Usability Testing:

Developing Useful and Usable Products

Produced by:

Students in the Master of Technical and Scientific Communication Program

Miami University of Ohio

October 2002

Edited: October 2004

Table of Contents

List of Figures viii

List of Tables ix

Executive Summary 1

Introduction 4

About the Authors 4

Purpose of This Paper 5

Organization of This Paper 6

How to Use This Document 6

Defining Usability 7

An Operational Definition of Usability 8

Introducing Usability Testing 11

Characteristics of Usability Testing 11

Usability Testing and the Design Process 13

Recognizing the Value and Limitations of Usability Testing 15

Recognizing the Value of Usability Testing 15

The Value of Usability Testing to Companies 15

The Value of Usability Testing to Product Developers 17

The Value of Usability Testing to Users 18

Recognizing the Limitations of and Resistance to Usability Testing 18

The Limitations of Usability Testing 19

The Resistance to Usability Testing 20

The Value of Usability Testing Outweighs its Limitations 21

Conducting Iterative Usability Tests 23
The Value of Iterative Testing 23
Goals of Iterative Testing 26
Short-Term Goal 26
Long-Term Goal 26
Classification of Usability Test Types 27
Testing During Pre-design 29
Testing During Early Development Phase 30
Testing Through the Development Phase 31
Testing Before Product Release 32
Resistance to Iterative Testing 32

Conducting Front-end Analyses: Task Analysis and Participatory Design 34
Defining Task Analysis 34
Identifying Tasks and Objectives 35
Identifying Tasks and Sub-tasks 35
Identifying Testing Objectives 36
Documenting Task Analysis Data 37
Collecting Task Analysis Data 38
Defining Participatory Design 40
Communicating Through Participatory Design Workshops 41
Applying Participatory Design Strategies 41
Selecting the Participants 42
Planning the Workshop 42
Recognizing the Limitations of Front-end Analyses 44
Benefiting from Front-end Analyses 44

Preparing a Standard Usability Test Procedure 46
Developing the Test Plan 46
Choosing Tasks for the Usability Test 47
Types of Tasks in a Usability Test 47
Factors to Consider in Selecting and Prioritizing Tasks 48
Presenting Tasks 50
Using Scenarios 50
Field Testing 52
Deciding on the Type of Data to Collect and How to Report It 52
Preparing for the Usability Test 53
Checklist #1: Test Preparation 53
Checklist #2: Pre-Test 54
Checklist #3: Test Monitor Activities 54

Selecting the Right Test Monitor 55
Professionals Who Monitor Usability Testing 55
Qualities of Successful Test Monitors 57
Choosing Test Monitors 58
Implications of Choosing Test Monitors 59
Behaviors of Test Monitors 59

Choosing the Test Environment and Participants 61
Setting Up a Test Environment 61
Considering Test Environment Factors 62
Presenting Test Materials 62
Choosing Test Participants 63
Developing User Profiles 63
Identifying Desirable Characteristics 63
Determining the Most Important Characteristics 64

Determining Experience Level of Participants 64
Selecting the Appropriate Number of Test Participants 66
Using Incentives to Encourage Participation 66
Protecting Participants 67

Ensuring the Rights of Participants 68
Principles and Federal Laws Concerning Human Participants 68
The Belmont Report 68
The Federal Register 69
Federal Regulations for Accessibility 69
Rights of the Participants 70
Ensuring Minimal Risk 70
Ensuring Comprehension 71
Obtaining Informed Consent 71
Garnering Willingness to Participate 72
Ensuring Participant Privacy 72
Responsibilities of the Testing Organization 72
Requiring Non-Disclosure 73
Ensuring Confidentiality 73
Seeking Permission 73
Understanding Participant Expectations 73
Absolute Versus Discretionary Principles—The Gray Area of Usability Testing 74
Risk of Using Incentives 75
Risk of Using Internal Participants 75
Risk of Using an Internal Test Monitor 75

Collecting Data From a Usability Test 77
Monitoring User Activities During the Test 77
Audio and Video Recording Options 78
Data Logging Options 78

Using Data Sheets and Databases 79
Self Reporting 80
Automated Data Logging 81
Data Collection Methods 81
Using Questionnaires to Collect Data 81
Using Thinking-Aloud Methods for Data Collection 82
Using Constructive Interaction 82
Using Retrospective Testing 83
Using Focus Groups for Data Collection 83
Debriefing Test Participants 84

Analyzing and Interpreting Data 85

Defining Data Types 85
Collecting Qualitative Data 86
Collecting Quantitative Data 88
Using Performance Measures 88
Identifying Tasks That Did Not Meet Criteria 89
Identifying User Errors and Difficulties 90
Analyzing Data 90
Conducting a Source of Error Analysis 91
Using Inferential Statistics 91
Counting Frequency 91
Theme-Based Content Analysis 92
Data Collation 92
Theme Definition and Classification 92
Interpreting Data 95
The Importance of Data Analysis and Interpretation 96

Making Recommendations 97
Categories for Ranking Recommendations 98
Possible Formats for Ranking Recommendations 99

Presenting the Results 103
Presenting in Writing 103
Steps for Generating the Written Presentation 104
General Format for the Written Presentation 105
Section One: Executive Summary 105
Section Two: Objectives, Methods, and Procedures 106
Section Three: Results 106
Section Four: Explanation of Findings and Recommendations 106
Section Five: Appendices 106
Common Industry Format from a Written Report 107
Feedback from a Written Report 107
Presenting an Oral Report 108
Format 108
Effective Visual Aids 109
Feedback from an Oral Presentation 111
Presenting a Videotape 111
Responsibilities and Rewards of Presenting Results 111

Conclusion 113
A Message From the Authors 114

Glossary 115

Appendix A: Outline of the Common Industry Format (CIF) for Usability Reports 121

Title Page 121

Executive Summary 121

Product Description 121

Test Objectives 122

Participants 122

Context of Product Use in the Test 122

Design of the Test 123

Usability Metrics 123

Data Analysis 123

Presentation of the Results 124

Satisfaction Results 124

Appendices 124

Sources Consulted 125

List of Figures

Figure 1. Classification of usability tests. 29

Figure 2. Sample user profile for usability test participants. 65

Figure 3. Sample paper data sheet. 80

List of Tables

Table 1. Prioritizing task lists. 50

Table 2. Types of professionals qualified for usability testing. 56

Table 3. Qualities of successful test monitors. 57

Table 4. Appropriate behaviors for test monitors. 60

Table 5. Sample TBCA table of participant behaviors 94

Table 6. Sample recommendation format. Explanations (rationale) follow each recommendation. 99

Table 7. Sample recommendation ranking format. Table of findings, recommendations, and severity (indicated verbally). 100

Table 8. Sample recommendation ranking format. Table of findings, recommendations, and severity (indicated verbally). 100

Table 9. Sample recommendation ranking format. Table lists problem number (for easy reference), scope, severity level, and frequency. 101

Table 10. Sample recommendation ranking format. Table of tasks participants were asked to perform, the source of participant error, severity of problem, recommendations, and business rationale. 102

Executive Summary

Usability is the characteristic of a product that describes its ease of use. Usability testing is a method of gathering feedback about a product from potential users or people who have the same characteristics as potential users, to help product developers build products that more effectively satisfy the needs and desires of users. Creating more usable products through this process is often referred to as user-centered design.

Usability is important because consumer products become increasingly sophisticated, users often select products based on ease of use. Consequently, product developers who conduct usability testing to build ease-of-use into their products often gain a competitive advantage.

This paper provides an introduction to the concepts of usability and usability testing. It provides the background information necessary to understand usability testing and the role it plays in developing user-centered products. It describes the steps for planning and administering usability tests and describes strategies for analyzing the data obtained from them. It also suggests ways to present usability test results to a variety of people involved in different aspects of product development and at different levels in an organization to ensure that they will be understood and considered in developing the next release of the product.

This paper is presented in sections. The following is a brief description of the topic addressed in each section.

- "Defining Usability" begins with a simple definition of usability and illustrates some common ways in which usability problems are manifested in everyday products. It then offers several possible operational definitions of usability that provide a starting point for assessing a product.

- "Introducing Usability Testing" describes the elements of usability testing that distinguish it from other types of product testing that might be conducted

throughout the development cycle. Unlike some other types of tests, such as validity testing, which is concerned only with whether a product functions as designed, usability testing requires real users (or people who have similar characteristics to those in the product's intended target audience) to perform representative tasks so that product developers can be assured that the product meets users' needs.

- "Recognizing the Value and Limitations of Usability Testing" delineates the benefits that product users as well as product developers and corporations gain from usability testing. It also discusses limitations of usability testing and reasons for resistance to it.

- "Conducting Iterative Usability Tests" stresses that the best results are obtained if usability testing is conducted throughout the development process. This section also describes how usability testing can aid design decisions at various stages of product development, especially when it occurs early in the process.

- "Conducting Front-end Analyses: Task Analysis and Participatory Design" discusses two important techniques of early testing – task analysis and participatory design. It includes the definitions, procedures, limitations, and benefits of these two techniques.

- "Preparing a Standard Usability Test Procedure" focuses on the most used type of testing and provides an overview of developing a plan for designing a standard usability test. This section also describes the test's basic elements and includes a series of checklists to help the person conducting the usability test to prepare for the testing.

- "Selecting the Right Test Monitor" describes the role of usability test monitors – the persons who conduct usability tests. It lists the qualities of successful test monitors and describes behaviors that successful test monitors should adopt.

- "Choosing the Test Environment and Participants" expands upon two aspects of test preparation: selecting a test environment that is conducive to usability testing – one that is neutral and free of distractions; and defining the characteristics of and selecting the appropriate test participants.

- "Ensuring the Rights of Participants" describes the legal rights of both the test participants and the testing organizations to ensure that test monitors act, and test participants are treated, ethically. This section also discusses the ethical obligations a testing organization has to the test participants.

- "Collecting Data From a Usability Test" outlines various methods used to obtain usability data and suggests formats for collecting test data.

- "Analyzing and Interpreting the Data" describes strategies for extracting both qualitative and quantitative data from a usability test and suggests methods for interpreting the data.

- "Making Recommendations" describes ways to convert the information obtained from a test into a series of ordered recommendations that managers and the product development team can act upon.

- "Presenting the Results" offers suggestions for reporting usability test findings so the audience will respond favorably. Results can be presented in any combination of written, oral, or videotaped formats.

This paper concludes by urging professionals – technical communicators, in particular – to routinely incorporate usability testing into their product design efforts.

Please contact us at lutzja@muohio.edu to ask questions or offer comments about this paper. In keeping with the principles of usability, your feedback will be used to improve future versions of this document.

Introduction

Usability testing, the process by which products are tested by those who will use them, is intended to help product developers – including information product developers – create, modify, or improve products to better meet the needs of actual or intended users to make those products user-friendly.

According to Dumas & Redish (1993), authors of *A Practical Guide to Usability Testing,* usability testing helps product developers determine whether "the people who use the product can do so quickly and easily to accomplish their own tasks" (p. 4). This paper discusses common issues in usability and usability testing. It examines the benefits and limitations of usability testing, and it describes the various elements involved in planning a test, assessing results, and presenting recommendations.

About the Authors

As students in Miami University's Masters of Technical and Scientific Communication (MTSC) program, we recognize the importance of creating communications that are not only effective, but also usable. In many respects, we view ourselves as advocates for users, and one way to determine whether products actually meet users' needs is through usability testing.

Usability testing is especially important to technical communicators because, as Dr. Karl L. Smart, Senior member of the Society for Technical Communication, Intermountain Chapter says in his article "Quality and the Customer Experience: Methods of Collecting Data," one of our roles is to enhance the customer experience. "To help customers have better experiences," Smart says, "we often must redefine what we do as technical communicators. We're also facilitators of communication who assist real people in using technology to do work. Our goal is … to enhance customers' experiences in using a product to achieve their aims."

Purpose of This Paper

Usability testing is often used to improve or modify consumer products. To learn more about usability testing, we took on the task of researching the major issues in usability testing that we present in this paper.

The purpose of this paper is to introduce readers, through our research, to the definition, benefits, and issues involved in a commitment to test usability throughout the product development life cycle as a means of helping to ensure ease-of-use and ultimately the success of a product. This paper provides a variety of examples that illustrate how usability testing has helped product developers design more effective and usable products.

This paper does not provide an exhaustive and complete explanation of usability testing. It is not intended to be a "how to" or guide to conducting usability testing; Rather, it is meant to serve as an introduction to this important area of product development.

We have posted this paper on the MTSC website because we wanted to express our belief in the value of usability testing, especially in the field of technical communication. Although technical communicators are not always given the authority to make decisions about implementing usability testing, they can communicate its importance and benefits to product developers and project leaders. In addition to aiding technical communicators, the information contained in this paper can be useful to anyone involved in or concerned with the process of developing user-friendly products. It will be best used as a foundation from which to learn more about the effectiveness and benefits of usability testing.

Organization of This Paper

We organized this paper starting with general information about usability testing and then offering more specific information about various aspects of conducting usability tests. We begin by defining the concept of usability and explaining how usability testing can be used to help develop user-centered products. We describe the elements involved in conducting usability tests, from planning and designing usability tests to actually conducting them. We examine methods used to collect and analyze the data from usability tests. Finally, we explore ways to present usability test results so that they are most likely to be incorporated into a product's development processes.

How to Use This Document

This document uses the conventions that follow.

- In the electronic version of this document you can click any section name or page number in the Table of Contents to go directly to a section. Click the Back button to return to your original place in the document.
- Glossary terms appear as hyperlinks. In the electronic version of this document you can click any term to go directly to its definition in the "Glossary." Click the Back button to return to your original place in the document.
- Section names also appear as "hyperlinks." In the electronic version of this document you can click any section name to go directly to that section. Click the Back button to return to your original place in the document.

Defining Usability

Usability is defined as the characteristic of being easy to use ("Usability First," 2001). More precisely, the usefulness of a product can be assessed by two key features – its utility and its usability (Nielsen, 1993). Utility refers to a product's capability to carry out an intended function. Usability refers to how easy users find it to accomplish that intended function. For example, a VCR may have the capability to allow users to videotape television programs when away from home. But, users may find it difficult to program the VCR for that purpose. Consequently, while this VCR would have high utility, it would have low usability (Dumas & Redish, 1993).

