education, 2, no.2, pp. 84-113.

Reeves, T. C., & Harmon, S. W. (1994). Systematic Evaluation Procedures for Interactive Multimedia for Education and Training. In S. Reisman (Eds.), Multimedia Computing: Preparing for the 21st Century (pp. pp. 472-505). Harrisburg, PA: Idea Group Publishing.

Rettig, M. (1992). Interface Design When You Don't Know How. Communications of the ACM, 35, no.1, pp.29-34.

Roske-Hofstrand, R. J. (1989). Video in Applied Cognitive Research for Human-Centered Design. SIGCHI Bulletin, 21(2), 75-77.

Rowley, D. E., & Rhoades, D. G. (1992). The Cognitive Jogthrough: A Fast-Paced User Interface Evaluation Procedure. In CHI' 92, (pp. pp.389-395). Monterey, CA: Association for Computer Machinery.

Salasoo, A. (1991). Initiating Usability Methods with a New Engineering Design Tool. SIGCHI Bulletin, 23, no.1, pp. 68-70.

Shashaani, L. (1993). Gender-based Differences in Attitudes Toward Computers. Computers in Education, 20(2), 169-181.

Shneiderman, B. (1987). Designing the User Interface: Strategies for Effective Human-Computer Interaction. Addison-Wesley Publishing.

Siann, G., Macleod, H., Clissov, P., & Durndell, A. (1990). The effect of computer use on gender differences in attitudes to computers. Computers in Education, 14(2), 183-191.

Slaughter, L., Harper, B., & Norman, K. (1994). Assessing The Equivalence of the Paper and On-line Formats of the QUI5 5.5 No. Laboratory for Automation Psychology, University of Maryland, College Park.

Sutton, R. E. (1991). Equity and computers in the schools: a Decade of research. Review of Education Research, 61, 475-503.

Svendsen, G. B. (1991). The influence of interface style on problem solving. International Journal of Man-Machine Studies, 35, 379-397.

Virzi, R. A. (1992). Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough? Human Factors, 34, n. 4, pp. 457-468.

Wallace, D. F., Norman, K. L., & Plaisant, C. (1988). The American Voice and Robotics "Guardian" System: A Case Study in User Interface Usability Evaluation. In The Human/Computer Interaction Laboratory, University of Maryland.

Wharton, C., Bradford, J., Jeffries, R., & Franzke, M. (1992). Applying Cognitive Walkthroughs to More Complex User Interfaces: Experiences, Issues, and Recommendations. In CHI' 92, . Monterey, CA: Association for Computer Machinery.

Wright, P. C., & Monk, A. F. (1991). A cost-effective evaluation method for use by designers. International Journal of Man-Machine Studies, 35, 891-912.

APPENDICES

APPENDIX A

Consent Form

As a part of a research in the Department of Counseling, Educational Psychology and Special Education at Michigan State University, this experiment is being performed during the Summer of 1995. For this experiment, multimedia designers are being sought who will voluntarily serve as subjects. The purpose of this experiment is to determine the differences between target users and multimedia designers when evaluating a Computer Assisted Instruction Prototype about Telecommunications.

As a subject for this study, you will be expected to spend around three hours learning to use and evaluating the usability of an early teletraffic prototype. You will be videotaped and your voice and screen choices will be the only focus of it. These tapes will be used to record your comments, as well as the interaction between you and the computer. The amount of time to perform various tasks will also be recorded. Names of subjects will not be released in any way.

There is no coercion or demand that you take part in the study. It is solely your personal choice.

Any questions can be directed to the investigator. Pericles Gomes (email 22591mgr@msu.edu or phone 353 5497).

" I understand the above, and I voluntarily choose to serve as a subject for this experiment."

Name

Date

APPENDIX B

Preliminary Questionnaire

Name: _____

Phone: _____

General Computer use:

DOS Windows Macintosh other: _____

Applications: _____

Do you know how to program computers? _____

Do you own a computer: _____

How old were you when you first used a computer? _____

Nationality: _____

Native Language: _____

Age: _____

Gender: _____

Area of Engineering you are mostly interested in: _____

What was your last TOEFL result (if applicable): _____

How long have you been in the USA (if applicable) _____

How would you place yourself in terms of Telecommunication
Engineering:

Not Knowledgeable 1 2 3 4 5 6 7 8 9 Very Knowledgeable

How excited are you about using a computer as a learning
tool:

Not Excited 1 2 3 4 5 6 7 8 9 Very Excited

Have you used Computer Based Instruction before in your
academic life?

Never 1 2 3 4 5 6 7 8 9 or more Times

Preliminary Questionnaire

Name: _____

Phone: _____

Email: _____

Systems you are familiar with::

DOS Windows Macintosh Unix

Age: _____

Gender: _____

Your Background Education :

Area of Multimedia you are most
proficient: _____

Which of the following describes best your activity:

(Circle one or fill in the blank)

Instructional Designer	Software Engineer
Media Designer	Interface Designer
Educational Researcher	Hypermedia Designer
Programmer	Professor
Instructional Researcher	Instructional Technologist

Have you participate in usability evaluations before?

Never 1 2 3 4 5 6 7 8 9 or more Times

APPENDIX C

Description of Experiment for Participants

In this study, your expertise in using this application will help us determine current problems of this prototype interface.

This is a very early prototype. It is not a finished product, by any means. That means your suggestions will be taken into consideration by the creators of this application.

Rather than having your performance being evaluated, you will serve as a means of evaluating the program you use. Hence, we would like to use a technique called "Think aloud" which simply means that you should try say everything you are thinking and tell us any and all impressions you will have of the system: they are VERY valuable for this experiment, and for the designers of this software. Please express your opinions, point out whatever seems confusing to you, and try to explore as much as possible. (Do not feel afraid to err, because there is no right or wrong here, for you).

Your task is to explore the application as much as possible. The content area of this prototype is about teletraffic concepts one would need be to get a job in a telephone company.

APPENDIX D

Some Helpful Definitions:

- **Central office:** Telephone switch
- **Trunks:** Direct lines between two telephone switches
- **Route:** Direct Routing (without Tandem)
- **Tandem:** Alternate Routes
- **Full availability:** All trunks available
- **Sequential Hunting:** the process of searching for a free trunk
- **Erlang:** Unit of telephone traffic (universal)
(1 ERLANG = 1 hour of telephone line usage)

APPENDIX E

.

REEVES & HARMON Questionnaire**Pedagogical Dimensions:**

NA (Not applicable)

I) Goal Orientation:

NA	1	2	3	4	5
	Weak				Strong

II) Experiential Value:

NA	1	2	3	4	5
	Weak				Strong

III) Use of Error as a Learning Tool:

NA	1	2	3	4	5
	Weak				Strong

IV) Motivation (Intrinsic and Extrinsic):

NA	1	2	3	4	5
	Weak				Strong

V) Structure:

NA	1	2	3	4	5
	Weak				Strong

VI) Accommodation of Individual Differences:

NA	1	2	3	4	5
	Weak				Strong

VII) Learner Control:

NA	1	2	3	4	5
	Weak				Strong

VIII) User Activity:

NA	1	2	3	4	5
	Weak				Strong

IX) Pedagogy of objectives:

NA	1	2	3	4	5
	Weak				Strong

X) Cognitive Psychology :

NA	1	2	3	4	5
	Weak				Strong

User Interface Dimensions:

I)	Easy of Use:					
	NA	1	2	3	4	5
		Weak				Strong
II)	Navigation:					
	NA	1	2	3	4	5
		Weak				Strong
III)	Cognitive Load:					
	NA	1	2	3	4	5
		Weak				Strong
IV)	Mapping:					
	NA	1	2	3	4	5
		Weak				Strong
V)	Screen Design:					
	NA	1	2	3	4	5
		Weak				Strong
VI)	User Control:					
	NA	1	2	3	4	5
		Weak				Strong
VII)	Information Presentation:					
	NA	1	2	3	4	5
		Weak				Strong
VIII)	Media Integration:					
	NA	1	2	3	4	5
		Weak				Strong
IX)	Aesthetics:					
	NA	1	2	3	4	5
		Weak				Strong
X)	Overall Functionality:					
	NA	1	2	3	4	5
		Weak				Strong

APPENDIX F

Identification number: _____

PART 1: Type of System to be Rated

1.1 Name of hardware: _____

1.2 Name of software: _____

1.3 How long have you worked on this system?

- | | |
|--|---|
| <input type="checkbox"/> less than 1 hour | <input type="checkbox"/> 6 months to less than 1 year |
| <input type="checkbox"/> 1 hour to less than 1 day | <input type="checkbox"/> 1 year to less than 2 years |
| <input type="checkbox"/> 1 day to less than 1 week | <input type="checkbox"/> 2 years to less than 3 years |
| <input type="checkbox"/> 1 week to less than 1 month | <input type="checkbox"/> 3 years or more |
| <input type="checkbox"/> 1 month to less than 6 months | |

1.4 On the average, how much time do you spend per week on this system?

- | | |
|---|--|
| <input type="checkbox"/> less than one hour | <input type="checkbox"/> 4 to less than 10 hours |
| <input type="checkbox"/> one to less than 4 hours | <input type="checkbox"/> over 10 hours |

PART 2: Past Experience

2.1 How many different types of computer systems (e. g., main frames and personal computers) have you worked with?

- | | |
|-------------------------------|--------------------------------------|
| <input type="checkbox"/> none | <input type="checkbox"/> 3-4 |
| <input type="checkbox"/> 1 | <input type="checkbox"/> 5-6 |
| <input type="checkbox"/> 2 | <input type="checkbox"/> more than 6 |

2.2 Of the following devices, software, and systems, check those that you have personally used and are familiar with:

- | | | |
|--|---|---|
| <input type="checkbox"/> keyboard | <input type="checkbox"/> text editor | <input type="checkbox"/> color monitor |
| <input type="checkbox"/> numeric key pad | <input type="checkbox"/> word processor | <input type="checkbox"/> time-share system |
| <input type="checkbox"/> mouse | <input type="checkbox"/> file manager | <input type="checkbox"/> workstation |
| <input type="checkbox"/> light pen | <input type="checkbox"/> electronic spreadsheet | <input type="checkbox"/> personal computer |
| <input type="checkbox"/> touch screen | <input type="checkbox"/> electronic mail | <input type="checkbox"/> floppy drive |
| <input type="checkbox"/> track ball | <input type="checkbox"/> graphics software | <input type="checkbox"/> hard drive |
| <input type="checkbox"/> joy stick | <input type="checkbox"/> computer games | <input type="checkbox"/> compact disk drive |

PART 3: Overall User Reactions

Please circle the numbers which most appropriately reflect your impressions about using this computer system. Not Applicable = NA.