A wide variety of products have been tested for their usability, including written documents, websites, computer programs, consumer products, medical equipment, and voice response systems to name a few (Dumas & Redish, 1993). Usability is an important consideration for all types of products. The word "product" can be used in a broad sense and this paper uses that term to describe a variety of user goods – software programs, VCRs, children's car seats, websites, and documentation.

There are four signs that indicate to product developers that a product has a low level of usability (Dumas & Redish, 1993).

- User dissatisfaction with the usage experience.
- The presence of workarounds.
- The need for rework.
- Low levels of product usage.

The presence of even one of these symptoms can alert product developers to the need to make a product easier to use.

To illustrate, consider the usability issues that consumers face in purchasing tickets for concerts or sporting events through an automated phone system designed to make ticket purchases easier and more efficient. When a consumer calls to purchase tickets s/he is confronted with a sophisticated voice recognition technology system. The consumer is required to respond to a series of system-generated prompts as the transaction proceeds. In providing purchasing information, ticket buyers must pronounce, spell out, and confirm each element of their shipping address and credit card information. For example, when providing their address, buyers must pronounce the name of their street – Bennington Drive – and then, spell the name – B-E-N-N-I-N-G-T-O-N. The automated system repeats the street name – "I thought I heard you say Bennington" – and requests that buyers press 1 or 2 on their touchtone phones to confirm or change the entry. The above process must be repeated for each element of the buyer's address and for credit card information. This redundancy is used to ensure the order's accuracy, but to a user the redundancy is a nuisance. People who phone in ticket requests seek an efficient way to make a purchase and the inefficiency of the system described reduces the product's usability. People have choices and they will likely *not* choose a product they perceive as inefficient.

An Operational Definition of Usability

Recognizing usability problems in products is powerful, but product developers can go beyond simply recognizing them. Ideally, they can build usability into new products by obtaining user feedback on the product's ease of use throughout the development process. This feedback can be qualitative – user feedback about the quality of the product or some aspect of it – or quantitative – a measurable result such as the time it takes to complete a task. In either case, product developers must create an operational definition of usability for a product in order to collect feedback in a systematic way. Some criteria that can be used to develop an operational definition follow.

- Learnability describes how quickly a novice user can develop a basic proficiency with the product and generate results at some desired level. Users prefer products that allow them to be productive after only a short training period.

- Memorability refers to how well an occasional user can use the product after having not used it for a certain time. If the product has poor memorability, it may be difficult for the occasional user to remember how to use it effectively. High memorability can eliminate the need for retaining in order to use the product properly.

- Efficiency can measure how productive an experienced user can be with the product. For example, one measure of efficiency for an order entry software application might be the number of orders that customer service representatives could process per hour after they have become adept at using the software. Users desire a high level of efficiency so that they can be more productive.

- Error tolerability requires that the product helps prevent users from making errors and allows users to quickly recover from their errors. It is important that any errors that do occur are not catastrophic. For example, in a word processing application, users should not lose their documents when they make an error saving a data file. Likewise, users should not need to start over from the beginning of the process when an error does occur.

- Likeability is a subjective measure of how well users enjoy using the product. If their experience is a good one, users will likely continue to use the product regularly to accomplish their tasks.

Each of these criteria allows product developers to measure usability in both qualitative and quantitative terms. For example, learnability might be measured quantitatively as the time required for a novice user to correctly install a new software program. Efficiency might be measured quantitatively as the number of orders that an experienced customer service representative can process per hour. Likeability might be measured qualitatively on a subjective scale that asks users to

rate their satisfaction with the product. (See "Analyzing and Interpreting Data" for more information on collecting and assessing qualitative and quantitative data).

In order to measure their product's ease of use, product developers and/or usability researchers will need to tailor these five criteria of usability to the details of their particular application. For some criteria, they will also need to establish a pre-defined measurement to serve as the basis on which results can be evaluated.

Introducing Usability Testing

Usability testing can help product developers ensure that their products are easy to learn, that users can efficiently use the products to perform the desired tasks, and that users enjoy using the products (Rubin, 1993). In usability testing, product developers collect empirical data by making observations of users who are conducting representative tasks using a product or prototype (Dumas & Redish, 1993). Product developers use these data to decide which product performance features meet standards or require improvements. While the details of the usability test will depend upon the test's objectives, all usability tests share some common characteristics.

Characteristics of Usability Testing

There can be wide variations in the types of tests conducted, but effective usability testing shares four common characteristics: defined objectives, real users, real tasks and early and iterative testing. A description of each characteristic follows.

- **An effective usability test has defined objectives.** Every usability test – like any other experimental test – requires objectives. Each usability test objective should be clarified well in advance of the actual testing. Clear objectives allow product developers to choose test participants, test methods, and user tasks that are best suited to address the questions of interest.

  For example, if the primary objective of a usability test is to assess a product's learnability for first-time users, then the test will require selecting novice rather than experienced product users. If the objective focuses on eliminating the need for user workarounds by those users who have prior experience with the product, then product developers should choose experienced users who understand and who can compensate for a product's limitations. By having clearly defined objectives based on the experience of the intended users, then product developers improve the chances that their concerns will be addressed by a particular test.

"Collecting Data From a Usability Test", which appears later in this paper, outlines various methods for obtaining usability data in more detail.

- **An effective usability test uses real users**. Usability testing requires that test participants have similar characteristics to users in the product's intended target audience. Consequently, reviews by design experts are not considered usability tests, although such evaluations may be conducted as part of the overall development process.

  Likewise, product developers should not be selected as test participants because their understanding of the product exceeds that of most users. As Dumas & Redish point out (1993, p.23), "If the participants are more experienced than actual users, you may miss problems that will cause the product to fail in the marketplace. If the participants are less experienced than actual users, you may be led to make changes that aren't improvements for real users." The choice to test with real users allows users' needs, rather than the designers' preferences and biases, to drive the product design.

- **An effective usability test monitors real tasks**. Users will provide the best feedback when testing is conducted in the same context in which they will use the final product. This does not mean that each usability test must be conducted in the users' homes or at their workplaces; tests may be conducted in a testing laboratory or a facsimile of the "real" environment.

  Usability testing is particularly challenging with the globalization of markets. Factors ranging from cultural differences to differences in time zones can affect when and how testing should be conducted. However, if the users' feedback is to be helpful, the tasks must fairly represent the tasks users will actually perform with the marketed products, regardless of where they are located.

- **An effective usability test is conducted early and iteratively**. Usability testing should begin early in the product development process – from the time the design is put on paper – and continue throughout the process, through

prototyping and finally to the finished product stage. Early testing can help product developers refine specifications to ensure that the product's design fits the mental model that users have for it and to help it feel more intuitive to users.

When used to its best advantage, usability testing permits product developers to use the feedback to drive design choices and to make appropriate design modifications. Each time usability testing is conducted, it should incorporate results from the previous test into the next step of the design process.

Overall, usability testing is a way to ensure that users have a positive usage experience and are satisfied with the product. But it is important to keep in mind that usability testing is only part of an overall design process that focuses on meeting users' needs.

Usability Testing and the Design Process

Usability testing is a component of a broader development process known as user-centered design. User-centered design (UCD), which is known by other names, among them usability engineering, human factors design, ergonomics and computer-human interface design, places users' needs at the focal point of design development and incorporates a variety of tests in addition to usability testing. For example, human factors experts may conduct an expert evaluation of a prototype to prepare the prototype for initial testing.

Many factors are common across product lines and these factors can be considered up front without direct input from users. However, usability testing should be conducted on a prototype of the product to ensure that the human factors engineer has not overlooked a key user need.

Later in the development process, functional testing will be needed to ensure that the product works as intended prior to release. For example, a functional test of a website might be conducted to see if the links that appear on pages actually connect to the desired destinations. A functional test of a user manual for a software

application might be conducted to determine whether cross-references, index entries, and the table of contents reference the correct page numbers.

Occasionally, corporations conduct usability tests only near the end of the development process to see if users are satisfied with their usage experience. This type of testing is referred to as validity testing or verification testing. It is only a narrow use of usability testing because it is not iterative, which is the key aspect of usability testing that makes it successful. Validity or verification tests are referred to as usability tests only in the sense that they measure user satisfaction – which is also one of the key aspects of usability. But changes that are recommended as a result of usability tests conducted late in the development process may be superficial and may not necessarily be implemented due to cost and time constraints.

Early and iterative testing will yield the largest benefit. In fact, some professionals in the field of usability testing like Stephanie Rosenbaum, President of Tec-Ed. Inc., a user advocacy firm that was founded in 1967, claim that as few as eight cycles of testing during a product development life cycle, with just one or two participants, can increase user accuracy by 20%.

Recognizing the Value and Limitations of Usability Testing

Recognizing the value and limitations of usability testing is essential to companies and product developers. Companies cite valid reasons for not conducting usability testing – budget constraints and release dates to name just two. But there are convincing arguments for conducting usability testing that have equally important implications for the product and the company.

Recognizing the Value of Usability Testing

As stated in "Introducing Usability Testing," companies and product designers can easily recognize the value of usability testing from products that lack usability. Usable products support users' workflow and helps users avoid errors. Usable products can also provide users with feedback, give users control, and minimize users' cognitive load. Products that are not usable are both frustrating and inefficient for users.

Usability's greatest goal is to positively affect users' experiences, so it stands to reason that the greatest value of usability testing is to the users. But recognizing that usability testing directly affects the quality of users' experiences also makes it an integral part of the product development life cycle and essential to companies and product developers.

The Value of Usability Testing to Companies

Usability testing has short- and long-term value to companies. In the short term, usability testing can help companies minimize the cost and risk associated with releasing a product that has potential usability problems. In the long term, usability testing can help companies increase revenue, sales, and brand loyalty; acquire a competitive edge; and create historical records of usability benchmarks for future releases of products and future products.

- **Usability testing helps minimize cost.** When usability testing is performed, documentation and training costs as well as user support costs can all be minimized. Usability testing can help reduce the cost of unscheduled updates, maintenance, and product recalls that are required due to poor product design. Other costs that can be minimized include additional labor costs commensurate with increases in the support staff needed to handle high volumes of service calls.

- **Usability testing helps minimize risk.** Releasing a product after usability testing is far less risky than releasing a product with serious usability problems. Consider the implications of releasing children's car seats without conducting usability testing. If the car seats have serious usability problems, parents may use them improperly and subsequently, children may be injured. When this type of usability problem is discovered after product release, product recalls and personal injury lawsuits may ensue. The possibility of negative publicity and decreased consumer confidence are also risks that companies take when they do not implement usability testing prior to product release.

- **Usability testing helps increase revenue, product sales, and brand loyalty.** User satisfaction with a product may spark a chain reaction that leads to increased revenue for companies, increased product demand and sales, brand loyalty, and increased interest in future product releases and other company products. Users who have positive experiences with a company's products also tend to influence other potential users to try that company's products.

- **Usability testing helps companies acquire a competitive edge.** Usability has become a benchmark by which customers separate products in their minds. Scanning products on the shelves of local stores shows that product packaging often advertises increased usability. "Easy open top," "easy assembly, no tools required," and "just add water" are some examples of how advertising slogans stress usability. Usability testing results suggest that users prefer products that are as easy to use as their advertising claims.

- **Usability testing helps create a historical record of usability benchmarks for future release.** Companies can ensure that future products either improve upon or maintain current usability standards by keeping a record of usability test results.

The Value of Usability Testing to Product Developers

Usability experts like Jakob Nielsen and Michael Wiklund suggest that usability testing can be of great value to product developers in many ways: in more efficient use of their time, in the minimization of unscheduled updates, maintenance and recalls, and in reducing the time required for documentation and training.

- **Usability testing helps product developers use their time more efficiently.** Product developers use their time more efficiently when usability testing is part of the development cycle because adequate product definitions, user definitions, and user feedback can reduce or eliminate the time-consuming revisions associated with poorly designed products.

- **Usability testing helps minimize the need for unscheduled updates.** Highly usable products reduce the need for product developers to produce and document unscheduled updates and maintenance releases. In May 2002, the U.S. National Institute of Standards and Technology estimated that making corrections to inadequately designed software products alone cost development companies and users between $22.2 billion and $59.5 billion annually.

- **Usability testing helps make developing documentation and training easier.** Usability experts suggest that developing products that are consistent, predictable, and are in the users' language can make documenting a product easier. Dumas & Redish (1993) suggest that "the more that writing a task-oriented manual requires reinterpreting and reorganizing the product's structure and metaphor, the more difficult it is to develop a coherent and concise manual or a logical tutorial" (p. 17). Therefore, usability testing can make documentation

and training easier for product developers because it helps them analyze their users' needs well.

The Value of Usability Testing to Users

Product developers, including technical communicators who have been influenced by usability experts, including Joseph Dumas, Janice Redish, and Jeffrey Rubin, advocate using usability testing as part of a user-centered design approach. Dumas, Redish, and Rubin suggest that usability testing provides the greatest value to product users because it places the focus on developing usable products and helps increase user satisfaction.

- **Usability testing focuses on developing usable products**. Usability testing keeps product development teams focused on users' needs. Usability testing drives decisions and helps product development teams recognize that users' needs should guide the development process. Product developers can use the results of usability testing to outline usability goals that will lead to finished products that take less time to learn and require fewer steps when used to accomplish a task.

- **Usability testing helps increase user satisfaction.** Usability testing can increase user satisfaction especially when products live up to their advertised promises of increased usability. Also, because usability testing helps product developers to detect product flaws before they release the products, users can feel confident that the products they purchase have minimal manufacturer defects.

Recognizing the Limitations of and Resistance to Usability Testing

Just as it is important to recognize the value of usability testing, it is equally important to recognize and understand its limitations and reasons why usability testing is sometimes resisted and occasionally omitted from the product development life cycle.

The Limitations of Usability Testing

Recognizing and understanding the limitations of usability testing can help companies and product developers make the most of every usability test they conduct. Usability expert Jeffrey Rubin states, "[Usability] testing is neither the end-all nor the be-all for usability and product success and it is important to understand its limitations" (Rubin, 1994, p. 27). Rubin suggests that even when 100 percent certainty that products work during testing is achieved, there still can be no absolute assurance that those products will be usable when they are released. One limitation of usability testing is that, however close to reality it comes, usability testing is still only a simulation of a product's use. Another limitation is that testing is not proof that a product works. And a third limitation is that test participants may not truly reflect actual users.

- **Usability testing is conducted in a simulated environment.** Whether testing is conducted in a lab or in the field, it is still conducted in a simulated environment. Even when usability testing adheres to strict test procedures, it still only artificially represents product use. As Rubin states, "the very act of conducting a study can itself affect the results" (p.27).

- **Usability testing does not necessarily prove that products work.** How products work, or whether they work at all, depends significantly upon the way testing is conducted. What is actually measured during usability testing is the *statistical probability* that products work; statistical probability provides no absolute guarantee that products will work when they are released. Although this limitation seems ominous, it can be minimized. "Choosing the Test Environments and Participants," which appears later in this paper, explains how.

- **Usability testing may include test participants who do not represent the target audience.** Product developers sometimes have difficulty identifying and describing their actual product users and thus choose test participants who do not represent the target audience. Rubin agrees when he states, "Participants

are only as representative as your ability to understand and clarify your target audience" (p. 27). Product designers can minimize this limitation by carefully analyzing their product's intended users and choosing test participants who best reflect their characteristics.

The Resistance to Usability Testing

Usability testing sometimes faces great resistance and is subsequently left out of the product development life cycle altogether. The three main arguments why usability testing may be resisted include the cost, the fact that product developers believe that they themselves can detect and correct any flaws, and the amount of time it can add to the product development life cycle that will delay the product's release.

- **Testing may be costly.** For many companies, usability testing costs include the cost of recruiting a usability specialist, the cost of space and videotaping equipment that might be needed to conduct testing, and the cost of enticing participants. Many companies are unwilling to factor these costs into the total cost of product development. However, conducting simpler usability testing can minimize testing costs. According to usability expert Jakob Nielsen "current best practices call for spending about 10% of design budgets on usability" (2003). Others, such as Usability by Design, a usability specialty company that is a member of the Usability Professionals Association, claim that usability requires just 2.2% of the development budget.

- **Testing is not always the best technique to use.** Usability testing is just one technique that can be used to evaluate and improve products. Companies often view other techniques, such as expert evaluations, as more time and cost effective. Additionally, companies may not need to bring in test participants to reveal obvious design flaws in product development.

  Consider software development as an example. One way a company can remedy obvious flaws in new or revamped applications is to have other employees, who are not on the product development team, test the software at various stages of

the product development life cycle. Developers can then remedy the flaws without conducting usability testing, and thus deem usability testing unnecessary.

- **Usability testing extends the product development lifecycle.** Resistance from management may occur because the iterative nature of usability testing requires extending the product development life cycle. Extensions may delay the product release date. Managers often argue that there is simply no time in the product development life cycle to plan iterative usability testing, allow usability specialists to become familiar with the product, observe participants, consider observations, and submit recommendations to product developers. It is for these reasons that managers often feel that usability testing is a luxury that time does not permit.

The Value of Usability Testing Outweighs its Limitations

In spite of these limitations and reasons for resistance, those who advocate and administer usability testing still find that it is a very effective indicator of potential problems with products and provides a means to resolve these problems. Many of these limitations can be minimized if usability testing is conducted with precision and care, at the appropriate times in the product development life cycle, and as part of an overall user-centered design approach (Rubin, 1994).

Rubin and other usability experts like Dumas and Redish suggest that conducting usability testing that is iterative may be more valuable than conducting one large test at the end because smaller tests require less time and money.

The value of usability testing –to companies, to product developers, and to users – far outweighs its limitations. Other usability experts agree with Rubin when he states "it is better to test than not to test" (p. 27). Case in point: When Microsoft Corporation developed the user interface for the new Windows 95 operating system in 1992, they found that, through iterative testing, no detail of the initial interface design survived unchanged. Said Kent Sullivan, author of The *Windows® 95 User Interface, A Case Study in Usability Engineering, "*Although doing design and

usability testing iteratively allowed us to create usable task areas or features of the product, usability testing the product holistically was key to polishing the fit between the pieces . . . If we had not done this testing, users' overall experience with the product would have been less productive and enjoyable."

Conducting Iterative Usability Tests

Usability testing is most effective when it is conducted iteratively throughout the product development life cycle. Iterative usability testing is a systematic method of repeatedly observing users as they use or attempt to use a product at different phases during product development. Usability researchers analyze results from each test to determine how easy or difficult it is for users to use the product.

Experts agree "usability testing is best when used early and often, not at the end when it is too late to make changes" (Redish & Dumas, 1993, p. 22). The effectiveness of usability testing is optimized when it is used to diagnose problems with a product, and not to verify at the end that the product works.

Renowned usability expert Jeffrey Rubin writes, "Usability testing is most powerful and most effective when implemented as part of an iterative product development process. That is, a cycle of design, test and measure, and redesign throughout the product life cycle has the greatest probability of concluding with a usable product" (1994, p. 30, 31). A product development life cycle that includes testing and involves users throughout helps ensure a highly usable and marketable product.

"Conducting Iterative Usability Tests" presents the following information.

- The value of iterative usability testing.
- The goals of iterative usability testing.
- Definitions for various types of tests that can be used at different phases of product development and the benefit of testing at each phase.
- Resistance to iterative usability testing.

The Value of Iterative Testing

By definition, iterative testing means conducting a few, small, periodic tests over the course of product development. The iterative process is more valuable than conducting one large test in the end because small tests are less costly and require

less time. Products that have been tested iteratively during their development have better marketability and companies that sell usable products have greater profitability. Mary Dieli, a former usability manager at Microsoft (Dieli, 1989, p. 2) said, "In Microsoft's usability group, we define our mission as working with writers, designers, and developers to help produce intuitive, easy-to-learn, and easy-to-use products. With each test that we run, we learn something new about our processes and methods."

Many cutting edge technology companies, such as Microsoft, IBM, and Hewlett-Packard have adopted usability testing as part of their product development processes by investing in usability labs. They have successfully dominated the market by producing highly usable products that have been tested repeatedly prior to being released to consumers. These companies have realized the many values of iterative testing, including the elimination of errors, the fact that errors can be fixed more easily earlier in the development process, that improvements suggested early are more likely to be implemented, and that prototype testing is less expensive and more effective than testing the final product.

- **The more a product is tested, the higher the chances are that designers and testers will eliminate errors.** According to Nielsen, it is better to distribute a testing budget across several small tests using fewer participants than on one large study in the end (Nielsen, March 2000). He claims that after testing a product with a fifth user, "you are…observing the same findings repeatedly but not learning much new." Thus, testing iteratively with a few users ensures optimum use of the testing budget, because every iteration is another opportunity to refine a product.

- **The sooner usability testers find problems, the easier it is to fix them.** If initial testing helps identify problems in a product while it is still on the design table, it is easier and less expensive to fix. Finding and fixing problems early will reduce rework later in the product's development. If product developers find a problem late in the development life cycle, it is more expensive to correct the product. Redesign requires time and costs that were not part of the original

scope.

Consider Forrester Research Inc.'s estimates for making corrections to a website after development is complete, which were reported in *CIO Web Business Magazine* by Sari Kalin (1999). Forrester estimates that websites that are difficult to use lose up to 40% of return visitors. And fixing a website after it is up and running can be expensive.

Forrester estimates that quick fixes – changes that can be made in three to six weeks, such as making vocabulary consistent or ensuring that site navigation is reliable – can range in cost from $8,500 to $17,000. A complete re-architecture of the website to correct problems such as changing content so that it is organized by user goals and provides functionality that users demand, can take 24 to 39 weeks, and can cost between $780,000 to $1.56 million. According to Nielsen, the first round of testing with as few as five users can uncover up to 85% of usability problems in website design (2000).

- **Changes are more likely to be approved if they are recommended early in the development process.** Testing will often help identify problems and serve as a check to confirm that problems identified during previous tests have been corrected. Iterative testing is conducted from the predesign or planning stage to post development just before the product is released. "If a product is tested early and iteratively, there will be checks and balances into the process" (Dieli, 1989, p. 4). Thus, if usability researchers miss a product flaw in one iteration of testing, they can identify it in the next round of testing. In addition to identifying flaws in each iteration, corrections made to the previous iteration as a result of testing can be checked in the next iteration.

- **It is possible to test usability and design issues with prototypes and models of the actual product.** Product developers can create mockups of products, which is far less expensive than testing a finished product. For example, web developers might use a paper outline of a website to test navigation and organization.

Iterative testing reduces the learning curve for users, allowing them to learn and use more functionality of the product with less effort. It helps reduce the need for product recalls, or, in the case of software applications, updates and maintenance releases that companies often provide free of charge to users. Iterative testing that is conducted throughout a product's design and development results in a product that is easy to learn and easy to use.

Goals of Iterative Testing

Usability testing is part of a larger concept in product development: usability engineering, which is an approach to product development where a product is customized to the needs and desires of its users. The main goal of both usability testing and usability engineering is to improve the usability of a product. In usability engineering, usability is engineered into a product "through an iterative design and development process. (Dumas & Redish, 1993, p8). Iterative testing contributes to developing highly usable products; therefore, usability engineering is incomplete without iterative testing.

Short-Term Goal

The short-term goal of integrating iterative usability testing into the development cycle is to discover ways to improve the product. In the case of computer and computer-related products, the goal is to create a list of "general and specific recommendations for improving the hardware/software, training and/or other collateral materials provided to end users" (www.usability.gov, accessed on 9/18/02). Analyzing the results of iterative testing enables testers to recommend changes for improving the product.

Long-Term Goal

The long-term goal of iterative usability testing is to continue to create "highly usable and marketable products by increasing and managing user input through the product development cycle" (www.usability.gov, accessed on 09/18/02). A company

that continuously creates usable products enjoys many benefits, such as enhancing the company's reputation, increasing sales, increasing consumer and investor confidence in the company's products and thereby, increasing profitability.

Classification of Usability Test Types

Usability researchers normally use different types of usability tests at different points in a product's development cycle. Every iteration of testing will have different goals, thus during every iteration, a different type of test will be administered. During these tests, the focus is on answering questions like the ones listed below.

- Can the user complete a task using the product successfully?
- How fast is the user able to do each task?
- Is the speed efficient enough to satisfy the user's goals?
- What paths does the user take to complete the assigned task?
- Where does the user stumble, encounter problems, and/or get confused while using the product?
- Are the problems identified in the previous iteration of testing still surfacing in this iteration?

(Adapted from http://usability.gov/methods/usability_testing.html, 09/18/02)

The main types of tests conducted during iterative testing are listed below.

- Exploratory tests are used early in the product development life cycle and help to establish the validity of conceptual or high-level design prior to the development of fine details.
- Assessment tests are used as information-gathering tools during early development to evaluate the "usability of lower-level operations and aspects of a product" (Rubin, 1994).
- Validation or verification tests are conducted near the end of the development cycle and are used to confirm a product's usability.

- Comparative tests can be used in conjunction with any of the above tests. They are used to compare two or more aspects of a product. For example, a design element and a functional element. Comparative tests are used to determine the best design that enables ease of use; they are also used to understand the advantages of using a particular design over other designs.

During the initial phases of product development, exploratory and assessment tests are ideal formative tests. Formative tests are designed to detect flaws in a product that can be fixed. Using paper ideas and prototypes and implementing only high-level functionality can help developers find usability problems early. Because the product is still in its infancy, changes are easy to make.

In later stages of product development, assessment and validation tests can be used to find usability problems when the product is used in context. These are referred to as summative tests – tests that are designed to obtain measurements that indicate usability when the product is used in the environment for which it was designed.

Figure 1 illustrates different test goals, types of tests, and the stage of development at which each test is most effective.

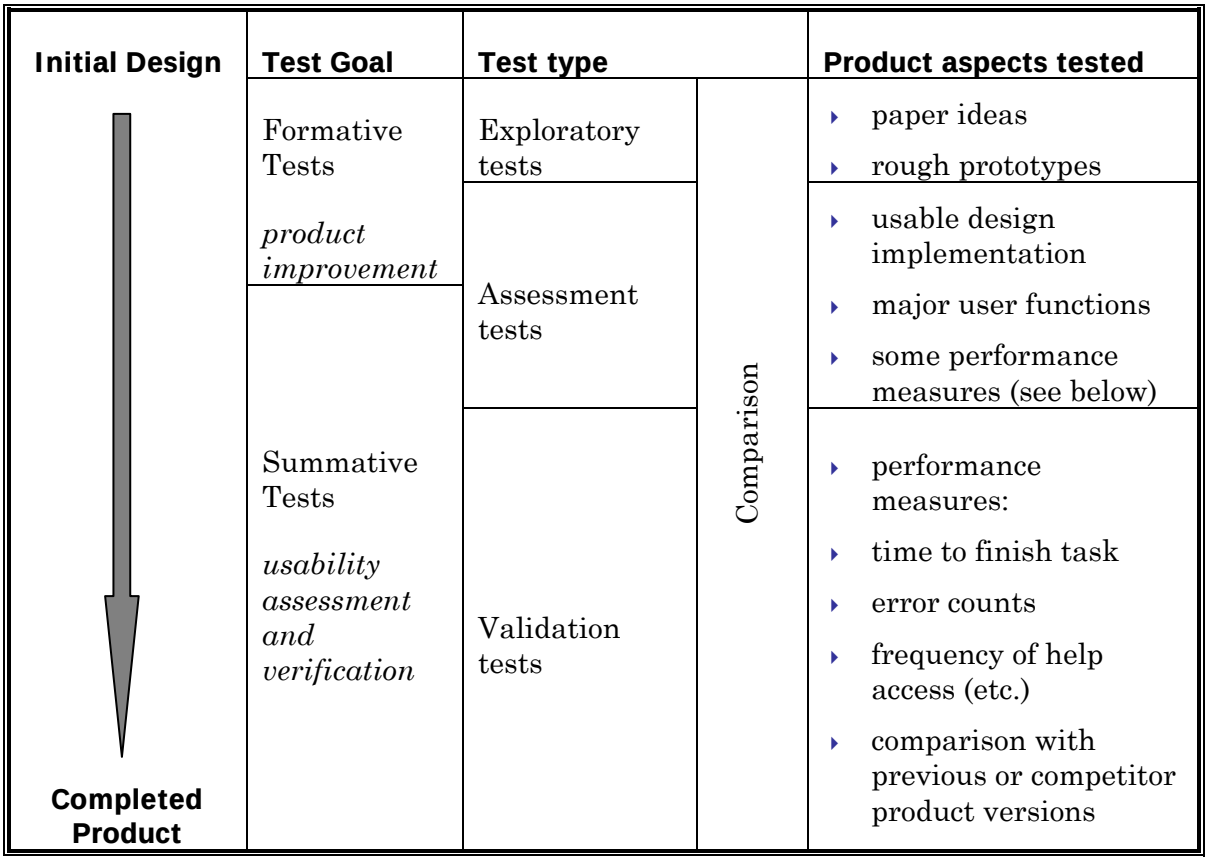

Figure 1. Classification of usability tests.