Overall reactions to the system:

3.1	terrible 1 2 3 4 5 6 7 8 9	wonderful 1 2 3 4 5 6 7 8 9	NA
3.2	frustrating 1 2 3 4 5 6 7 8 9	satisfying 1 2 3 4 5 6 7 8 9	NA
3.3	dull 1 2 3 4 5 6 7 8 9	stimulating 1 2 3 4 5 6 7 8 9	NA
3.4	difficult 1 2 3 4 5 6 7 8 9	easy 1 2 3 4 5 6 7 8 9	NA
3.5	inadequate power 1 2 3 4 5 6 7 8 9	adequate power 1 2 3 4 5 6 7 8 9	NA
3.6	rigid 1 2 3 4 5 6 7 8 9	flexible 1 2 3 4 5 6 7 8 9	NA

Please write down any comments that you have about your overall reactions:

APPENDIX G

Meta-Evaluation Questionnaire

- 1) How useful was the Preliminary Questionnaire?
Very Useful 1 2 3 4 5 6 7 8 9 Not Useful
- 2) How useful was the Reeves/Harmon Questionnaire?
Very Useful 1 2 3 4 5 6 7 8 9 Not Useful
- 3) How useful was the QUIS Questionnaire:
Very Useful 1 2 3 4 5 6 7 8 9 Not Useful
- 4) How useful were the Users Comments from QUIS:
Very Useful 1 2 3 4 5 6 7 8 9 Not Useful
- 5) How useful were the Audio/Screen compilations:
Very Useful 1 2 3 4 5 6 7 8 9 Not Useful
- 6) How useful were the Navigations Maps:
Very Useful 1 2 3 4 5 6 7 8 9 Not Useful
- 7) How useful were the different ethnic groups results
Very Useful 1 2 3 4 5 6 7 8 9 Not Useful
- 8) How useful was the Overall list of errors (by screen location)
Very Useful 1 2 3 4 5 6 7 8 9 Not Useful
- 9) How useful was the Report of Usage?
Very Useful 1 2 3 4 5 6 7 8 9 Not Usefull
- 10) How useful was to have the quantitative data available for statistical analysis?
Very Useful 1 2 3 4 5 6 7 8 9 Not Usefull

Thank you so much for taking the time to participate in this study!

APPENDIX H

Korea 01

Characters on the screen : I hope that program has more helpful loption like initial letter. dont use initail letter, it is hard to memorize at once and more table to use calculate for some formular, and more example. like when I calculate some variable number, computer explains why get solution number and explain what solution number means to us and our society. also on diagam, to better make understandable diagram to use this program, don't use ike A, B, C , and T. use like favorite place and favorite car to get more interesting.

Just put undo option to make useful computer. when I calculate some number, then retype that value to the other space that screen is gone so I have to memorize result even it is big number.

And the option screen is like boring. when I look at it first time, I jsut feel man, it is going to be boring. hence if put atractive stuff on screen then it is more neat program. just don't make them boring and tried. Sometimes I lose position i mean what should i suppose to do for next step. so put previous back option to get back not go back to beginning screen.

Terminology appropriateness : put it more option to more usfull. like more interesting option even it is program for study

Keeps you informed : when I calculate some information , it does not say what I am doing and what is correct result and what that result for.

System Speed : I hope this program has more step by step

Sounds and noises : say information with easy words

Korea 02

Characters on the screen : If the color of the characters were black or some differnt color than blue, it might be better for the users to understand.

Highlighting on screen : The highlight was good. However, if the path was highlighted with the color something darker than yellow, it might be better.

Screen layouts : The layouts were pretty good. But if the small windows of the tables, calculator, and the program stays while the user write the answer, that would be great.

Sequence of screens : It was okay. No specific comments.

Use of terms : The terms were sequencially comming out , so it was very good. Reminds the users in every few monent whether they creally understands the concepts or not.

Terminology appropriateness : If the user can pop up the useful screens (i.e. calculators, programs) anytime they want, and also the definition of the terms, it would be better.

Messages on screen : it was clear and easy to read.

Messages to users : it was clear, but if the screen shows all the equations for the answers was shown in the examples, it might be better.

Keeps you informed : not really. I wanted to go back to the specific screen, but I couldn't. If I can find with the keyword, it would be great.

Learning to operate system : Not that hard, but confusing sometimes. no specific comments.

Exploration of features : If there were some kind of the help screen, it might be easier to explore the features.

Remembering names and commands : I never entered any commands

Tasks performed in a straight-forward manner : no equations for the tasks. there was just numerical answers. I want some straight forward equations for the tasks and the answers.

Help messages : I couldn't find where the help screen was. I never used.

System Speed : No comments. It was fast enough.

System Reliability : There was no warning while I was using the program. However, the program was easy and reliable to use.

Sounds and noises : Was good. But kind of boring. who cares!!! this is not a computer game.

Error correction : NO UNDOs. I tried, but I couldn't.

Experienced and novice users : Good enough!! easy to use.

India 03

Characters on the screen : it's a well processed tool but is frustrating at times and sometimes a bit confusing too. direct access to pages is not available unless scrolling through the whole lot. doesn't intend to give instantaneous solutions to all errors. has a limited menu.

Highlighting on screen : it's very useful in making out the priorities.

Screen layouts : screen layouts have been excellent.

Sequence of screens : sequence has been organised well. but a over-view after certain portions would be very helpful making things look more sequential.

Messages to users : sometimes it gets very confusing as to what is being done wrongly.

Keeps you informed : it doesn't specify exactly which part has to be reviewed more but just brings lot many pages to be reviewed again.

Learning to operate system : initially is slightly confusing as both the arrows keep pointing. after going for a couple of minutes we get the hold of the instructions directing us.

Remembering names and commands : might be easy after going through the whole for once but definitely not at the first attempt.

Pakistan 04

Characters on the screen : I would prefer if there was more material in hard copy (paper). Also, a way to bypass the sound part so the user can have more flexibility in choosing the parts of the program he/she needs to review.

Use of terms : a few grammatical errors / typos were found and reported

Terminology appropriateness : as i said in one of the previous comments, more information in hard copy. would help. also, if more chances to review just the definitions on the screen are provided, it would help.

Messages on screen : a few more messages will help. e.g, in the first screen: click here to continue

USA 05

Characters on the screen : There is a lot of user interactive parts to the program, but many times I found myself just clicking the arrow key to move on. I liked the screens where you had to move the cursor over the diagrams to get the explanation or definition to appear on the screen. That made you more active , made you move around and see the program and also directly gave an explanation for what you were specifically looking at. Instead of just looking at the whole screen and clicking the arrow key to get the definitions up and then moving on.

Highlighting on screen : Highlighting, blinking, and changing the color of certain parts of a diagram or sentence are very useful as long as the color is

pleasant to the eye. Otherwise it will still grab your attention but you will want to move on to avoid looking at the screen instead of studying the screen and learning the main point of concentration.

The reverse video screen is a nice touch also but should not be over done because it could confuse the viewer into thinking the wrong thing.

Screen layouts : The screen layouts were easy to follow, and easy to read. They were about the right size, big enough to be the point of emphasis, yet there was enough room for them to add definitions at the side. I think that maybe other explanations could of been offered. Like more definiton boxes, or maybe specific example boxes could be brought up on screen.

Sequence of screens : The sequence of screens was not necessarily what I expected. It was easy to move forward but not to go back. Many times I found myself lost or back at the beginning when I tried to go back one or two blocks. Then instead of being able to continue where I left off I had to go all the way through the program again. That was not very convenient.

Use of terms : The terms I thought were very consistant and easy to follow. I understood the message that was being offered and it remained consistant throughout the use of the program. I did not understand some of the notation but there was a definition offered or a box would come up and explain it to me in those situations. That made it easy to continue with the program with a knowledge of the terms without having to stop using the program to research the topic or look up the definition of the word.

Terminology appropriateness : The terms I thought for the most part were appropriate. I did not think that some of them necessarily were specific to telecommunications, but they were convienent for the use of the public in general. There were not too many computer terms and the ones there were basic enough that if you were able to use the program you would be able to understand what they meant.

Messages on screen : The messages that appeared on the screen were good and helpful, but they were not always consistant in where they were going to be, the length they were going to be, the number of them there were. i.e. sometiones I thought maybe an extra diagram or message might help convey the idea a little better but that option was not there. The position of the messages (boxes) is best when it appeared directly over or under the word or directly next to the diagram that it related to. moving along at diagonal angles is not convienent.

Messages to users : Directions for correcting errors were not good at all. I couldn't ever figure it out. It would tell me to go review or to try again and my old choices would still be present, yet there was no clear option to erase my previous answers. I tried typing over them, and I tried pressing the arrow keys, but they did not work. Usually I just ended up moving on and not correcting my mistakes.

Keeps you informed : I didn't feel like I could control the amount of feedback and I didn't feel that the computer was letting me know what was up next. I just followed the path of the program and hit the arrow keys. It was easy to maneuver through the program but I didn't always know what to expect next.

Learning to operate system : I felt the program to be easy and quick to learn. The advanced features were well introduced and well explained. The arrow keys made it quick and easy to move from screen to screen and the menu allowed you to pick up in different parts of the program.

Exploration of features : It was easy to see all the options the program had to offer, the menu was simple and if you did happen to "mess" up you could easily find your way back to where you left off. It was encouraged by the program to try problems or to practice some of the concepts you were going over. And if you wanted out of the practice you could continue on and go to the menu where it was possible to continue where you left off. Sometimes you would have to go through a lot of screens but it really was not all that time consuming or difficult.

Remembering names and commands : It was very easy to remember names and how to enter numbers or check boxes as to choices you might want to make. Often times they would remind you numerous times the name of what you were working with. And entering numbers was no problem. One time though I thought that by clicking on the plus button I could put more numbers on the screen and this was not the case. It wasn't hard to figure out I just didn't know that you couldn't do that.

Tasks performed in a straight-forward manner : The steps were well outlined and clear as to what the message was. They followed a logical order and were easy to understand. I think that some could be emphasized better or be bolder to give a stronger message, but for the most part they were very adequate.

Help messages : The amount of help was probably just right but sometimes it left you wondering what to do. then you would continue on and it would make sense. Maybe the messages should inform you that if you keep going on with the program it will make sense. It was easy to get the messages and to get help when you needed it and it was easy to understand the message once it had been brought up on the screen.

System Speed : I am very impressed with the speed of the system. There is nearly no waiting inbetween screens or messages brought up, and when a calculation is done it is immediately displayed in the appropriate spots. response to the mouse is quick and continuous.

System Reliability : The system appears very reliable and stable. It was not disrupted when I tried clicking numerous options on the menu and

immediatly responded to what I did move to. The system did not fail so I do not know how it reacts in that situation.

Sounds and noises : The system is quiet. In the program some of the tones are loud and surprising. I did not expect to here chimes or beeps especially at the loudness that I did. I don't think that there is anything wrong with them, they are just kind of loud.

Error correction : I did not have an easy time correcting my mistakes and I did not have a very good time getting back to where I had left off once I made a mistake. It wasn't that it was hard or all that time consuming. I just didn't like going through the entire menu process again. Errors I had would usually end up skipped instead of showing an example of how to properly do it.

Experienced and novice users : I think that it is meant to be run for someone with more of a background than a new user. It is not difficult so a new user could get on the program and with a fair amount of ease learn the system and eventually, with some quickness master it. but I think that if someone was already familar with the system then they would be able to accomplish a lot more and be more successful with it.

Bangladesh 06

Characters on the screen : it was interesting!

Highlighting on screen : very helpful, but if there is a meaning that we could find clicking on it ,that should be mentioned.

Sequence of screens : the sequence of the screens were not correct. I could not go back to the previosus item that I had just seen or forward to the item I wanted to see and that shoud be afer the one I am on. It did not happen always , but most of the times. So it seeems to be a bit inconsistent.

Messages to users : You could make it a little more clearer in some cases I think. But otherwise they were fine. But sometmes when I was doing the practice exercise and wanted formula I could not get it, instead I had to go to tutorial.

Learning to operate system : it was interesting and with time could be learned quite easily

Remembering names and commands : it wasn't that tough

Tasks performed in a straight-forward manner : yes

Experienced and novice users : needs more information

China 07

Characters on the screen : Based on my evaluation, it's principally a nice system. I figured out, I believe, the main structure of this system in the case that I had not had the user manual. The interface is pretty friend with some minor things need to be improved. Also a suggestion is that if the menu structure can be more friend such that a user who has some experience in using some current popular software would be feel much more comfortable.

Venezuela 08

Characters on the screen : very clear and with good contrasting backgrounds

Sequence of screens : Main menu and navigation is VERY unclear and confusing

Sounds and noises : the beep when you reach the end of a scroll screen is very annoying

Error correction : it is hard to return to a prior screen sometimes

USA 09

Highlighting on screen : The highlighting on the screen helps to explain individual terms. It is a good help method.

Screen layouts : The way the screens were set up was ok, but the amount of information was not easy to use. If you wanted to see more of the data, it was hard to scroll through the data. If you did you would lose your calculations on the calculator, because you would click on sometime else.

Sequence of screens : The sequence of screens was fine. It was logical, but it was hard to go back and look at examples, because you would lose your calculations.

Messages on screen : The messages that appeared on the screen were helpful in the step by step routing sequences. But when doing calculations, you never received any positive feedback. You only knew when your calculations were wrong.

Messages to users : There were no messages to the user for correcting errors. It told you if you were wrong. I assumed if it said nothing, then the calculations were correct. After one try, it seemed that it didn't check your answers anymore.

Keeps you informed : On the problems it had you work out, it only told you when you were wrong. It made it difficult to know if you corrected your errors and fixed the problems.

Learning to operate system : It was overwhelming at first, I did not know exactly what I was supposed to be doing. It would be easier to break the examples up and then combine it into one problem.

Exploration of features : It was hard to go back and look at previous problems, when you did your calculations were lost.

Tasks performed in a straight-forward manner : When doing calculations on the calculator, you could not scroll through the data. It made it hard to do the complete calculations, and it took at least one or two tries to figure that out.

Error correction : I don't think that you could correct your mistakes, it would do it for you and show you what you should have had but you could not compare them. If you tried to correct them on your own, you never knew if you got them right or not. The computer only told you when your answers were wrong.

Pakistan 10

Characters on the screen : For the first time, I am encountered with this sort of experience, it was a very interesting and fascinating one. It was a good experience this software is a good introduction for the new entry in telecommunication services especially in telephony. It could be channeled for the better services in telephone department.

USA 11

Characters on the screen : Was not easy to navigate, using calculator was frustrating as the answers kept disappearing.

Help messages : There was some difficulty in obtaining help with subject matter.

USA 12

Screen layouts : the windows sometimes covered each other up which made reading multiple windows difficult. I shouldn't have to re-arrange windows.

Messages on screen : arrangement was poor because it blocked room for other windows such as calculator or program

Characters on the screen: They are fine and I don't think it needs any adjustment.

India 13

Highlighting on screen: I have a comment about the blinking of characters on the screen. The Blinking Speed is an important factor and

should be adjusted based on the importance, length of the statement and the it could be a good idea to just change the colors which will bring the effect of blinking.

Messages on screen : The positioning sometimes caused hinderance to read the other relavent info on the screen.

Help messages : I did not use help at any time so I really don't know anything about help.

USA 14

Screen layouts : At one point, I was unsure which diagram the comments were referring to.

Sequence of screens : ok, but sometimes difficult to see purpose in structure.

Use of terms : more definitions would have been helpful, or easier access to these at any point in the program

Messages on screen : instructions could sometimes be spaced better in relation to the diagram

Messages to users : messages were clear and understandable

Keeps you informed : feedback was not always predictable

Learning to operate system : fairly easy

Remembering names and commands : I found it diffucult to remember new terminology later in the program and would have liked to have the definitions easy to access at any time.

Tasks performed in a straight-forward manner : Examples of how to perform certain calculations would have been helpful

Error correction : able to rework calculation easily

Experienced and novice users : instructions usually, but not always, easy to understand for the novice

China 15

Characters on the screen : It is OK.

Remembering names and commands : Good.

Sounds and noises : excellant

China 16

Remembering names and commands : not very easy

Expert 1

Sequence of screens : not sure how to get back to earlier screen or exercise, not sure where I was within whole program and how to jump around and sometimes arrow for forward kept me on screen with more information and sometimes same arrow went forward

Messages to users : sometimes not clear on how to get more information of to continue a demonstration to continue

Keeps you informed : one exercise I didn't get feedback on my choice until I selected to move forward, I would like feedback on my choice immediately, audio feedback was somewhat intimidating and feels more judgemental also somewhat unexpected, I would like choice of audio or print feedback

Help messages : help or explanations of content came where I needed it but I didn't bring them up by choice, they came when I selected right arrow to continue - ie when I had already cognitively moved on to next task, I would like to get extra help by choice (ie further explanation or diagrams that did follow when I was wrong or confused but I didn't know they were there until I continued)

Expert 2

Highlighting on screen : bad highlight color

Screen layouts : some pop-up windows were unnecessary

Terminology appropriateness : some inconsistency with regard to which terms need to be defined
would like to have easy access to definitions of any term at any time

Keeps you informed : there was inconsistency in how to interact (e.g. move mouse over capital letter, vs. click on capital letter). Didn't always tell you if you were correct

Exploration of features : set of "features" was very limited (perhaps due to the nature of the tutorial)

Help messages : not much help offered

System Reliability : not enough evidence regarding possible system failures

Expert 3

Screen layouts : Pop-up media appeared covering the media it referred to. I found I had to make mental notes in order to understand the pop graphics.

Terminology appropriateness : Terms were used but not defined.

Keeps you informed : With problems of navigation, you didn't know what the computer is going to show next. Also, I went through a few review questions that shot me back to the beginning of the block, supposedly to make me study the whole block again. It didn't tell me that though.

Learning to operate system : There were no instructions on how to use the system.

Expert 4

Characters on the screen : there was little or no anti-aliasing used around the characters, at times they looked "jaggy" --- also, that font is boring, dull, from a design standpoint.

Different fonts could have been used to display different TYPES of information.

Screen layouts : poor use of space at times, too rigid and also very dull from a design standpoint. boring and not used to display information in an intelligent way

Sequence of screens : The arrows meant something different from every screen, very confusing. Sometimes when I made an incorrect answer, I was asked or forced to review the same information I had seen already. and in other sections if my answer was incorrect I was allowed to proceed. why?

Use of terms : very little implementation of hypertext to define terms. some but not enough

Terminology appropriateness : the formulas and some of their terms were explained, but there was no way to go back and review what you learned.

Messages on screen : often appeared in different places, when those ugly light blue rectangles appeared at the bottom of the screen, they were graphically ugly.

Messages to users : inconsistent

Keeps you informed : I was often lost, thus un-informed.

Learning to operate system : feedback was inconsistent, resulting in my confusion and inability to predict what would happen next.

Exploration of features : no navigation or "help me" featured are used. everything is trial and error, and often when you make an error your result is inconsistent with what happened before. Remembering names and commands : block + and block - were confusing to me

Tasks performed in a straight-forward manner : no, because of disorientation of where you were, you never knew, so you never knew what you had to complete to move on

Help messages : needed "show me how to do this by example" button, or a "help" feature. or at least a voice-over explaining what to do.

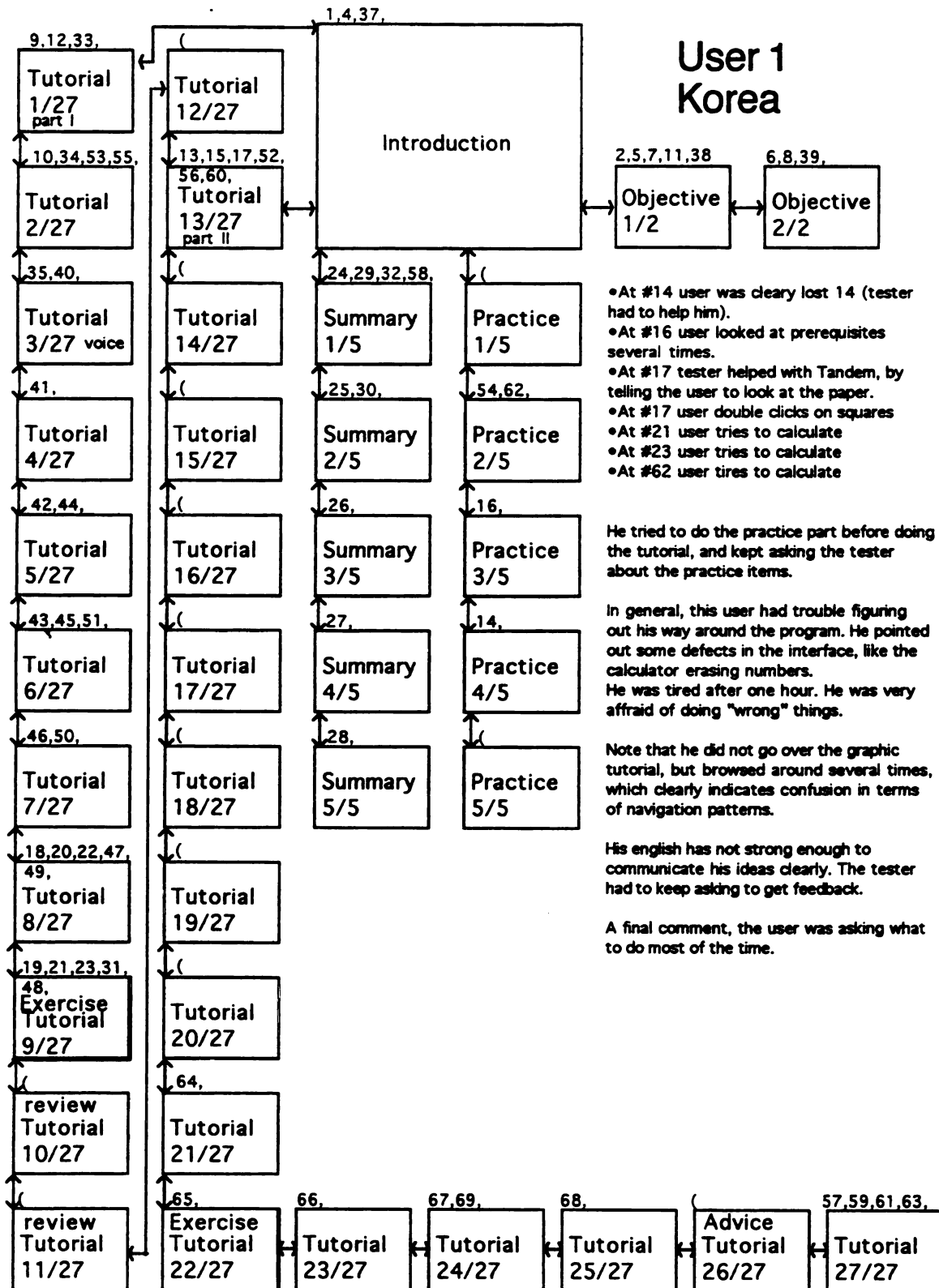
Error correction : when I made a mistake I was removed from figuring out that problem and made to force-learn more operations.