Testing During Pre-design

In the pre-design or planning stage, usability researchers conduct exploratory tests "to evaluate the effectiveness of preliminary design concepts" (Rubin, 1994). Early analysis and research are critical at the planning stage when vital design decisions are being made.

Usability researchers conduct initial usability tests using ideas on paper, prototypes, previous versions of the product, alternative designs, or even competitors' products. They may use testing to identify the flaws of previous versions, as well as to plan enhancements to the new version.

Conducting a usability test during the pre-design stage helps to provide input into the design of the product before development starts. Here is an example: Usability

test experts at Terra Lycos, an international network of websites based in Barcelona, Spain, use a card-sorting technique to design sites. A tester jots down ideas and terms on index cards. A participant then arranges the cards. "Out of that comes a hierarchy about how things should be put together, which helps with making menus and taskbars," explains Dave Hendry, the company's network manager for user interface research (Oreskovic, 2001). This example demonstrates how participant input in the design phase helps Terra Lycos arrive at decisions for organizing websites.

Testing During Early Development Phase

Through the early development stages, usability researchers conduct exploratory and assessment tests with available materials and prototypes.

Prototypes let product developers evaluate how well the structure and functionality of the product meets users' needs before companies invest too much time, money, and effort. Even though the prototype usually does not include all the functionality, it does mimic the actual product, enabling testing even if the product is not fully developed. The ideal prototype has enough detail or functionality that a user can complete a task; but the design can be changed easily based on the results of usability testing. Prototype testing enables designers to easily fix errors that might prove to be expensive if they are caught at a later stage.

Here is an example: writers developing instructions for a product can use a mockup or create sample pages to test the effectiveness of page design features, or test using one section or chapter even before the instructions are written in their entirety.

Website development is another example of how early prototype testing can enhance the finished design. IBM's Jeanette Fuccella, a human factors engineer, and Jack Pizzolato, website designer, advocate enlisting early and repeated user input into the layers of web design. Getting iterative feedback from users about the content and graphics of a website separately – through two independent tests – helped Fuccella and Pizzolato shorten the design process and enabled the IBM website designers to develop more usable websites.

To test website content, the IBM designers created an outline of the website content in a 'wire frame' – a simple HTML model of a proposed website – that identifies the navigation and location of content within the proposed website. Placeholders were used for visuals that would be added later so they would not be distracting to test participants. This prototype helped the designers test only the content and navigational elements of the site.

Fuccella and Pizzolato then tested the graphics before adding them to the site. The designers tested with prototypes using greeked text, which is dummy text used as a placeholder for the real text. In this test, usability testers were able to gather important information about the graphics without distracting users with content and navigation.

The website designers followed up with usability testing on the website as a whole after they implemented changes from the previous tests (adapted from Fuccella & Pizzolato, June 1999). Testing for content and graphics separately during the early stages of website development enabled these IBM designers to create highly usable websites.

Testing Through the Development Phase

As development progresses, more details are added to the prototype, which can still be used to test. Testing at various predetermined points throughout development helps identify problems with the product design and usability.

Usability testers use functional testing to ensure that the products are working as designed. For example, in website development, testers may verify that the website displays on screens and functions consistently using different browsers. If the targeted user group consists of vision-impaired users, testers might check the website for its ability to support large fonts. Also iteratively checking the integrity of the links in a website, especially after content changes, is a valuable step in maintaining accuracy, as evidenced in the example that follows.

Testers on a documentation team used storyboarding to test links in the content that it was creating for an intranet website. There were numerous individual procedures containing links to other procedures. The team printed out paper copies of all the procedures, and arranged them on a wall according to the way the procedures were linked on the website: a procedure was followed by one or more procedures that were linked to it. This arrangement helped the team identify faulty links, faulty file names, and missing links. Before the final outcome, the team conducted this type of test two more times until all the links were correct.

Testing Before Product Release

Testing at the end of a product development cycle is the most common type of usability testing. At this stage, though, testing serves only to validate the product's functionality. It is usually a checkpoint to ensure that all the problems identified in the previous tests have been corrected.

At this stage, making major changes to the product's design might jeopardize its release schedule. Information that comes from testing at the end of product development may not impact the current product very much. However, this information will provide valuable input into designing future versions of the product.

Resistance to Iterative Testing

Usability experts believe that iterative testing is enormously effective in producing highly usable products. Even so, many companies are reluctant to invest in regular usability testing at various stages in the development cycles of their products. It is difficult enough to obtain companies' commitment to usability testing at the end of product development; it is more challenging to get them to commit to iterative testing.

Companies may view iterative usability testing as a nice-to-have, not an absolute necessity. Many companies feel that iterative usability testing is an economic drain.

Moreover, building repeated testing into the development schedule delays the time when the product will be market ready.

It is true that integrating usability testing into the product development cycle requires some investment in terms of money and time on the company's part; however, this investment may be justified when the company reaps the benefits of marketing a highly usable, successful product. Consulting firms that administer usability testing for their clients, such as Usability by Design based in the United Kingdom, claim that key cost metrics –support calls, for examp0le – fall by at least 10% when companies include usability testing as part of the development process.

If a company releases a product with flaws because it is unwilling to invest in usability, it might find itself making additional expensive investments in recalls, updates, maintenance, and service demands from customers. Committing to usability and implementing iterative testing is a good investment to ensure the company's credibility in the market. Furthermore, as customers realize the usability of the company's product, consumer confidence in the company will grow, helping to ensure success with future products. As products continue to improve throughout the development process as the result of iterative usability testing, the benefits of investing in usability engineering become increasingly clear.

Conducting Front-end Analyses: Task Analysis and Participatory Design

Including users during a product's earliest developmental stages through task analysis and participatory design – collectively referred to as front-end analyses – are integral elements of usability testing. Task analysis is a process used to determine the objectives users have about a product and the tasks users expect to be able to perform using it. Participatory design is an approach to product development that includes users in the discussion of the design and development of a product. Front-end analyses encourage developers to understand the product users' needs and expectations so that the final product satisfies them.

To illustrate the importance of front-end analyses, usability expert Jakob Nielsen conducted a study in which he asked 13 usability engineers to rate 33 different usability methods in regard to their importance to the overall usability of a product. On a scale of one to five (five being an absolute necessity) iterative design and task analysis ranked number one with a rating of 4.7. Participatory design followed closely behind with a rating of 4.4. Therefore, usability experts consider front-end analyses essential to producing a usable product (Nielsen, 1993, p. 112).

Defining Task Analysis

The development of a user-centered product – one in which users' needs and expectations are the focus of the design – should begin with task analysis, a process that focuses on how users work with a product to accomplish a goal. Task analysis is conducted to identify the specific tasks and sub-tasks that a user will complete using the product. Through task analysis studies, developers are steered away from traditional, product-centered design strategies where design focuses on the product itself or the technology used to create it. Instead, developers seek to place the product in the context of an end-user scenario to ensure that the product is designed for the user, instead of forcing the user to adapt to the product (Johnson, 1998).

For example, Johnson discusses a 1980's scenario in which the Seattle highway department was unable to control increasing traffic congestion. In an attempt to alleviate the problem, Seattle engineers implemented a roadway system that mimicked that of another large city, Los Angeles, but Seattle's traffic problems persisted. Eventually, a group of technical communication students tried another strategy; they questioned the people of Seattle who drove the cars and got stuck in the traffic. The students' study, which consisted of surveys, interviews, focus groups, and observations, pinpointed the drivers' tasks in getting to/from work. The drivers said that the roadway system was not the major problem; it was actually a lack of traffic information. Many drivers watched television news stations and/or listened to the radio news to identify congested traffic areas before they drove to/from work, but none of the news programs offered much information. Consequently, drivers did not know when to take alternate routes. They erred because of a lack of information, not because of a faulty road design. Had the engineers conducted a task analysis, they would have known the tasks that the drivers were completing to reduce traffic congestion, and the problem could have been solved more easily (1998).

Identifying Tasks and Objectives

Through task analysis, developers can identify the users' objectives, the task(s) users consider the most important, the chronological order in which the users perform the tasks, and which tasks are dependent upon the completion of other tasks in order to proceed. Based on the tasks they have identified, developers can also establish testing objectives in order to better align the product with user tasks to increase overall product usability.

Identifying Tasks and Sub-tasks

The types of tasks identified through task analysis will vary depending upon the context and purpose of individual products. However, regardless of the product, every task has the characteristics listed below.

- The task achieves an objective.

- The task has a defined starting point.
- An action (or set of actions) is followed to complete a task.
- The task has a stopping point when it is complete.
- The task has identifiable, but not necessarily tangible, output (Dumas & Redish, 1993, p. 42-43).

Tasks can usually be broken down into sub-tasks or sub-steps that must be performed to complete a major task. For example, the major tasks that must be completed to print a document might include the steps that follow.

1. Type a document in a word processor.
2. Turn on the printer.
3. Choose "**Print.**"

Each of these sequential, major steps also has sub-step(s) that must be performed to reach a goal. For example, step 3, "Choose "**Print,**" usually consists of sub-tasks, including deciding how many documents to print and choosing the appropriate print speed/quality. Product developers should use task analysis to not only understand the major steps, but also to understand the minor steps, or sub-tasks, involved in completing a task (Dumas & Redish, 1993, p. 41).

Identifying Testing Objectives

After identifying tasks, developers should identify testing objectives. Many practitioners are committed to a strong link between test objectives and test planning because these objectives are usually a reflection of the users' reactions to the tasks and sub-tasks associated with a product. Therefore, test objectives should be established after the developers have observed users interacting with the product.

A problem-solving strategy taught in the Miami University MTSC program stresses setting objectives for testing documentation before proceeding with the alpha draft because many objectives stem from user complications with one or more of the tasks.

Nielsen suggests that these user-identified weaknesses often present product developers with opportunities for improvement so that the users can better accomplish the associated tasks (Nielsen, 1993, pp. 75-77).

For example, assuming that the developers of the aforementioned printer were conducting a task analysis, their user observations probably uncovered more than just the three major steps. Perhaps one user had to replace an ink cartridge in the printer but neither the printer nor the instruction manual included information about how to insert the cartridge. Most likely, the user was not able to complete the major task of printing because of a sub-task that the developers had not considered. Consequently, the developers should have generated an objective similar to "A user should be able to insert a new ink cartridge into the printer without having to ask for help." Once this objective was established, the developers should have offered a solution, such as putting a diagram of how to load an ink cartridge underneath the printer's lid.

Not all objectives stem from user complications. For example, perhaps the developers simply wanted to test the effectiveness of the printer's instruction manual. An objective similar to "A user should feel more confident about printing after reading the instruction manual than he/she did before reading the document" could be tested through post-test user surveys and interviews. Or, maybe the developers wanted to find out if the users could easily locate the help desk contact information. An objective similar to "A user should be able to quickly (within a specified period of time) locate the help desk's telephone number within the instruction manual" could be tested by asking the user to find the help desk information and timing how long it takes the user to complete the task. Identifying tasks, followed by identifying testing objectives, helps developers improve the quality of their task analysis data.

Documenting Task Analysis Data

As developers document task analysis data, they must try to identify all user tasks and test objectives, whether major or minor, to develop a product that satisfies users

at all task levels. Typically, user tasks and test objectives are recorded in chronological order and in a hierarchal format. For example, a task analysis data sheet should list the items that follow.

- The major and minor outputs produced throughout the process.
- The major tasks that must be completed, presented in the order of their occurrence and/or importance.
- The sub-steps that are required to complete each major task, including any special skills/knowledge required.
- Major and minor user goals.
- Instances where users failed to accomplish tasks and/or achieve test objectives, if applicable (Callahan, 1985, pp. 14-15).

Task analysis data can be gathered through many methods, but five of the most common strategies are discussed next in "Collecting Task Analysis Data."

Collecting Task Analysis Data

Because every product/project varies, there is more than one way to conduct a task analysis. Many practitioners combine several strategies in an effort to capitalize on the strengths of each. These strategies, which are further discussed in "Collecting Data From a Standard Usability Test," can also be used to gather information during a usability test. Here they are discussed as they apply to task analysis only. Common task analysis strategies include, but are not limited to, protocol analysis, focus groups, scenario building, questionnaires, and on-site observation.

Each of the strategies described in this section are discussed in the context of following task analysis case: A website designer was developing a site that would function as an on-line instructional guide to teach users how to burn music CDs. The developer needed to learn the tasks associated with CD burning to ensure that users would consider the website effective.

- **Protocol analysis allows developers to observe users interacting with a product.** Often the users are asked to talk aloud while they use a product (Sullivan, 1989, p. 260). To apply this strategy, the website developer asked people that already knew how to burn CDs (subject matter experts) to progress step-by-step through the process so that the developer could observe and take notes.

- **Focus groups employ a moderator who leads a discussion about a product with a group of participants.** The developers observe a discussion of the product from a non-disclosed location, such as behind a two-way mirror (Dumas & Redish, 1993, pp. 44-46). To apply this strategy, the website developer hired a test monitor and invited both subject matter experts and people who did not know how to burn CDs to the discussion. The moderator explained the purpose of the website to the group, and the participants were asked to describe the tasks the site should include to ensure its effectiveness.