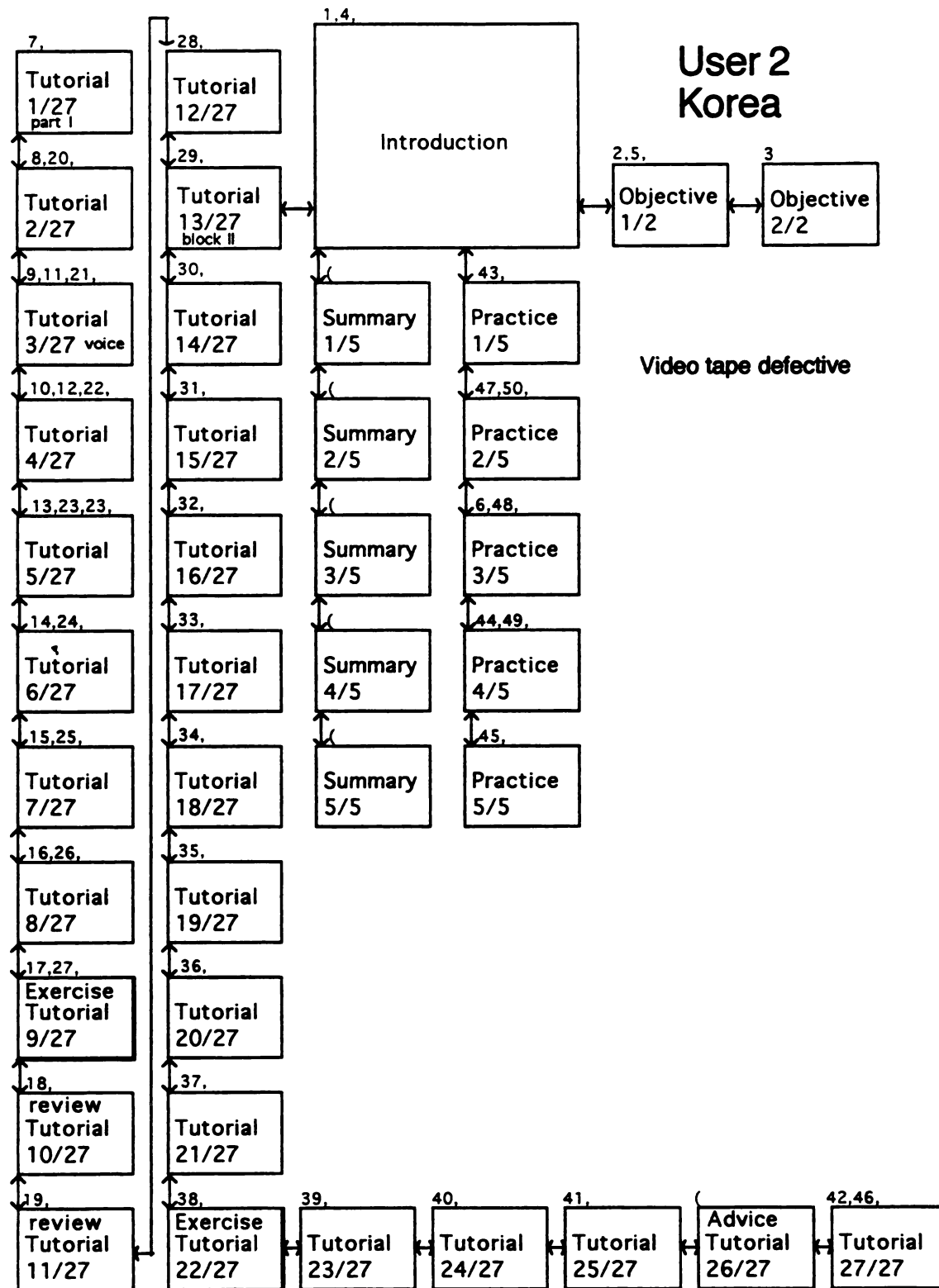
Expert 5

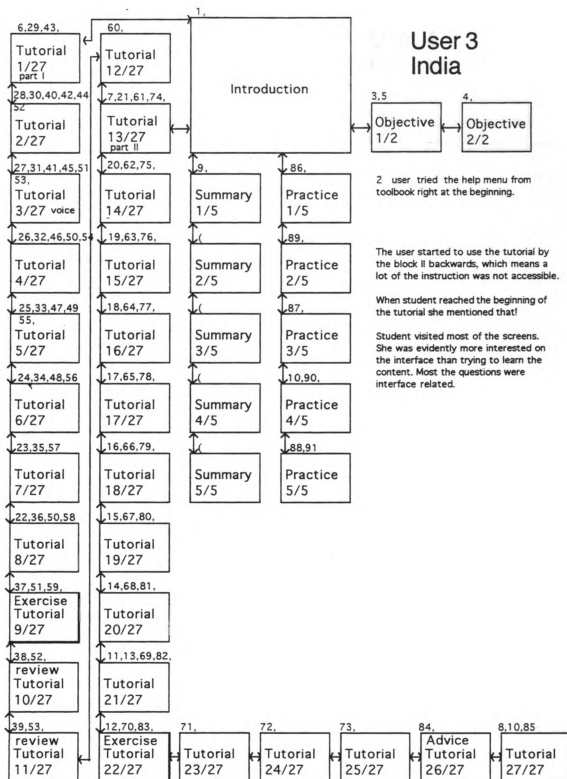
Characters on the screen : a bit boring, could have used some color coding at times

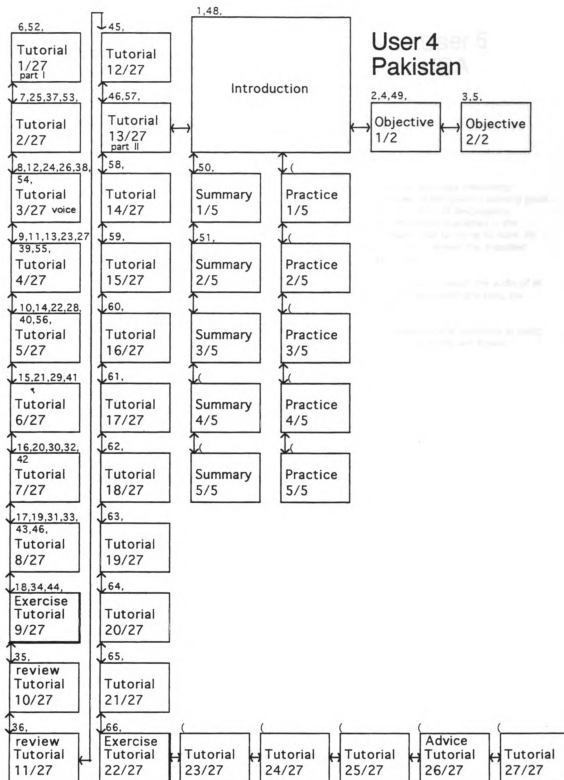
Highlighting on screen : good use with diagrams to correspond with text or buttons