- **Scenario building is a technique that allows test participants to create a context for the product.** Participants are asked to envision the product as it would be received and used by a user (Nielsen, 1993, pp. 99-100). To apply this strategy, the website developer asked the same participants who attended the focus group to try to imagine how and where users would interact with the website, i.e., in their homes, at their places of work. The developer posed broad questions like, "What background knowledge will the users need to burn the CD?" and "What materials will the users need to complete the task?"

- **Questionnaires usually consist of questions that focus on detailed information about tasks and objectives.** This strategy varies from protocol analysis because it provides the participants with anonymity and usually lends itself to more candid responses than interpersonal discussion (Callahan, 1985, p. 4). To apply this strategy, the website developer asked subject matter experts to create a written list of all the tasks that must be completed to burn a CD, including the potential complications of each task, and the associated objectives.

- **On-site observation provides an excellent opportunity to view a user interacting with a product in an actual setting**. On-site observation is typically conducted in the user's home or at their workplace. It is wise to observe especially effective users and their strategies or "workarounds" as hints to improving an existing product (Nielsen, 1993, p. 75).

  For example, Procter and Gamble (P&G) has studied the use of Tide® laundry detergent through on-site observations. When P&G conducted the in-home research, they noted a problem with the powder detergent; some users felt that it was necessary to stir the detergent so that it was thoroughly dissolved. Consequently, some of the concerned users kept a large stick beside the washer so that they could stir (unnecessarily) the detergent before each wash. The users' "workaround" was something that P&G scientists/developers had not considered before they conducted task analysis.

  Another example of the on-site observation strategy is the website developer who observed subject matter experts in their homes and/or at their places of work to look for workarounds or especially effective methods of CD burning. The developer also observed people who did not know how to burn a CD to better understand their comfort level with computers and their associated background knowledge. After conducting the five task analysis strategies, the developer had learned much about the site's potential users and about the tasks associated with burning CDs.

Defining Participatory Design

Participatory design methodologies further the idea of developing a user-centered product by advocating the inclusion of the user in the design of any new product so that it accurately reflects the expectations of its target audience. Ideally, participatory design consists of a workshop(s) where users, product developers, and management/marketers come together to develop a product that incorporates the ideas of each group.

Participatory design has evolved so that its primary purpose is to increase communication between product developers and users. Originating during the 1960s in the Scandinavian countries, it grew out of the realization that traditional methods of product design were deficient at introducing new technologies in the workplace and factory floor. In an attempt to increase workplace productivity and efficiency, Scandinavian business owners generally agreed that opportunities for increased worker participation were necessary ("Participatory Design History," 1998).

The participatory design movement spread to America, but its focus eventually changed from worker/manager centered to developer/user centered. America's technology-driven economy depends upon the effective dissemination and translation of information from developer to user and vice versa. Consequently, participatory design has gained respect because it initiates dialogue between the developer and the user, helping to reduce the risk of user dissatisfaction.

Communicating Through Participatory Design Workshops

Participatory design workshops require communication: a dialog between developer and user. Given that the thoughts and ideas of an individual are influenced by their experiences and that no two people have shared exactly the same experiences, individuals are typically able to provide some original insight in any design situation. In participatory design workshops, users often ask questions and raise concerns that product developer never would have considered. Through this group dialogue, the developers are exposed to ideas and possibilities that they probably could not generate on their own. Participatory design workshops are a communication medium that allows developers to view the product through the eyes and experiences of users (Nielsen, 1993).

Applying Participatory Design Strategies

Participatory design practitioners apply many different strategies to product development. However, according to the corporate website of Information & Design,

a usability consulting company in Melbourne, Australia, many of these strategies have similarities, including how to select participants and plan the workshop.

Selecting the Participants

Every participatory design methodology should focus its efforts on recruiting participants who are representative of the product's target audience. Studies suggest that workshops with more than 15 participants are not as effective as workshops that limit the number of attendants to eight or nine. Generally, participatory design workshops should include the following people.

- A minimum of two and a maximum of 15 representative users. Include as many as necessary (a maximum of 15) to adequately represent the audience.
- One of the product's developers; preferably, all of the developers would attend so that each understands the users' ideas.
- One businessperson; possibly a manager, CEO, or marketing representative to remind the participants of the corporate budgets, timelines, and objectives.
- One unbiased facilitator who attends the meeting to mediate problems and keep the conversation on track.

For more information about selecting the workshop participants, see "Choosing the Test Environment and Participants."

Planning the Workshop

After participants are selected, planning the workshop is the next step in participatory design. Plans are often sketchy and should be. An outline format works well because conversation-based workshops will probably not adhere to a strict plan. The following is a recommended plan for a participatory design workshop divided into three sequential sections: beginning, middle, and end.

- **Beginning the workshop**. During the beginning of the workshop, introductions are made and background product and/or usability information is disseminated.

- Introduce participants. Typically, the participants will introduce themselves, but a facilitator should initiate this process if the group seems shy.
- Present the users with background information about usability to explain the workshop's purpose.
- Explain specific workshop objectives and expectations. For example, a workshop objective might be to improve upon an existing product to reduce complaints and store returns.

- **During the workshop.** During the workshop, participants discuss product issues and workshop goals and develop scenarios to discuss in small groups.
  - Identify issues with the product (if the product already exists) or with the development of a new product.
  - Explain the pre-determined goals of the workshop but allow participants to elaborate on or dismiss goals as they see fit.
  - Introduce scenarios so that the discussion focuses on the product's users. Ask the participants to explain their own experiences with the product or a similar product.
  - Split the group into smaller groups and allow them to work independently on solutions to the scenario(s).

- **Ending the workshop**. At the end of the workshop, the participants present solutions and discuss whether the objectives of the workshop have been met.
  - Reconvene and allow individual groups to present their ideas so that all groups can discuss advantages/disadvantages.
  - Decide how to use the remaining time most effectively depending upon the outcome of the first small group session.
  - Review the workshop objectives to see whether they have been met before allowing everyone to leave.
  - Document the workshop outcomes ("Information & Design," 2002).

Recognizing the Limitations of Front-end Analyses

Despite the value of task analysis and participatory design, limitations occasionally steer businesses away from conducting them. Richard Farson, an organizational communications expert, discusses the following limitations to front-end analyses (or "participative management" as he refers to it) from a businessperson's point of view.

- Time and cost factors tend to scare organizations away from front-end analysis. Further, front-end analysis often has to be repeated, which causes production deadlines to be pushed back. In addition, participants generally expect to be compensated for their time, usually in the form of money.
- Managers and developers tend to resist and/or resent front-end analysis because there is a fear that they are opening themselves up to criticism and attack from the users. Front-end analyses occasionally stir anxiety among developers because it can mean that managers consider their work to be lacking in some way.

Front-end analyses depend on trusting the group. It means investing in and relying on the participants to agree on a workable solution. Some managers and developers do not have the time or the patience to develop that kind of trusting relationship.

Even when front-end analyses are conducted, the activity does not guarantee that an extremely successful product will be developed (1997, pp. 77-80).

Benefiting from Front-end Analyses

Even with its limitations, most product-driven organizations can benefit from front-end analyses. For example, The CoHousing Company, based in Berkeley, California, is a construction company that uses front-end analyses to develop community living homes that reflect the wants and needs of their customers. Specifically, they "…work closely with their clients through a unique participatory design process that allows the group to build the community, not brick by brick, but decision by decision." First, The CoHousing Company shows potential customers "An Introduction to CoHousing" slide presentation. If the customers remain interested, the firm invites them to participate in a "Getting-It-Built" workshop that includes the customers in

organizing, planning, and designing the community housing. The CoHousing Company's practices have proven successful; they have already built over 35 different communities throughout the United States ("The CoHousing Company," 2002).

As the success of The CoHousing Company shows, front-end analyses provide developers with coveted consumer reviews before the product is marketed; these reviews increase the quality of the developers' work and enhance the reputation of their company. But most importantly, the users benefit because their investment in the early stages of product development helps to ensure the product is designed to satisfy their wants, needs, and expectations.

Preparing a Standard Usability Test Procedure

There are several ways to assess a product's usability and there are several different types of tests that can be performed to gather data that meets test objectives. But regardless of the product being tested and the type of test that will be performed, the procedure for developing the test should be the same. The standard usability test format has elements that every usability test should include.

Developing the Test Plan

The first step in conducting any usability test, is drafting a test plan. The test plan states the purpose of the test, the problem, and specific objectives that the usability test should achieve. The usability researcher should consult with managers and product developers to identify and agree upon the goals of the usability test. The test plan should address their different concerns from the start, so that conflicts over the test design and future test results can be minimized.

Rubin (1994) suggests that the usability test plan include the following elements.

- Target user profile: a description of the characteristics of potential users of a product.
- Test design: the types of tests that will be administered, procedures used to conduct those tests, and the instruments or procedures that will be used to collect data.
- List of tasks to be tested: a description of the tasks that users will be asked to perform and the way those tasks will be presented. See "Choosing Tasks for the Usability Test."
- Required test environment and equipment: a list and/or description of the space and equipment needed to perform the test. For example, the test might include videotaping users as they use a product or might require a computer.
- Role of the test monitor: tasks that the person trained to observe and record meaningful data should perform.

Choosing Tasks for the Usability Test

The usability test assesses different usability aspects of a product, including unpacking and installing (especially for electronics equipment), as well as basic and advanced use of the product.

The following are some aspects of the product that can be tested.

- Self-evidence of the interface (how much the user can use the product without having to read instructions).
- Relative ease with which users perform certain tasks with the products.
- Benchmark values that indicate the speed of performing certain tasks and other quantitative measurements.
- Frequency and success of accessing documentation and/or online help in products where that is an option.
- Understandability of instructions.

Types of Tasks in a Usability Test

Tasks that are chosen for the usability test may fall into one of three general categories: performance, location, and understandability.

- Performance tasks (or tests) ask test participants to use a product's functions to complete a task, such as successfully setting up a video recorder. Performance tasks can also test whether participants can understand and correctly implement instructions in product documentation, such as software manuals, instructions for assembling home furniture, and electronics installation and operation manuals. The focus of performance tasks is on the correct execution of the task to get the results that the user wants or to achieve the goal.

- In a location test, participants are asked to find something in or on the product – an on-screen button in a software application, for example, or a switch on a mechanical device. Location tests can be used for reference materials such as

legal documents (for example, taxation handbooks), product manuals, and online help systems (Anderson, 1998), information in documents, online help systems, or websites. The test monitor will record the participants' success and possibly the time it takes for them to locate the information.

- Understandability testing is used to test instructions or documentation. Understandability tests determine whether instructions are easily and accurately comprehended by the reader (Anderson, 1998). Usability researchers who conduct this type of test might ask participants to read a document, and then ask them questions about what they read. Understandability tests are useful for written records such as legal consent forms or waivers used in hospitals. Writers need to test these documents to make sure that those who sign understand their legal rights and what it means for them to sign.

  The following is an example of an understandability test: A reader testing the understandability of instructions for operating an expensive piece of equipment may be asked, "If you push this lever when the alarm goes off, what will happen?" Readers may be asked to answer in their own words or to choose from several items in a multiple choice-type list. Technical writers may also use true-or-false questionnaires to determine the understandability of their document.

  Usability researchers can use understandability tests in lieu of performance tests to avoid high expenses, especially if the product requires very expensive materials to produce. Understandability tests may also take the place of performance tests when dangerous materials are being tested and when it is difficult to obtain properly trained participants for the test.

Factors to Consider in Selecting and Prioritizing Tasks

In planning a usability test, usability researchers carefully select tasks to reflect the goals and objectives they set during the planning stage of the product. Dumas and Redish (1993) describe four criteria that aid in task selection.

- **Tasks should be actual tasks that users will perform with the product or for which the product is intended.** These tasks are logical choices for a usability test. The test can focus on selected new or modified product features, as well as frequently performed tasks and tasks performed under special circumstances and environments.

- **Tasks should focus on areas where potential usability problems are likely to occur.** Product developers usually have a good idea where the usability flaws in a product may lie and can be good sources of information when selecting tasks to uncover major problems in the design. However, product developers may not be attuned to some basic problems that users might have with the product.

- **Test tasks should present problems that have been past concerns.** Usability researchers may also draw from concerns and experiences they have encountered in their past work to select common tasks that have been problematic for users.

- **Test tasks that may result in catastrophic consequences or essential test items.** Usability researchers should include tasks that, if performed incorrectly, could result in catastrophic events that would destroy the product or render any work unrecoverable.

Usability researchers should also consider including special tasks performed under pressure or under some unique circumstances (Dumas & Redish, 1993). An example is use of the product by those who are colorblind (Nielsen, 1993). Environmental factors (such as lightning, noise, and heating) for products that require or may need to be used under special circumstances should also be considered in designing the test (Law, Barnicle, and Henry, 2000).

Time and practical constraints prevent every possible task from being included in the usability test. Thus, tasks need to be prioritized in order to determine which can be included. The table that follows shows four suggested factors to consider when prioritizing a task list (Rubin, 1994).

Table 1. Prioritizing task lists.

Priority	Task Description
Frequency of task use	tasks that are performed 75-80% of the time
Criticality of the task to product operation	tasks that, if missed or incorrectly done, lead to damaging consequences
Vulnerability areas of the product	tasks identified to have potential usability problems
Readiness of task functions	tasks for which the necessary product features and documentation have been incorporated in the prototype in time for testing

Presenting Tasks

Tasks to be performed in a usability test are most often presented to participants in scenarios or via field testing. Scenario building, which gives participants a way to envision the different situations in which a user might use a product as described in "Conducting Front-end Analysis: Task Analysis and Participatory Design." Likewise, usability researchers find scenarios to be an effective way to present tasks to test participants in order to simulate actual use of the product being tested.

Field testing is another option that allows usability researchers to observe test participants in their own environments.

Using Scenarios

Scenarios are "situations in which the task is embedded in a reasonable and very short story" (Dumas & Redish, 1993, p. 172). Instead of merely describing a task to be performed, such as "set up the video recorder to record tomorrow's show on Channel 18 at 3:30 p.m.," scenarios place the task into a context that reduces the test's artificiality, making the conditions in which the usability tests are performed

more realistic. The following is an example of a scenario for the performance task described above.

> "You have just made an appointment to see your doctor tomorrow. However, you later realize that the appointment is at the same time as your favorite TV program on Channel 18 (which airs at 3:30 p.m.). Set up the video recorder so that it will automatically record the program while you are out."

Location tasks can also be administered and evaluated in the context of a scenario. For example, the usability test may include only items such as, "Find the page that tells you how to use the time-cook feature of the microwave oven." Usability researchers can and will sometimes embed a task into a scenario. An example follows.

> "You are cooking several dishes for a party: one dish is in your microwave oven and two are on the stove. The microwave instructions for your recipe specify a four-minute cooking time on low power, then five minutes on medium, followed by two minutes on high. You want to program the microwave so that you can be free to concentrate on the dishes cooking on the stove. Find the page in your microwave guide that contains the programming instructions you need."

In whatever way scenarios are used, writing a good scenario requires careful attention to detail. A good scenario is short and unambiguous. It addresses the needs of the tasks to be tested but does not give clues about how to do them. Scenarios should be written with the user, and not the product, in mind.

For example, a user may not understand terms specific to the product. Thus, terms in the scenario should not be based on particular product functions, such as what a button does (e.g., "append"). Product-specific terms may serve as clues to the test participants. For example, if the participants are given a task to append a file, they may press the "append" button, knowing it is for the function specified in the test but without really understanding what function it is. This action will give the impression that the test participants understood the task they just completed.

Scenarios should reflect the test participant's experiences in order to be effective. For example, when scenarios are written to include more than one task in the

usability test, tasks should be sequenced in the order they are usually or naturally performed.

Written scenarios may be distributed among the test participants, or be may performed all at once by each participant. If the test monitor(s) need(s) some time in between scenarios to record results and observations, scenarios may be distributed to the participants one at a time. Scenarios may also be delivered orally, as in role-playing, to enhance the realism of the scenario. Either way, scenarios should clearly indicate when the task is over (Dumas & Redish, 1993; Rubin, 1994).

Field Testing

If the environment in which the product to be tested is particularly important in assessing usability, a field test may be conducted. This on-site test involves direct observation of the user in his or her natural working or home environment. Tasks may be given as described in the usability test, or the test monitor may observe while the user performs usual tasks with the product. To approximate natural conditions as much as possible, there is very little interaction between the test monitor and test participant in the field test. In addition to providing a realistic setting, an advantage of the field test is that it can reveal problems of integrating the product into the workplace (King et al., 1995).

Deciding on the Type of Data to Collect and How to Report It

The test plan should include the formats for presenting the results of the usability test to all appropriate audiences, such as managers and product developers, as well as the usability test format report. The test plan should also include pre-determined criteria by which to compare results. For example, the test might establish five minutes as the time it should take a user to complete a task. End results can be compared to the pre-defined time to determine whether the user succeeded or failed to perform the task, and thus to come to conclusions about the usability of the product. "Analyzing and Interpreting Data," has more information about identifying tasks that do not meet criteria during testing.

Keeping the end goal of a usability test report in mind during test plan development helps to ensure that the usability test will contain the necessary elements to answer the questions and achieve the objectives of the usability test. "Presenting the Results" has details on creating the usability test report.

Preparing for the Usability Test

The test plan serves as the foundation preparing for the usability test. Once the test plan is drafted, the orientation script for participants and the test script for test monitors can be written.

The orientation script should introduce the test and its purpose, inform participants of their rights and provide an overview of the tasks participants will be asked to perform and how they will be expected to respond, such as thinking aloud or completing a questionnaire. It should also include an introduction of the test monitor and his or her role with the company conducting the test or the product being tested.

The test script is a script the test monitor will use to conduct the test. It includes instructions that the test monitor will give to participants to perform tasks and may have a place for the test monitor to make notes or record results.

At this time, arrangements can be made to select and recruit test participants and to reserve any equipment and testing venue required. (See "Choosing the Test Environment and Participants.") The checklists that follow contain items suggested by Rubin (1994) that need to be accomplished as the test date draws near.

Checklist #1: Test Preparation

Approximately two weeks before the test, the usability test itself should be tested. The first two items on the list are essential in making sure that the usability test itself does not contain any items that test takers may find confusing and that will lead to unreliable results.

Task	Complete
Perform a self-test on each part of the test to ensure that the test addresses objectives and that the tasks and questions that will be asked of test participants are complete and clear.	
Perform a pilot test to detect flaws in the usability test before administering it to a large number of test participants. Employees of the product manufacturer who represent typical users may perform the pilot test.	
Revise the test plan based on the self and pilot tests.	
Reserve any monitoring equipment required, such as videotape recorders, and the room to be used during testing. The test monitor should check equipment at this stage to make sure it functions properly.	

Checklist #2: Pre-Test

Complete the second checklist the day before the test.

Task	Complete
All video equipment is set up and ready.	
The products to be tested are in place and functioning as designed.	
Materials for test monitors have been assembled, such as the orientation script, test script, and data tracking materials.	
All written test materials for participants have been assembled, including consent forms, any necessary questionnaires, instructions and the like.	
Status of the participants is checked. It is possible that some will cancel; thus, replacements should be found for them.	
Double-checked the test environment and equipment.	

Checklist #3: Test Monitor Activities

A third checklist details the activities of the test monitor, which will be specific to each test, before, during, and after the test. Usability researchers should devise their own checklist for the test monitor(s).

Selecting the Right Test Monitor

Test monitors play a pivotal role in usability testing because they help the test developers and product developers meet the test objectives. According to Nielsen, "To reach the goal of making technology truly suited for humans, the world will need about half a million new usability professionals over the next 20 years. The sooner their training begins, the better off we'll all be." ("Becoming a Usability Professional," 2002).

"Ensuring the Rights of Participants," discusses the additional role of test monitors as the person responsible for overseeing the usability testing and ensuring the rights of those involved.

Professionals Who Monitor Usability Testing

A number of different types of professionals can successfully monitor usability tests. Technical communicators and other professionals such as human factors and marketing specialists possess the knowledge and skills that are needed to properly monitor usability testing (Rubin, 1994). The table that follows lists types of professionals who are well-suited to be test monitors and the skills they posses that make them successful in that role.

Table 2. Types of professionals qualified for usability testing.

Type of professional who monitors usability testing	Why this type of professional is a successful test monitor
Usability Engineering Specialists	▸ can anticipate problems and can recommend solutions ▸ can suggest a testing environment most conducive to maximizing participant performance
Technical Communicators	▸ are user advocates ▸ are focused on writing, designing, and creating information to make it more user-centered **Note**: When choosing technical communicators, one must ensure that they know how to monitor usability testing and that they are able to be objective.
Human Factors Professionals	▸ usually have a degree in psychology or engineering and have extensive experience in testing, methodology, and people ▸ can evaluate the quality of usability tests and recommend any changes needed ▸ are often skilled at designing usability tests
Marketing Specialists	▸ are consumer advocates who are concerned with improving products for customer satisfaction ▸ may be intensely involved with the test product **Note**: Involvement with a test product does not disqualify marketing specialists from test monitoring as long as they are able to be objective.
Colleagues	▸ have product knowledge that can be useful during testing ▸ understand the company's goals for testing

Type of professional who monitors usability testing	Why this type of professional is a successful test monitor
External Consultants	▸ can offer a great deal of expertise in conducting usability tests ▸ generally are objective about the company's products

Qualities of Successful Test Monitors

It is important to evaluate qualities when reviewing candidates who will monitor usability testing. The table that follows lists qualities of successful test monitors and the reason each quality is important to testing.

Table 3. Qualities of successful test monitors.

Quality	Why this quality is important
Effective communicator	▸ must communicate with team members, test participants, developers and others involved in the development process ▸ must be able to explain suggestions for improvement and why such changes are necessary ▸ must communicate critical to effectively to present test results
Attentive listener	▸ must listen to participants' voice inflections to determine frustration, satisfaction or indifference ▸ must eliminate any bias that may cause them to listen selectively
Fast learner	▸ must become familiar with the test product ▸ should also learn the product history and concepts surrounding the product ▸ needs to learn about past problems, market share, and competition

Quality	Why this quality is important
Flexible	▸ should be open to any participants' suggestions or ideas and should expect the unexpected ▸ must be able to alter, or in extreme cases, stop the test if problems arise and participants are not able to perform
Good memory	▸ must be able to recall participant actions and behaviors for team members or developers who may not examine the recordings
Long attention span	▸ must be attentive during testing sessions, which can be long and repetitive, and which may have times when participants are reading, thinking, or relaxing ▸ must observe participants at all times, during all sessions, even when different participants are performing the same tasks
At ease with ambiguity	▸ must be able to deal with gray areas ▸ Bust be able to deal with observation inconsistencies that are not a part of the original test plan and that could make data inconsistent and inconclusive; test monitors must be able to indicate when these situations arise

Choosing Test Monitors

It is important to choose the appropriate test monitors to manage usability testing. According to Nielsen, a test monitor should have knowledge of the product, the test, the experiment, and the methodology. For example, if a software test requires a high level of system knowledge, it may be preferable to use a system developer rather than a test monitor who may not have as much knowledge about the system (1993). Depending on the type of test and the results desired, it is crucial to choose a test monitor who has experience in the types of methodologies necessary for conclusive data.

Implications of Choosing Test Monitors

According to Rubin, a test monitor is the captain of the team. He or she must be knowledgeable, competent, efficient, effective, and confident in monitoring the participants. If the captain is not able to manage the project, the test and test results can be skewed or even negated. Finding the appropriate test monitor can make the test run smoothly and ultimately make the test a success or a failure. Failure to choose the appropriate test monitor could potentially "make or break the test" (2000). An incompetent monitor could waste money, time, and the participants' efforts.

According to Nielsen, a bad usability specialist will give ambiguous results. For example, the test monitor may report, "User One liked this, but User Two did not." This statement is not helpful to the product developers; a good usability professional combines observations among many participants and arrives at a conclusion that is helpful to product developers ("Becoming a Usability Professional," 2002). Because test monitors are a critical part of usability testing, it is important to be aware of the consequences of selecting a person to perform those duties.

Behaviors of Test Monitors

Test monitors' behavior is important to consider when reviewing candidates for the role. Behavior can ensure that the test runs smoothly and help increase the participants' confidence to perform. According to Rubin there are some key behaviors that are valuable in good test monitors (1994).

Table 4. Appropriate behaviors for test monitors.

Behavior	Why this behavior is crucial
Assist rather than lead	▸ should not ask leading questions or make leading suggestions ▸ should be aware of tone of voice, facial expressions, and body language; these must remain neutral environment so they do not influence results
Be aware of participants	▸ should observe consistently to decrease the risk of missing important actions and behaviors ▸ should be aware of all participants and should not be overly concerned about data collection
Connect with all participants	▸ should connect with all participants, regardless of age, gender, size, ethnicity, and demeanor of participants ▸ must help foster friendly relationship between participants to help maximize participants' performance
Be flexible with test plan	▸ should know when or if it is necessary to deviate from test plans when test objectives are not being met so that time, money, and participants' efforts are not wasted

Test monitors should not act on test results prematurely and should form conclusions only after data have been compiled. Doing so will help keep the product development team from making major changes prior to data compilation. If test monitors appear too knowledgeable, participants may ask questions rather than perform tasks themselves. Participants may feel embarrassed or intimidated if test monitors appear to have all the answers. Test monitors should assure participants that the actions they perform are helpful.

Professionals who monitor usability testing can play many roles. It is important to choose test monitors who have knowledge and experience and who possess qualities for successfully monitoring usability testing. Choosing a qualified test monitor may increase the potential for meeting test objectives that can later result in conclusive data that can be used to improve product usability.

Choosing the Test Environment and Participants

Usability testing can be conducted in various types of settings, but because certain surroundings can affect test participants' concentration and actions, test environments should be well chosen and carefully arranged to avoid biasing results and causing distractions. Likewise, test participants should be selected carefully; they should represent the intended audience as closely as possible to provide the most realistic test results.

Setting Up a Test Environment

Testing environments may range from users' actual work environments (the field) to a controlled laboratory. According to Thomas Barker, author of *Writing Software Documentation* (1998, pp. 202-203), "Your best chance to learn about actual use in the context of the user's work and information environment comes from field testing. But, if one is available, the laboratory offers a greater degree of control and testers may find it more convenient."

Although field testing is ideal, it is not always possible. Regardless of where testing takes place, when setting up the test environment, it is important to remember that participants' surroundings can affect how they perform during a test and how they judge a product.

Testing environments need not be elaborate; it is more important that they are functional and that they support test-taking activities. For example, when testing a piece of software documentation, test participants only need to be provided with the appropriate computer equipment, instructions, and a distraction-free environment. A quiet office or an existing computer lab would suffice.

Considering Test Environment Factors

Test developers must consider a variety of factors when choosing or setting up a test environment. For example, will noises, pictures or decorations, or other features in the test area influence or distract test participants? Although some companies furnish their test areas with elaborate, high-tech, and visually stimulating décor, such distractions can draw participants' attention away from the test and possibly cause bias or partiality.

To attain accurate and unbiased results, testing should be conducted in an area that contains neutral wall colors, few decorations, minimal noise, and comfortable seating. If possible, testing should take place during a time of day in which participants are unlikely to be preoccupied by physical distractions such as hunger, sleepiness, or restlessness; mid-morning or mid-afternoon tests usually work well.

Test participants should know how long a test will take so that they can plan their daily activities accordingly. During long tests, test participants should be allowed to take breaks. Studies show that after a long period of time, test participants start to lose focus and become more easily frustrated when they encounter problems.

Presenting Test Materials

When presenting test participants with the product to be evaluated, test monitors should remain neutral in both their actions and dialogue. They should also emphasize that test results do not reflect upon the participants in any way; the product is being tested, not the participants' abilities.

Consider Karen A. Schriver's comments on testing written instructions: "When instruction guides mislead readers, as they often do, readers tend to blame themselves for their confusion. Contrary to popular belief, young people and old, males and females, are about equally likely to blame themselves for the troubles they experience" (1997, p. 222).