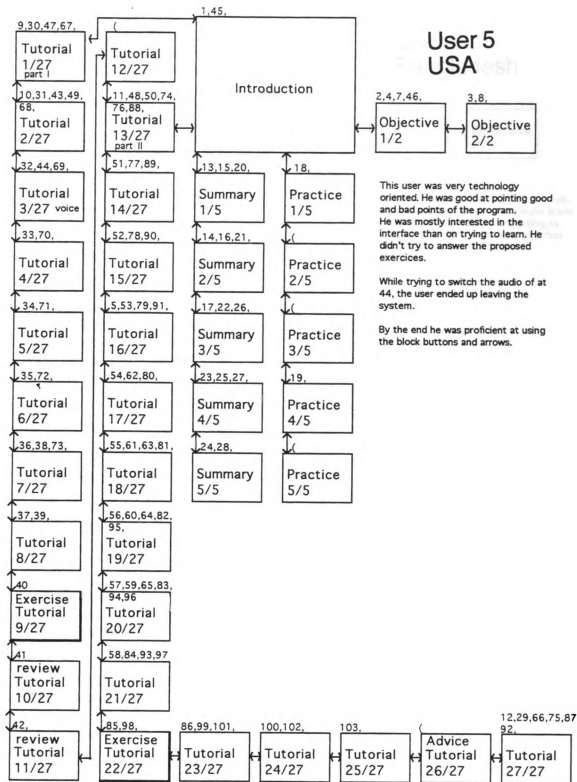
APPENDIX I

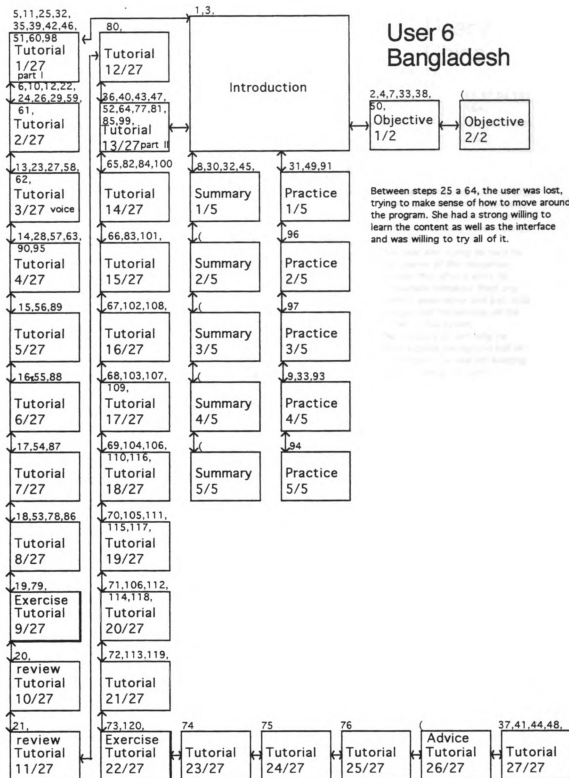


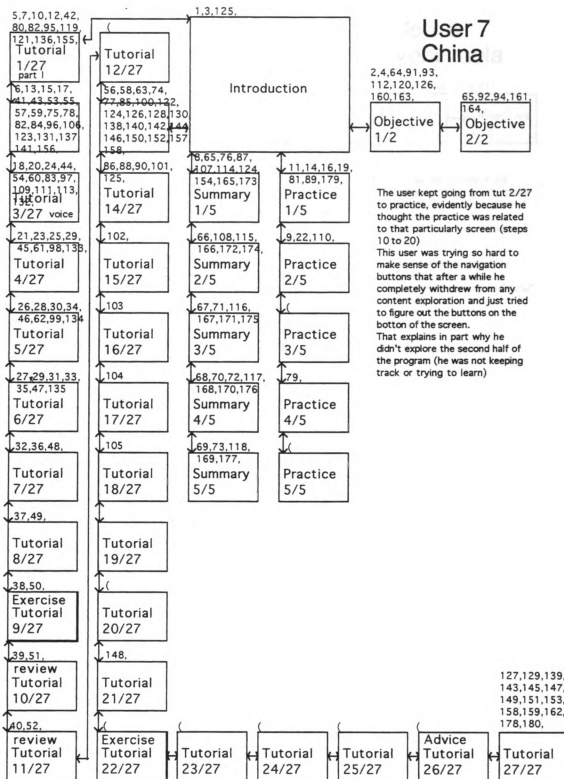


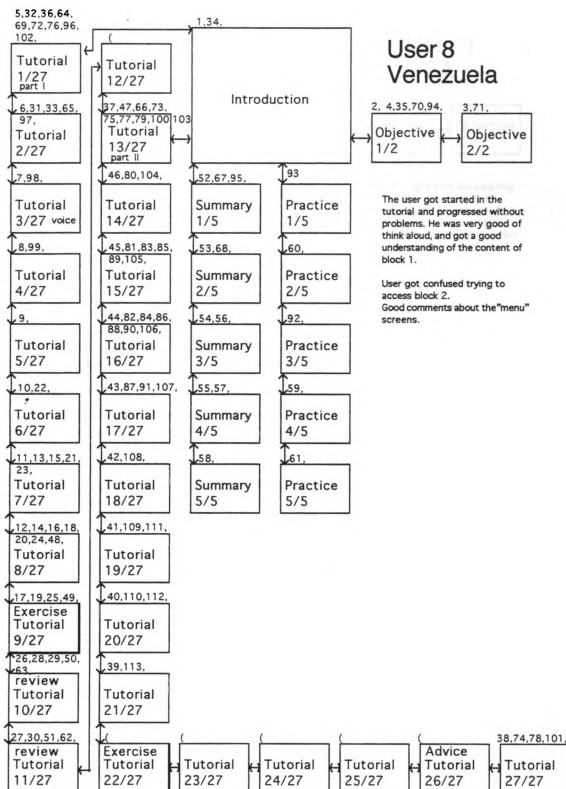


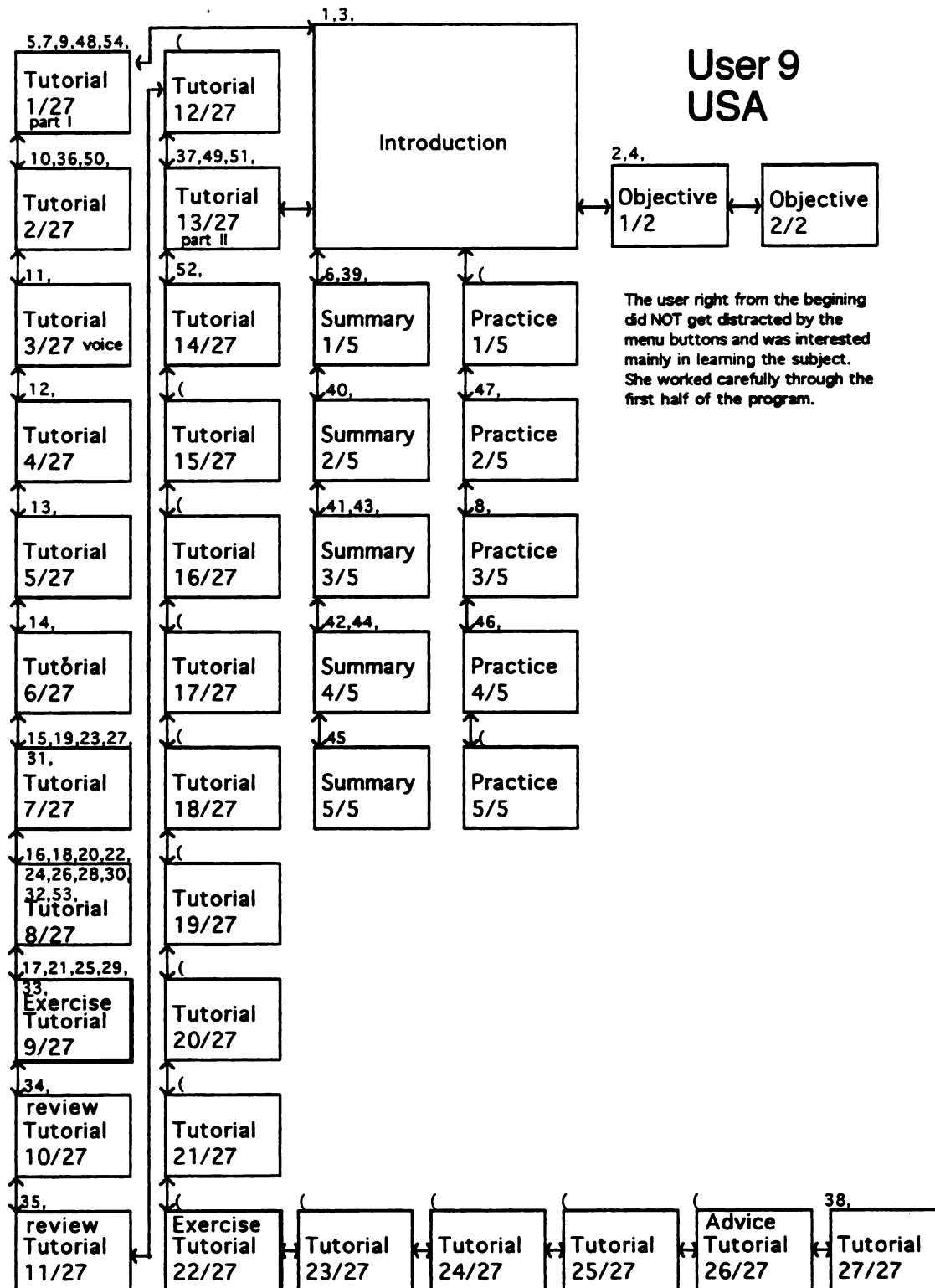


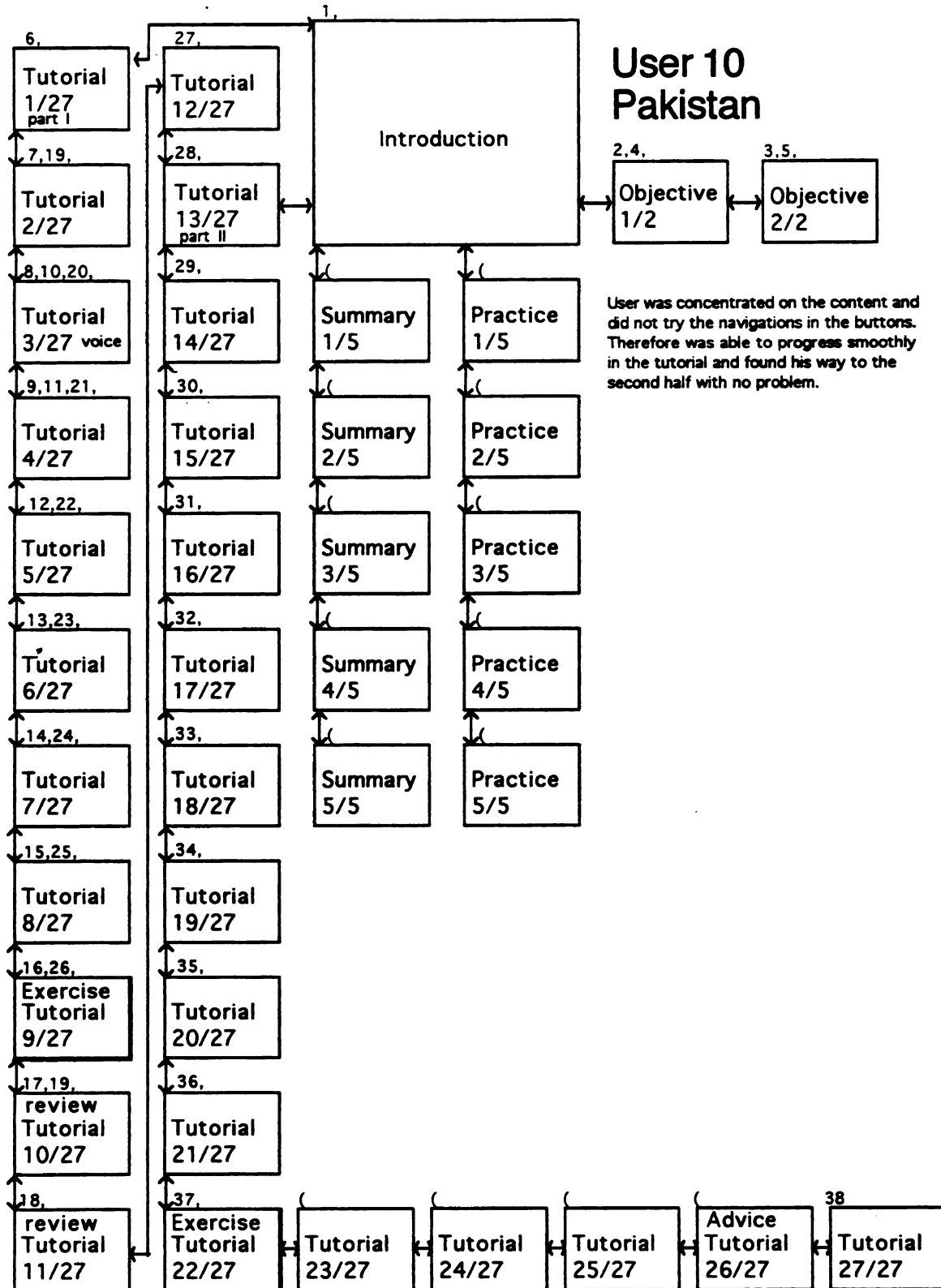


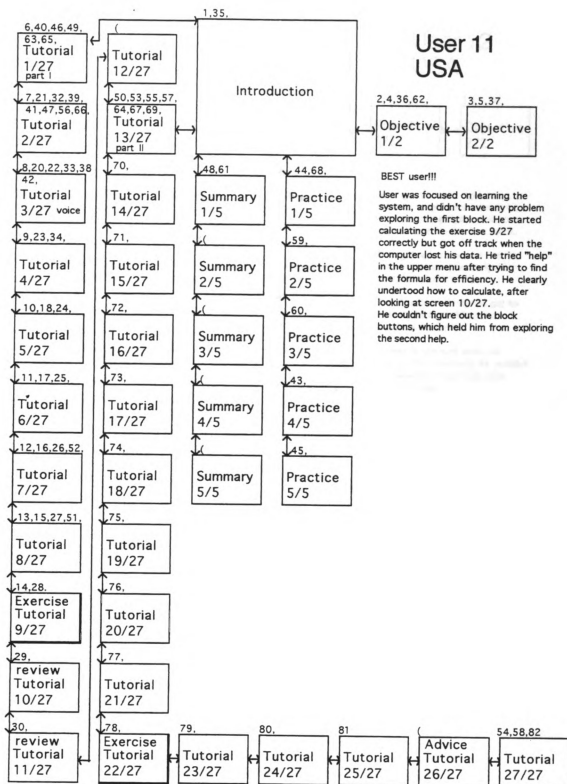


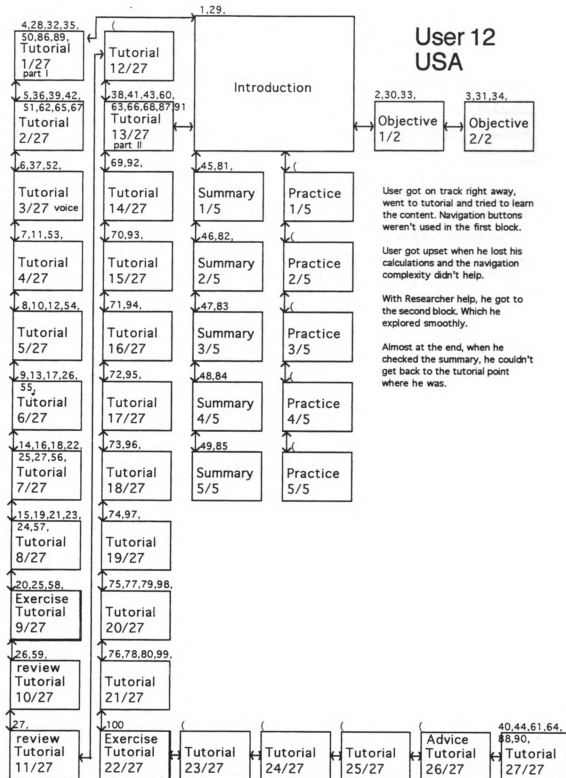


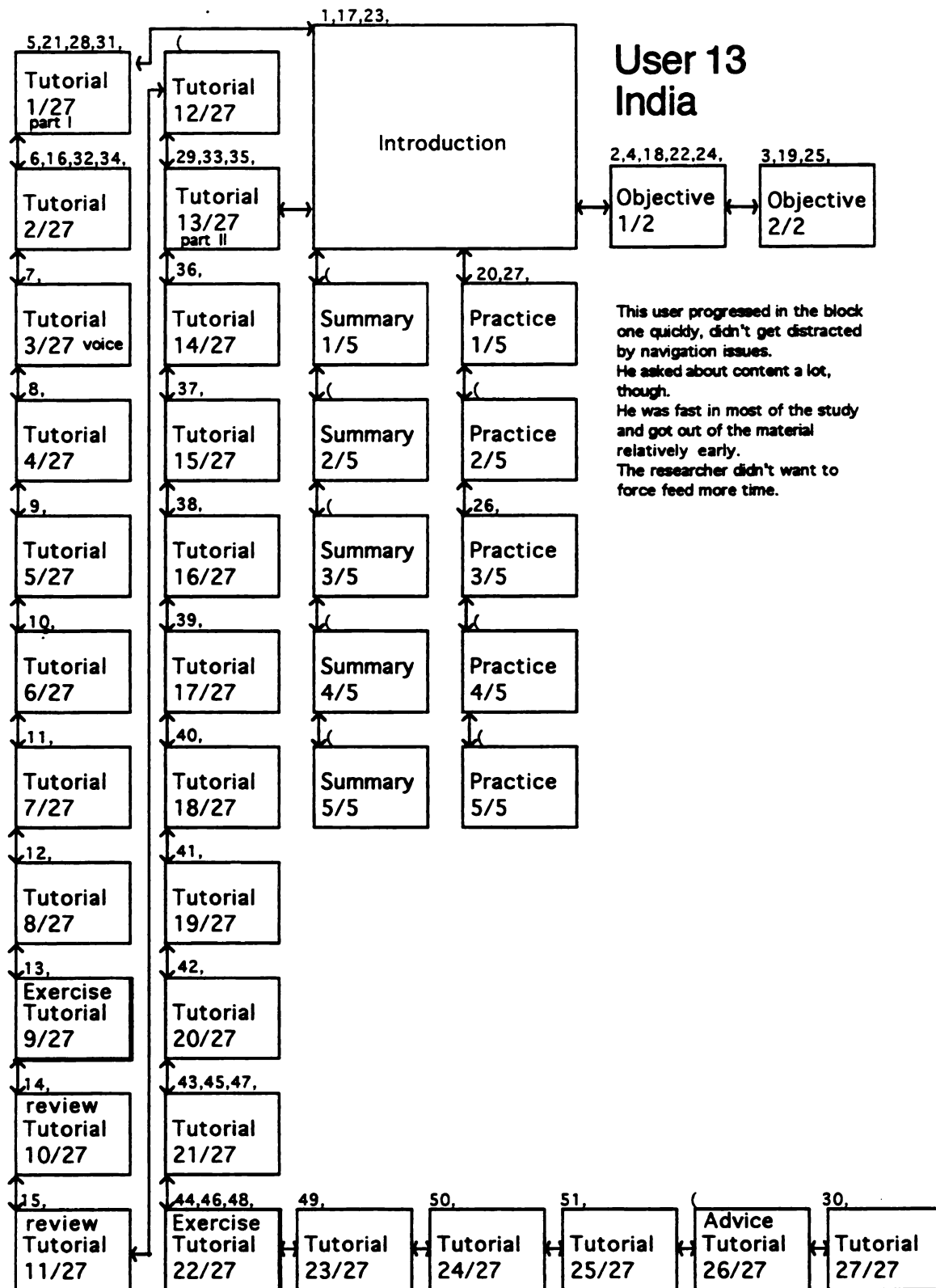


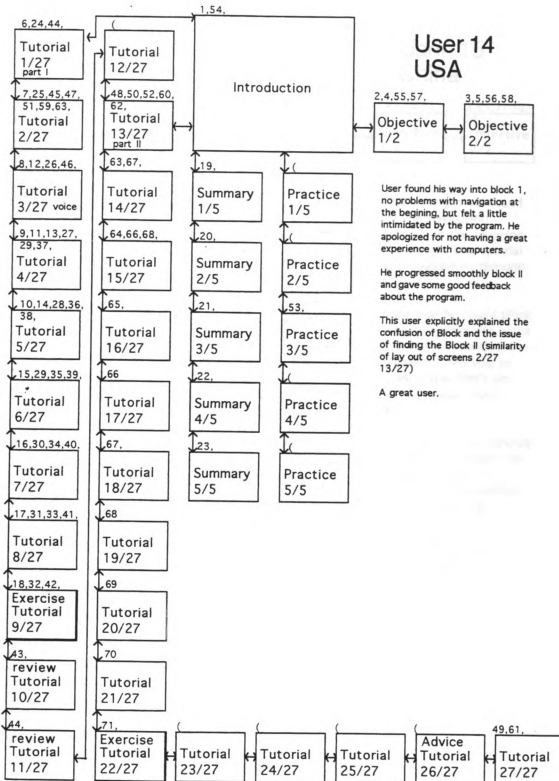


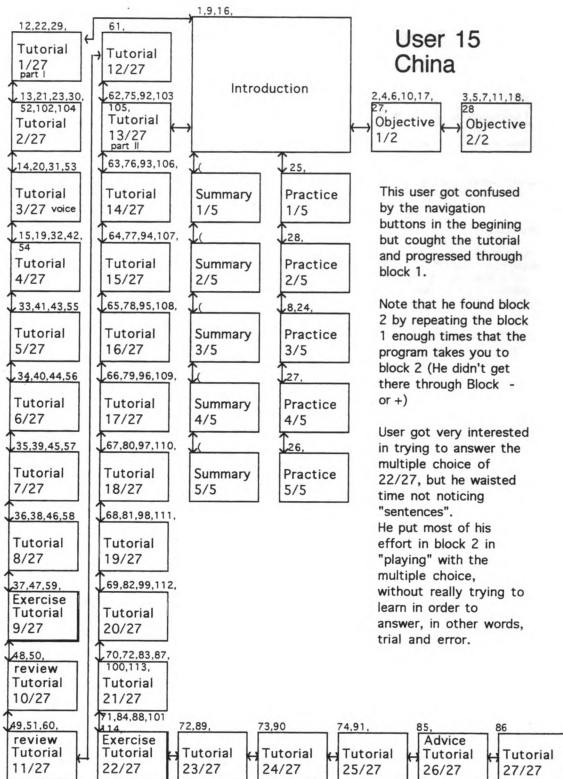


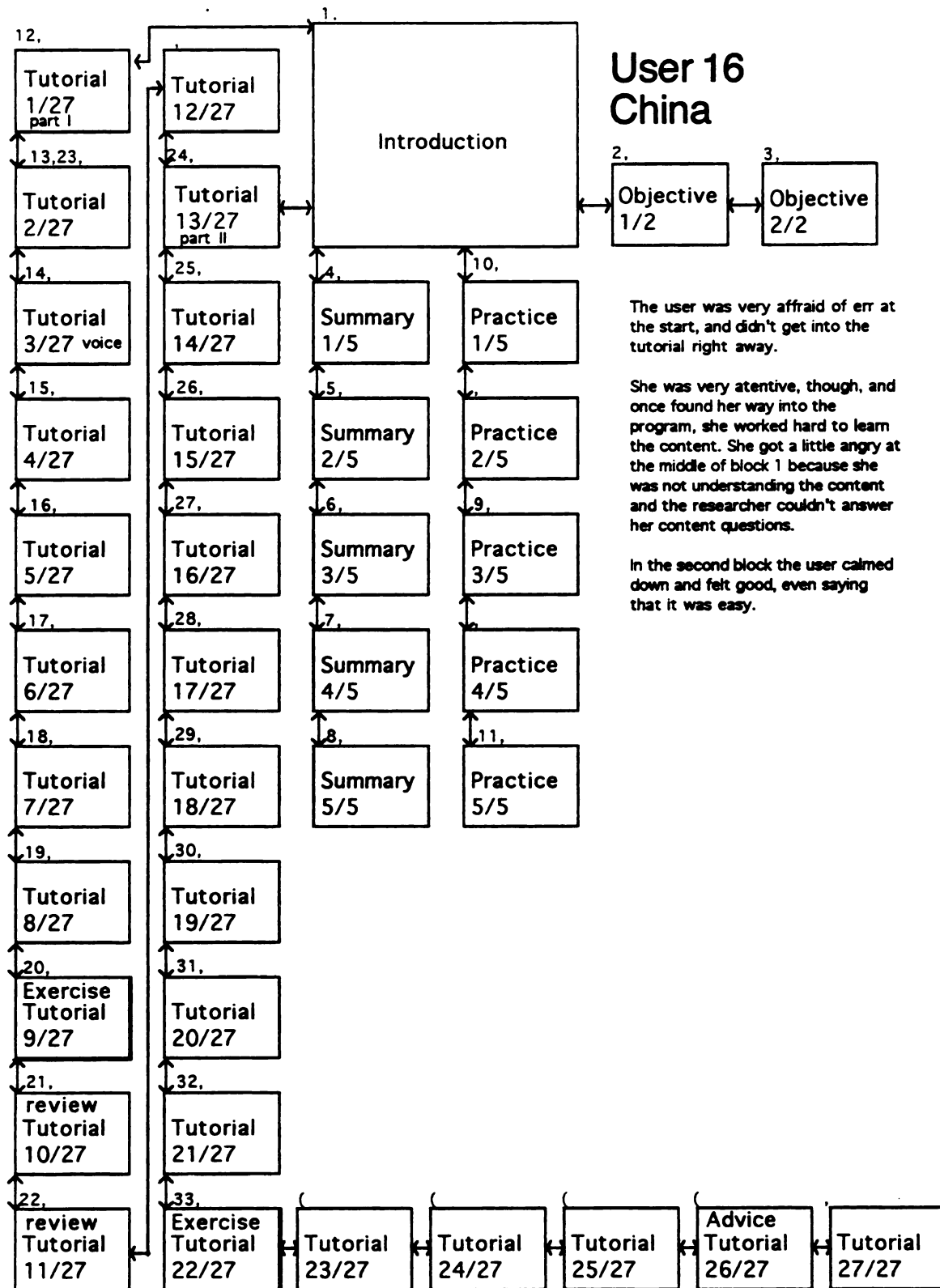


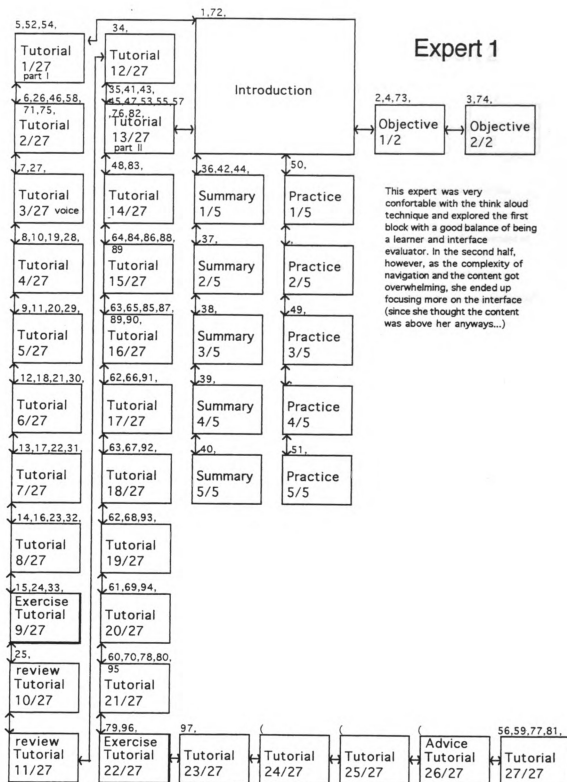


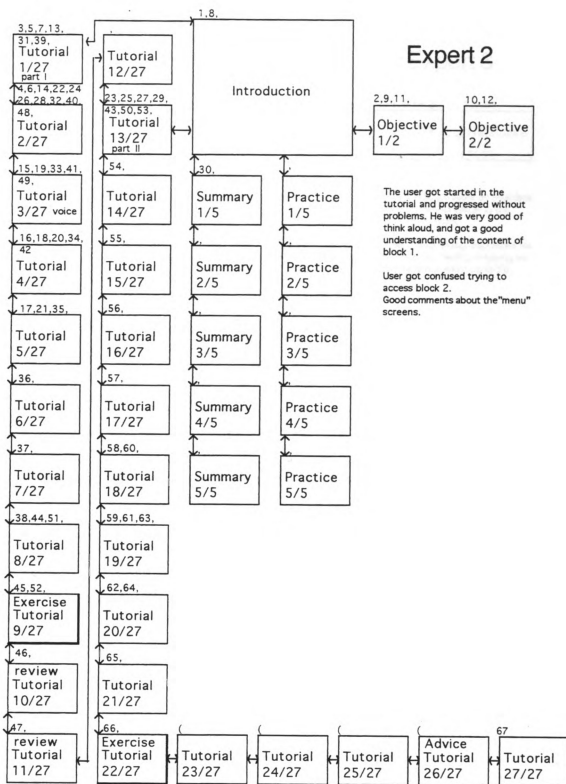


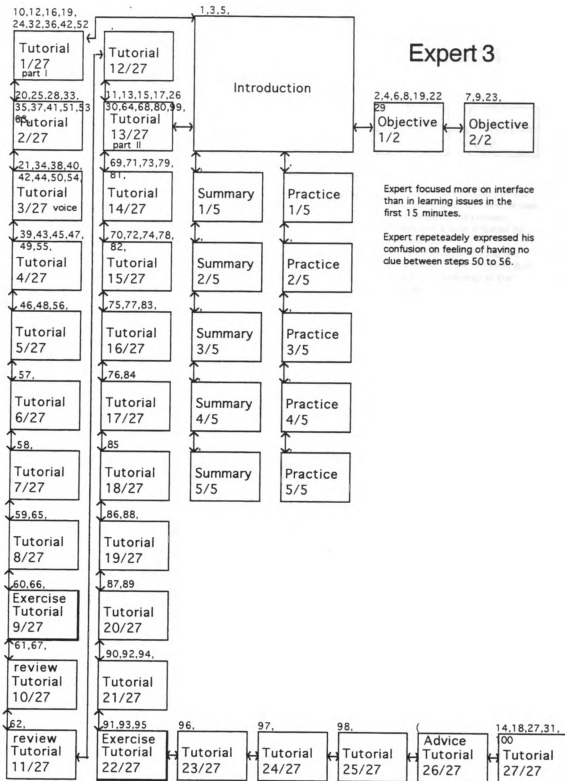


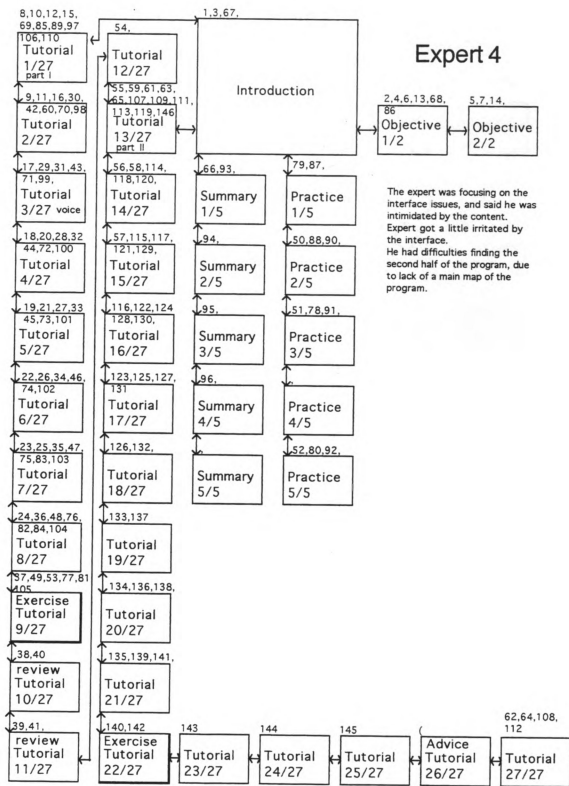


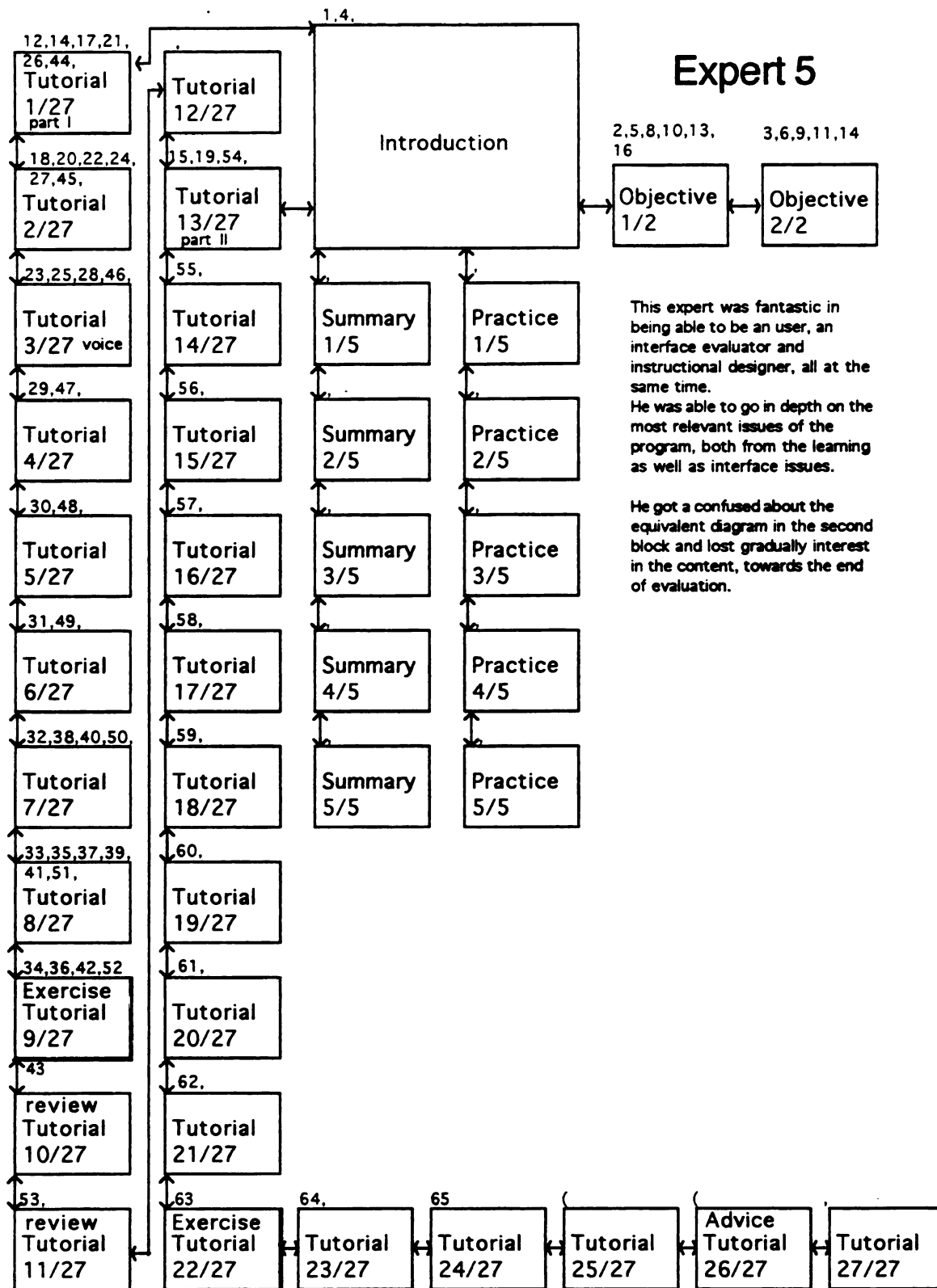












APPENDIX J

LOCATION:	PROBLEM TYPE:	DESCRIPTION OF PROBLEM:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	Total	%
Tutorial 9 of 27	Interface	"See the following review" feedback confusing	x	x		x	x		x					x							x	x		8	40%
Summary 2 of 5	Instructional	Formula confusing	x															x						2	10%
Tutorial 9 of 27	Instructional	not enough information (formula not available)	x	x								x		x						x	x	x	x	8	40%
Tutorial 3 of 27	Interface	check boxes with sound button misuse	x																					1	5%
Tutorial 4 of 27	Instructional	description of formula elements	x												x									4	20%
Tutorial 2 of 27	Interface	block buttons confusing (and - and + not legible)	x	x	x	x	x	x	x	x				x				x	x	x	x	x	x	13	65%
Tutorial 11 of 27	Interface	calculator erases numbers	x		x						x	x	x	x										7	35%
Tutorial 25 of 27	English	"right" typo	x			x	x	x	x	x	x	x	x	x						x	x	x	x	12	60%
Tutorial 17 of 27	English	instructions are misleading		x																				1	5%
Tutorial 17 of 27	Interface	lack of feedback		x																	x			2	10%
Tutorial 15 of 27	English	Capital letter instructions confusing	x		x							x	x							x	x	x	x	7	35%
Tutorial 3 of 27	Interface	sound doesn't turn off	x	x	x	x	x	x				x	x	x						x	x	x	x	12	60%
Tutorial 11 of 27	Programming	typing gives password message										x						x			x	x		5	25%
Tutorial 9 of 27	Interface	lack of automated input and Calculation	x	x						x					x	x						x	x	7	35%
Tutorial 3 of 27	English	Clicking instruction is confusing		x	x																			2	10%
Tutorial 5 of 27	Instructional	Mismatch between terms with program (Eri,Cth)		x															x					2	10%