Participants should be assured that problems they encounter are more likely due to a faulty product rather than to their actions during a test. After testing begins, test monitors should refrain from interacting with participants unless absolutely necessary; they should try to be as unobtrusive as possible. (See "Selecting the Right Test Monitor.")

Choosing Test Participants

Effective usability testing depends a great deal upon test participants. Selections should be based on a user profile that details characteristics of potential users of the product. The number of participants is also significant; it depends on the type of testing being conducted and can affect the validity of results. Finally, how test participants are approached and asked to participate in a test can affect the test's outcome. Thus, selecting test participants should be done carefully.

Developing User Profiles

One of the first steps in planning a usability test is developing a profile of actual or intended product users. According to Dumas & Redish, there are two types of characteristics test developers should consider when developing user profiles.

- Characteristics that all users will share.
- Characteristics that might make a difference among users (1993, p. 120).

Identifying Desirable Characteristics

In most cases, test participants should closely represent the intended audience. In some instances, test developers may want to consider choosing a combination of participants—some who represent the primary users and some who might represent a secondary group of users.

For example, when testing the usability of a high school's website, test developers would want their test participants to comprise a combination of students and

faculty—the primary audiences for the website. But test developers might also consider testing students' parents, who could represent a secondary group of users.

Determining the Most Important Characteristics

In the beginning stages of developing the user profile, test developers should brainstorm relevant characteristics that the intended audience might share. From that list, test developers can determine what significant characteristics are unique to the audience. Test developers can determine which characteristics are most prominent in the intended audience and narrow the list to those that are most important. They should keep in mind that "specific, relevant experience and motivation matter more for understanding differences in how people interact with products than do demographic factors like education, income level, and age" (Dumas & Redish, 1993, p. 121). The following are some factors to consider when selecting common characteristics.

- Work experience.
- General experience with the product.
- Experience with similar products (Dumas and Redish, 1993, p. 122).

Determining Experience Level of Participants

When developing user profiles, test developers should consider the level of experience test participants should have with the test product or similar products. In a discussion of usability testing for software applications, Nielsen says, "almost all interfaces need to be tested with novice users, and many systems should also be tested with expert users. Typically, these two groups should be evaluated in separate tests with some of the same and some different test tasks" (1993, p. 177). The same holds true for other products.

In the planning stages, test developers must determine whether they should primarily test novice users or whether including expert users in the testing phase would provide them with more accurate feedback about their products. Rubin states

that all usability tests should contain one or more "least competent users" (LCUs). He defines an LCU as "an end user who represents the least skilled person who could potentially use the product" (1994, p. 129).

Rubin (1994) provides the following fully developed user profile and includes all of the elements discussed in this section (pp. 108-109).

USER PROFILE

A total of 10 participants will be tested during the week of September 25, 1995 at the Market World facility in New York City. Two participants will be tested per day. Two alternative participants will be acquired in case one of the participants is unable to attend at the last minute. The participants will be divided according to background as follows:

Two participants who are both computer and printer novices. While these participants are not necessarily the intended buying audience, they represent the least competent user who will use the product. If they are able to perform, that is a strong indicator that more qualified end users will also be able to perform.

- Eight participants who have previous computer and printer experience. Participants in this group will have the following characteristics:
  - Work with a computer on a regular basis or plan to purchase one within three months.
  - Have previously purchased 0 to 2 printers.
  - Have never purchased or used a laser printer on a regular basis.
  - Have limited font sophistication (i.e., do not use more than three fonts at a time).
  - Spend the majority of computer usage performing word processing tasks. Do not perform spreadsheet or desktop publishing tasks more than 20 percent of their computer usage time.
  - Do not require a wide-carriage printer, or one that can print multipart forms.

Figure 2. Sample user profile for usability test participants.

Selecting the Appropriate Number of Test Participants

According to Rubin, the number of test participants chosen depends on many factors, such as those listed below.

- The degree of confidence in the results required.
- The number of resources available to set up and conduct the test.
- The availability of the type of participants required.
- The duration of the test session.
- The time required to prepare for the test.

Time and resources may limit how many test participants can be recruited. The number of participants may also depend upon how statistically reliable the results must be. If the tester's aim is to produce consistent and precise data, test developers will need to select a large enough number of participants to allow for the generalization of test results to a targeted audience. If, however, test developers are just trying to expose as many usability problems as possible in the shortest amount of time, the general rule is to test at least four to five participants (Rubin, 1994, p. 128). Recent research shows that four to five test participants will encounter the majority of potential problems in a product.

When choosing the number of participants, test developers must also consider how many tests will take place throughout a product's development. If a series of tests are conducted, rather than just one, a solid degree of accuracy can be maintained with fewer participants in each test.

Using Incentives to Encourage Participation

Introducing the concept of usability testing to potential participants can be challenging because, in most cases, the testing will not directly benefit them. To account for potential participants' self-interests, some companies use enticements, such as monetary rewards or discounts, to attract participants. Test developers

should keep in mind that enticements might potentially bias the results of usability testing. Enticements should be offered with caution.

Protecting Participants

Once participants show interest and agree to test a product, test developers must take specific steps to protect participants' rights before, during, and after testing. It is the testing organization's responsibility to know, understand, and to do everything they possibly can to ensure that participants are treated ethically.

Ensuring the Rights of Participants

Organizations and individuals conducting usability testing have an ethical responsibility to treat test participants with respect. Various principles, laws, and responsibilities outline organizations' legal and ethical obligations regarding test participants (Burmeister, 2001). These legalities can be perceived as an opportunity for organizations to test their products with confidence and to improve the value of their products.

Principles and Federal Laws Concerning Human Participants

Principles and federal laws for using human participants in testing range from general to specific. Three primary areas directly relate to technical communicators and usability testing.

- The Belmont Report is a set of ethical principles for researchers that outlines the treatment of test participants.
- The Federal Register is a governmental law that protects all human participants' rights (Dumas & Redish, 1993).
- Federal law regulates accessibility and protects the rights of participants with disabilities, which may interest organizations that want to test specific products or have a target audience that includes these individuals.

It is important for organizations to know and understand government mandates, such as these, before conducting usability testing (Dumas & Redish, 1993).

The Belmont Report

The Belmont Report, drafted by the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research in the late 1970s, identifies three basic ethical principles for research practice: respect for persons, beneficence, and justice. "Basic ethical principles" refers to "those general judgments that serve

as a basic justification for the many particular ethical prescriptions and evaluations of human actions" (1979, page 1).

These principles suggest that researchers acknowledge and protect the rights of participants equally by upholding their obligation to secure participants' well being, thereby maximizing benefits and minimizing harm to participants (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979).

The Federal Register

The Federal Register's Notice of Proposed Rulemaking protects human test participants' rights. Through The Federal Register, the government acts as an advocate for participants to ensure they will not be harmed or mistreated for the sake of research or testing.

Federal Regulations for Accessibility

A recent United States mandate addresses the needs and rights of participants with disabilities. It requires hands-free, eye-free, and ear-free products and encourages organizations to re-evaluate the accessibility of both their products and their testing environments. The regulation requires accessibility in several areas.

- Access to public services or buildings cannot be refused to people with disabilities.
- Access to telecommunications equipment and service products is mandatory, when easily achievable.
- Access to electronic and information technologies purchased by the federal government is required whenever it is not an undue burden. (Law and Barnicle, 2001).

Rights of the Participants

Laws and regulations ensure participants' rights, and it is an organization's ethical obligation to respect those rights. But organizations may deal with people who are not familiar with usability testing or testing legalities. It is always incumbent upon the organization to inform participants of the following rights before testing begins.

- Minimal risk means participants will not be harmed in any way during testing.
- Informed consent is the participants' agreement to participate in the test and to know what will happen during the test.
- Comprehension is indication from participants that they understand how the test will be conducted.
- Willingness to participate means that the participant is free to participate or leave the test at any time.
- Participant privacy means that names will not be revealed.

Ensuring Minimal Risk

Minimal risk means participants should not be exposed to any harm – physical, mental, or emotional – that is greater than they would normally experience in their daily lives. It is always important for an organization to always gauge the risk involved in their usability test.

Though many usability tests easily comply with minimal risk, such as when participants are testing computer software or testing instructions on how to program a VCR, other tests could involve danger, such as instructions on how to jump-start a car battery. Testing these instructions could expose participants to hazardous materials, like sulfuric acid or to an explosion if participants do not understand what they have read. Organizations have a responsibility to decide whether to eliminate risks or to forego testing.

Ensuring Comprehension

Ensuring that participants comprehend test procedures and their rights is an important component of informed consent. Providing clear and complete information orally and on a consent form optimizes participants' comprehension. Discussing test procedures thoroughly is the key. Participants may be confused or feel nervous if discussion is rushed before testing begins (Dumas & Redish, 1993).

Obtaining Informed Consent

Informed consent protects the rights of both organizations and test participants. It also gives test participants some amount of control over what they will allow to happen to them. Before the test begins, test monitors should explain several aspects of the test – the test's purpose, testing procedures, participants' free will (they can excuse themselves from the test or a portion of the test at any time), and any potential risks to participants. Test monitors should also distribute consent forms, which provide the same details about the test.

The consent form is the test monitor's way of asking participants to state that they understand test procedures and that they will participate in the test. This agreement/contract is especially important when a company videotapes or audio tapes testing: participants must be made aware of the fact that the use of video and audiotaping for documenting test proceedings is common and will be part of the test in which they are participating.

The test monitor should state and the consent form should reiterate that the test monitor has "no personal stake in the [product] being evaluated." Emphasizing the test monitors' distance from the product may put participants at ease and may encourage participants to speak their minds without feeling as though they will be punished or judged (Nielsen, 1994).

It is also important that participants understand that it is the product being tested and evaluated, not them. Participants should also be reminded that they are an integral part of product development. Dumas and Redish suggest emphasizing the

difference with the following statement (1994, p. 207): "The purpose of this test is to make the product (computer program, manual) easy to use…. We want you to help us improve this product."

Garnering Willingness to Participate

Willingness to participate refers to participants' right to leave at any point during a test, including taking breaks. Dumas and Redish point out to test monitors, "When participants are considering withdrawing from a test, remain calm and neutral in your manner so that you do not unduly influence them" (1994, p. 207). Furthermore, coercing participants for their consent, or to continue with an ongoing test, violates ethical research principles and participants' rights.

Ensuring Participant Privacy

It is important to reassure participants that they have the right to privacy. For example, test monitors may offer the option of having participants' names omitted from a report of the test results (Burmeister, 2001). This is another way that test monitors can reiterate and stress to participants that only the product is being evaluated.

Responsibilities of the Testing Organization

The role of the test monitors can be tricky. They must not only consider how to best accommodate and inform participants about testing procedures but must also keep their organization's desire for product feedback in mind. While organizations and the individuals who represent them must uphold their ethical obligations to test participants, they must also protect themselves and their products.

Requiring Non-Disclosure

When a product is still under development, organizations usually require each participant to sign a form stating that the participant will not discuss information about the product with anyone. This procedure is referred to as non-disclosure.

While discussing non-disclosure with participants is paramount, participants may be concerned how to account for their time to family and friends. Dumas & Redish, (1993, p. 208) suggest one way to handle the situation: ". . . you can tell people that you were evaluating a new product (computer program, manual) by using it and giving your opinions of it, but you cannot tell them anything about the product . . ."

Ensuring Confidentiality

Oliver Burmeister distinguishes between the participant's right to privacy and confidentiality in *Rights of Participants* (2001). Confidentiality refers to how data about the participants will be stored. The Association for Computing Machinery Council's Code of Ethics and Professional Conduct specifically addresses limiting access to certain data. The code of ethics mandates that organizations ensure data are accurate and that organizations uphold their commitment to participants about how data will be used (Burmeister, 2001).

Seeking Permission

Many usability researchers recognize that organizations have different policies regarding the scope of permission they seek. Most often, permission is sought to videotape or audiotape participants during a test to evaluate test results and to educate product designers about usability issues (Dumas & Redish, 1993).

Understanding Participant Expectations

Because of growing cross-cultural interactions in usability testing, whether within the same country or during remote testing through the Internet, organizations need

to recognize the expectations of different participants. Burmeister points out that problems can arise when participants have different work practices or social class systems. Usability test procedures that may be encouraged in one culture, such as discussing a product's weaknesses, may not be considered socially acceptable in another culture. Burmeister offers one example of a usability test in Singapore where a participant broke down and cried (2001, p. 8). "A post-test interview revealed that the test participant's behavior was attributable to Eastern culture in which it is not acceptable to criticize the (developer) openly, because it might cause the (developer) to lose face."

Absolute Versus Discretionary Principles—The Gray Area of Usability Testing

In addition to upholding principles and laws for the rights of the participants, other discretionary principles concerning the rights and responsibilities of the test participant should be practiced depending on the context of usability testing (Burmeister, 2001). Burmeister states that "laws concerning the treatment of human participants vary between countries. Not only this, but such laws also vary between states of the same country (as is the case in the US)" (2001, p. 8). Therefore, it is important for organizations to recognize their state's and/or country's specific laws regarding test participants and pay particular attention to laws that may apply while conducting usability testing abroad.

By taking the proper measures to ensure legal and ethical practices, usability testing can help protect organizations from unnecessary lawsuits, thereby saving time and money involved in litigation. But it is also a matter of integrity and accuracy. The more companies adhere to proper procedures and practices, the more credibility their products will have.

But even though participants are treated ethically during testing, there are other risks outside the actual test environment that must be considered. Risks of violating ethical standards are inherent in offering incentives for participating, using

employees of a company to test its own products, and using an employee of a company as a test monitor for its own products.

Risk of Using Incentives

A primary risk of usability testing involves enticing potential test participants. Encouraging participation by offering incentives can affect participants' behavior and performance, because it can be construed as a form of coercion. For example, some technology companies, like Microsoft, offer participants software for their time (Microsoft Corporation, 2000).