Tutorial 6 of 27	Instructional	Graphic labeling not good		x							x	x	x	x	x									5	25%
Tutorial 7 of 27	English	Phrasing is confusing (Supposed an offered...)		x																				1	5%
Tutorial 7 of 27	Instructional	Graphic labeling not good		x																				1	5%
Tutorial 7 of 27	Interface	Hyperlinks not too obvious		x	x															x	x	x		4	20%
Tutorial 9 of 27	Interface	table covers problem instructions		x										x		x						x		4	20%
Tutorial 19 of 27	English	"seeded" typo		x										x							x			3	15%
general comment	Interface	Lack of better way for moving around		x	x	x	x									x		x	x	x	x	x		9	45%
Tutorial 19 of 27	English	"definity" typo		x					x					x								x		4	20%
Tutorial 20 of 27	Instructional	Graphic labeling not good		x										x	x	x	x		x	x	x	x		7	35%
Tutorial 20 of 27	English	"send" verb conjugation		x																				1	5%
Tutorial 21 of 27	Instructional	Poisson traffic and Offered traffic should be the same		x																	x			2	10%
Tutorial 22 of 27	English	"show" verb conjugation		x																				1	5%
Objective 1 of 2	Interface	button "resume" misleading		x															x				x	3	15%
Objective 2 of 2	Interface	User thinks it is the mapping of program		x																		x		2	10%
Tutorial 3 of 27	Interface	yellow lines are not good visually		x											x						x	x	x	4	20%
Tutorial 3 of 27	Interface	audio info should be written also		x																	x	x	x	4	20%
Tutorial 22 of 27	Interface	"try again" causes confusion		x	x						x	x	x	x	x	x					x	x	x	11	55%
Tutorial 21 of 27	Programming	Programming bug in the # of trunks + key							x															1	5%
Practice 1 of 5	Interface	in the program, button "calc" was not obvious							x															1	5%
Tutorial 2 of 27	Interface	Practice button being available confuses user							x													x		3	15%
Tutorial 5 of 27	Interface	arrows don't bring you to "last" content (dual purpose)							x	x										x	x	x	x	6	30%
Tutorial 10 of 27	Interface	Your answer box when blank is misleading							x														x	2	10%
Tutorial 2 of 27	Interface	"Block -" Button should be dimmed like the arrow							x					x									x	3	15%
Tutorial 6 of 27	English	"the" is missing (position of trunk)							x															1	5%
Tutorial 11 of 27	Instructional	Congestion probability calculation in "show program"											x	x	x							x		4	20%
Summary 4 of 5	Interface	Instructions location and "capital letters" issue							x															2	10%
Tutorial 4 of 27	Instructional	Meaning of Poisson											x		x	x	x	x	x	x	x	x		7	35%
Summary 5 of 5	Instructional	Screen is useless																				x		2	10%
Practice 3 of 5	Interface	location of instructions is misleading											x		x	x								3	15%
Tutorial 11 of 27	Instructional	Lack of Calculation structure											x										x	2	10%
Tutorial 11 of 27	Interface	"review this topic" misleading											x											1	5%
Tutorial 11 of 27	Programming	MICROSOFT Calculator bug (input0.0003)											x											1	5%
Practice 4 of 5	Interface	"Incorrect" misleading (if user tries again)											x	x	x								x	5	25%
Tutorial 18 of 27	Instructional	Graphic (Arrows to nowhere)												x										1	5%
Tutorial 9 of 27	Instructional	Help menu is not related to course													x							x		2	10%
Tutorial 17 of 27	Interface	Sensitivity of dragging area													x	x			x	x	x	x		6	30%
Tutorial 8 of 27	Instructional	1.136% number (where from?)													x									1	5%
Tutorial 27 of 27	Interface	screen was confused with 212 obj													x									1	5%
Tutorial 18 of 27	Interface	What to do until clicking arrow->														x					x			2	10%
Tutorial 19 of 27	Instructional	Variance formula (hyperlink) confusing														x							x	2	10%
Tutorial 3 of 27	Instructional	"T" in graphic not clear														x						x		2	10%
Tutorial 11 of 27	Interface	Table should be seen entirely															x							1	5%
Tutorial 25 of 27	Interface	"Sentences" Plural (more than one right to choose)															x	x						2	10%
Tutorial 5 of 27	Interface	blinking cursor gone after clicking																x						1	5%
Tutorial 13 of 27	Interface	screen looks like 2/27 and confuses user																	x		x		x	3	15%
Tutorial 13 of 27	Instructional	Graphic explanation not clear (or oversimplified)																			x		x	2	10%
Tutorial 21 of 27	Instructional	lack of feedback																					x	1	5%

List of Problems Found by Users

APPENDIX K

USERID:_____

Time spent on the application _____

Number of Screens visited _____

Explored Block I?	Yes	partially	No
-------------------	-----	-----------	----

Find block II without help?	Yes	partially	No
-----------------------------	-----	-----------	----

Explored Block II?	Yes	partially	No
--------------------	-----	-----------	----

Exercise 9/27 in Block 1:

Resolved	Understood	Tried	not Tried
----------	------------	-------	-----------

Equivalent Graphic 16/27:

Resolved	Understood	Tried	Not Tried
----------	------------	-------	-----------

Equivalent Graphic 17/27:

Resolved	Understood	Tried	Not Tried
----------	------------	-------	-----------

Multiple choice at end of block II (22/27):

Yes	No
-----	----

APPENDIX L

Column	Name	Count	Missing
C1	DOS	16	0
C2	Windows	16	0
C3	Macintos	16	0
C4	Unix	16	0
C5	Geograph	16	0
C6	KnowPrgm	16	0
C7	Owner	16	0
C8	FirstTim	16	1
C9	Age	16	0
C10	Gender	16	0
C11	Eng.Area	16	0
C12	TOEFL	16	7
C13	IN USA	16	0
C14	TelcoKnw	16	0
C15	Excited	16	0
C16	CBTbefor	16	0
C17	Goal	16	0
C18	Experien	16	0
C19	ErrLearn	16	1
C20	Motivati	16	0
C21	Structur	16	0
C22	IndDiffe	16	0
C23	LrnCtrl	16	0
C24	UserActi	16	1
C25	Pedagogy	16	1
C26	CognPsyc	16	4
C27	EasyUse	16	0
C28	Navigati	16	0
C29	CognLoad	16	2
C30	Mapping	16	1
C31	ScrnDsgn	16	0
C32	UsrCntrl	16	0
C33	InfoPres	16	0
C34	MediaInt	16	0
C35	Aesthets	16	1
C36	OverFunc	16	0
C37	Minutes	16	0
C38	ScreenVi	16	0
C39	Scrn/Min	16	0
C40	BlockI	16	0
C41	Find BII	16	0
C42	BlockII	16	0
C43	9/27Calc	16	0
C44	16/27Gra	16	0
C45	17/27Gra	16	0
C46	MultChoi	16	0
C47	Problems	16	1
C48	NumMachi	16	0
C49	Wonderfu	16	0
C50	Satisfac	16	0
C51	Stimulat	16	0
C52	Easy	16	0
C53	Power	16	0
C54	Flexible	16	0

C55	EasyRead	16	0
C56	Sharp	16	0
C57	Fonts	16	0
C58	Hilites	16	0
C59	RevVideo	16	4
C60	Blinking	16	3
C61	Layouts	16	0
C62	AmntInfo	16	0
C63	ArrjInfo	16	0
C64	Sequence	16	0
C65	NextScrn	16	0
C66	Goinback	16	0
C67	TaskBMEs	16	0
C68	TrmOverA	16	0
C69	TASKStrm	16	0
C70	COMPRtrm	16	4
C71	YourWork	16	0
C72	COMPRter	16	2
C73	trmsScrn	16	0
C74	Messages	16	0
C75	PosInstr	16	0
C76	MsgsCfsg	16	0
C77	CommCfsg	16	0
C78	CorrtErr	16	1
C79	CptrInfY	16	1
C80	PrdctRes	16	0
C81	FdbckCtr	16	1
C82	LrnSystem	16	0
C83	LrnStart	16	0
C84	LrnAdvan	16	1
C85	TimeLrng	16	0
C86	T&EEncou	16	0
C87	ExplFeat	16	0
C88	DscvFeat	16	0
C89	RembrN&C	16	0
C90	RmbrRule	16	0
C91	TskManer	16	0
C92	TskSteps	16	0
C93	TskLogic	16	0
C94	StepsSeq	16	0
C95	HelpScrn	16	3
C96	HelpAces	16	3
C97	HelpCont	16	4
C98	HelpAmou	16	4
C99	SysSpeed	16	0
C100	RespOper	16	0
C101	RateInfo	16	1
C102	Reliable	16	1
C103	Dependab	16	1
C104	SystemFa	16	4
C105	WarnsPro	16	6
C106	SysNoisy	16	1
C107	MechNois	16	4
C108	Beep,ton	16	0
C109	CorrMist	16	0

C110	CorrTypo	16	1
C111	UndoOper	16	3
C112	NeedExNo	16	0
C113	Novices	16	1
C114	Experts	16	7

No Constants Used

MICHIGAN STATE UNIV. LIBRARIES



31293014216406