Incentives themselves do not violate any particular principles, but there are appropriate ways to offer compensation. Burmeister suggests that test monitors offer payment or gifts up front so participants do not feel obligated to stay through the entire test. Other usability researchers suggest test monitors wait until after testing is complete to offer complimentary gifts (2001). The choice of when to offer an incentive lies with the organization.

Risk of Using Internal Participants

Using incentives to encourage employees of a company to test its products may influence test results, especially if the participants' superior will observe testing. Employees may feel they could jeopardize their jobs if they are critical of the product (Dumas & Redish, 1993).

Organizations can directly avoid this risk by seeking participant volunteers from the public or from different departments that are not related to product development.

Risk of Using an Internal Test Monitor

The test monitors' role in usability testing is extremely important because they potentially have great influence over the participants depending on how the test is conducted. There are no ethical laws barring product developers from acting as test

monitors, but product developers can be biased because they are involved in product development.

It is important that product developers who act as test monitors do not make participants feel inadequate, express their feelings to participants, or divulge too much information about the product that would skew results. Many usability researchers suggest that test monitors refrain from interfering with participants and allow them to discover solutions to problems on their own (Nielsen, 1994).

Collecting Data From a Usability Test

There are almost as many methods for collecting data as there are types of data produced in a usability test. Choosing a method depends on the availability resources and the testing needs of the product development team and usability researcher. The standard usability test produces the following types of data.

- Users' actions while performing tasks, including errors committed.
- Users' comments (such as expressions of frustration) and behavior (eagerness, concentration on the task).
- User preferences (such as the choice of one over several ways of performing tasks).
- Users' opinions about the product.
- Computer or equipment activity.

Data can be collected in several ways, and at different time periods: before, during, and after the test.

Monitoring User Activities During the Test

Test monitors can observe and record the actions and comments of test participants directly when there are relatively few test participants (one to two per test monitor). Direct observation entails a close interaction between the test participant and test monitor that can distract the test participant and affect test results. To remedy that situation, monitoring equipment can be used to view the activities of test participants from an observation room. In addition to eliminating potential distractions, using monitoring equipment also produces permanent records of the test, which are useful references that can be used to enhance the usability report and that can be consulted when resolving usability problems long after the test was administered. Two ways to monitor test participant activities are using audio and video recording equipment and data logging.

Audio and Video Recording Options

Equipment such as cameras, audio and video recorders, and cameraless video feeds can be used to monitor participants' actions and comments during a usability test. One or more cameras focused on the user-product interaction can record the actions of test participants. During the test, one or several test monitors record specific actions based on the live video feed that they see.

Equipment for cameraless videotaping can be used to monitor tasks performed on a computer (Nielsen, 1994). A computer, equipped with a graphics card and a *video out* terminal, can be connected directly to the *video in* of a video recorder and TV monitor. Test monitors can see every user interaction; they can view the onscreen events, and can record data, such as where test participants click the mouse and the windows or error messages that appear as a result.

Cameraless videotaping is a way to capture onscreen information, but it leaves the test participants out of the picture. The advantage to this method is that test participants are generally more comfortable or less self-conscious when there is no camera aimed at them. Cameraless videotaping is also less expensive than camera recording because cameras and camera operators are not needed. The disadvantage is that video transfer from the PC screen to a TV format often results in poor resolution.

Audio recorders are a simple piece of equipment that can assist the test monitor during interviews or record test participants' comments during a usability test. Audio recorders enable the test monitor to concentrate on other observational tasks instead of writing down comments.

Data Logging Options

There are several options for logging activities that happen during the usability test. Usually, test monitors record their observations into a database or on paper data sheets. They can also use a system that enables test participants to record their own results during the test. When usability researchers conduct computer-based

usability tests, they have the additional option of recording activities done through the computer interface, such as mouse clicks and use of function keys on the keyboard. Several data logging methods are available: data sheets and databases, self-reporting, and automated data logging.

Using Data Sheets and Databases

One method of data logging is completing data sheets, which are designed with performance objectives in mind. Data can be recorded on paper, or through an online database. A typical data sheet or database might include sections for the following elements.

- Tape time (or test time elapsed).
- Data type (for example, “error”, or “comment”).
- Checklist of actions to be completed by the test participant.
- Comments of the test participant or test monitor.
- Test monitor’s observations or qualitative description of the events.

Short (one- to two-) letter codes can be developed to facilitate quick entry of different data types on paper data sheets. Writing down time tracks from audiovisual recordings may produce useful referents for reviewing audio and video records for specific problem events.

The figure that follows shows a typical paper data sheet used to record events in a usability test for an e-mail program (Rubin, 1994; Dumas & Redish, 1993).

Usability Test of the Electronic Mail Program									
Participant #		**M** - menu error					**O** - online help		
Date:		**S** - select from list error					**H** - help desk		
Recorder:		**E** - other error					**F** - frustration		
Task:	Time	M	S	E	O	H	F	Participant's comments	Notes
Task 1: (Example: Log in)	start: stop:								
Task 2:	start: stop:								
Task 3:	start: stop:								

Figure 3. Sample paper data sheet.

If online data collection is used, a checkbox or radio button for selections may be provided to record results for each data type. Online data entry provides the possibility of computerized analysis, such as automated counting of error frequency.

Self Reporting

Self-reporting is a data logging method that is especially useful when personnel resources for data collection are limited. Test participants are asked to answer a questionnaire or complete a checklist after finishing each task. As with data sheets and databases, self-reporting can be done on paper or online. Paper reporting – questionnaires, in particular – is more disruptive for computer-based usability tests because it is less integrated with the participants' testing tasks (Rubin, 1994); thus

the questionnaires used should contain minimal numbers of questions (Dumas & Redish, 1993).

Automated Data Logging

Automated data loggers can be used for computer-based usability tests (Rubin, 1994). Several independent data logging programs are available. Another option is for product developers to design programs for recording keystrokes, commands, menu and help access, and some program functions accessed during the test. A disadvantage to this method is the large amount of data that needs to be sifted through before relevant information can be analyzed. An advantage is that automated data logging is useful for tracing exactly where bugs occur in a product.

Data Collection Methods

To complement the previously mentioned observation and monitoring methods, usability researchers can also administer questionnaires, use the thinking-aloud protocol to obtain user feedback, and/or conduct focus groups.

Using Questionnaires to Collect Data

Usability researchers can ask test participants to complete questionnaires before, during, and/or after a usability test, as a means of obtaining test data.

- Pre-test questionnaires are used usually designed to assess the participants' prior knowledge about the product before the test, their backgrounds, and their initial impressions of the product.
- Post-task questionnaires are given out during the test or upon completion of a task. A post-task questionnaire may be administered for any of the following purposes: 1) as a method for data logging, 2) to obtain immediate reactions to the test at critical points, and 3) to obtain a view of how test participants' perceptions change as they spend more time testing the product.

- Post-test questionnaires gather the test participants' opinions about the product after testing. In addition, questions that address specific concerns or areas of the test that caused confusion may also be included (Dumas & Redish, 1993; Rubin, 1994).

Using Thinking-Aloud Methods for Data Collection

Thinking aloud is a valuable method of data collection in usability testing. It has several variations, but basically, test participants are asked to verbalize their thoughts while performing the tasks. Comments made by the participants are often valuable complements to observed behaviors in the test. Thinking aloud can help participants communicate what they are feeling about a product and problems they may encounter while using it.

Some have argued that a disadvantage of this method is that the process of thinking aloud may slow the participants down while they try to verbalize their thoughts (Nielsen, 1994). However, Berry and Broadbent (1990) found that test participants who were asked to think aloud performed their tasks faster. The authors argued that the thinking aloud method actually helped the participants process information about the tasks they were asked to perform better, thus enabling them to complete them more efficiently.

The counteracting effects of using the thinking-aloud method, mean that interpretation of performance measurements, such as speed of task completion, must be done carefully. To permanently capture the data, it is recommended that test participants' comments be recorded using audio equipment or that observers or test monitors write comments during the test.

Using Constructive Interaction

Constructive interaction**,** a variation of thinking aloud, is a data collection method that may be used to alleviate the awkwardness participants may feel when they try to verbalize their thoughts. In constructive interaction, a pair of test participants talk to each other while performing the specified tasks. Because it is usual for two

people to talk while solving problems together, this method provides a more natural setting that encourages thinking aloud (Nielsen, 1994).

Using Retrospective Testing

Retrospective Testing is another variation on the thinking allowed theme. In retrospective testing the test monitor and test participant view and talk through the recorded video of the completed usability test. Participants' comments during retrospective testing are typically more extensive than think-aloud comments obtained during the test. Although this process can take a long time – at least twice the length of the usability test during which test participants are actually performing tasks – valuable feedback can be derived from test participants. The playback can be paused at critical points, allowing the test monitor to question the test participants in more detail (Nielsen, 1994). Retrospective testing is most effective when only a few test participants are available and usability researchers would like to get more information from them.

Using Focus Groups for Data Collection

Focus groups can provide valuable insight into the usability of products. They can be used in the exploratory stages of development or after the product has been used or tested.

In a focus group, a minimum number of participants (usually six) are brought together to discuss the product's concept and the usability issues that may arise when using the product. A moderator usually prepares a list of the issues to be discussed and sets goals for the information that the group will provide (Nielsen, 1994). When using focus groups, the evaluators need to be aware of the tendency to "group think." That is, members of the focus group may start to think alike as they spend more time with each other, thus limiting creativity (Morgan, 1997). Members of a focus group should be chosen from diverse backgrounds to maximize the information produced by the group. A videotape of the focus group's interactions provides a valuable record for the usability report.

Debriefing Test Participants

Usability tests usually end with a debriefing session, typically after participants have completed any post-test questionnaires. Debriefing, in this sense, is a process of asking questions to obtain more useful information from the test participants. The aim of debriefing in usability tests is to clarify and understand all problems, errors, and omitted steps that the participants may have encountered during the test. The test monitor reviews the events of the testing session with the participants, interviewing them about any problems they encountered steps (Rubin, 1994). A videotape of the testing session, if available, and information from datasheets or databases is particularly useful in reviewing events.

Analyzing and Interpreting Data

Data analysis in usability testing can be defined as transforming raw qualitative and quantitative data into aggregate results that can be used to make recommendations for improving the usability of a product.

Usability researchers conducting the study can easily identify trends and patterns (Campbell, 2001) using some forms of data. Other data may require more sophisticated methods of analysis.

There is no definitive method for data analysis because the same data may be analyzed in different ways. Usability testers should choose methods of analysis that are most appropriate for confirming or refuting test objectives.

There is also no definitive answer to the question, "Who should perform data analysis?" Many different types of professionals may perform the data analysis connected with usability testing. And regardless of the method of analysis used, it is important to always have more than one person look at the results so that no important trends or problem areas are overlooked (Campbell, 2001, p.14).

Defining Data Types

The two data types researchers analyze and interpret are qualitative and quantitative. Qualitative data is any non-numerical data, such as observations of participant behaviors, words, and drawings. Although actual data is not expressed as numbers, qualitative data can be counted. A researcher might count occurrences of a behavior, for example.

Quantitative data is numerical and is almost always analyzed by statistics – even if only by simple statistics, such as a mean or median. There are common ways that researchers collect and analyze the data. After the analysis, researchers can make inferences about usability issues regarding the product.

Collecting Qualitative Data

Usability researchers collect qualitative data to assess how well test participants can complete a task using the product and to observe where participants encounter problems. Qualitative data may also include participant perceptions of and opinions about the product. Collection methods provide data of various usefulness:

- Interviews provide flexible, in-depth data about participant attitudes and experiences with a product (Nielsen, 1993). Interviewers can structure questions to be open-ended to avoid binary answers. Binaries are direct opposites such as yes/no or easy/difficult, which may not fully utilize the participant's knowledge. An example of an open-ended question is: "Could you please guide me through the steps of adding page numbers to a document?" A less effective way to phrase the same question is, "Were you able to add page numbers in all specified places using the word processing program?" The answer to the first question provides more valuable data because participants are prompted to prove their knowledge, and researchers can obtain more complete information.
- Narrative descriptions, written or spoken responses to one or more questions, can be used to obtain information that researchers may not have anticipated about participants and products. Narrative descriptions are usually more in-depth than interview responses and are particularly effective because participants can speak more broadly about their experience with a product.

  For example, if a usability test is being conducted to evaluate the ease with which computer users can convert music files from .mp3 format to .wav, and then burn a compact disc, a narrative description may ask participants to describe that experience. The test monitor might ask participants to detail how they converted .mp3 music files from different directories of a computer into .wav files, and then created an audio CD. The narrative description may detail how accessible directories were, how the participant's lack of sleep the night before made it difficult to wait for .mp3 files to convert to .wav files, or any other feelings the participant may have had about the task. It should be noted here that interviews and questionnaires provide users' perceptions. These

perceptions should always be looked at in light of other information gathered from users.

- Surveys and questionnaires are more prescriptive methods of data collection because the structure of the questions limits participants' likelihood of providing unexpected or in-depth answers, says usability researcher D.T. Covey (2002). Surveys and questionnaires are best used for collecting large quantities of qualitative data on user preferences, attitudes, motivation, and satisfaction (Covey, 2002). The way questions are asked is important in considering what kind of participant data the survey or questionnaire can yield because the response choices offered strongly affect the types of responses participants can provide.
- Videotaping, as described in "Collecting Data From a Usability Test," is a data collection method that can provide valuable data in both qualitative and quantitative forms. Videotapes can provide qualitative data on participant behaviors. If researchers also use audio features, videotapes can record participant comments.

  One way videotapes can provide quantitative data is user behaviors are counted: Zimmerman and Muraski, technical communication specialists, say that researchers can use content analysis – counting words or other behaviors – to generate frequency data (1995) a type of quantitative data. For example, researchers could count the number of times participants scratch their heads or stop during tasks. Researchers could use these counts during analysis to look for trends.
- Thinking-aloud protocols, which are defined in "Collecting Data From a Usability Test," are an effective method for collecting qualitative data on certain portions of tasks that participants may forget or not remember correctly afterwards. It is highly useful because researchers can collect data about what participants do and what participants say they want to do to perform a task when they are actually doing it (Covey, 2002).

- Focus groups provide qualitative data about participant backgrounds as well as elaborate information about participant opinions, attitudes, and preferences on a product. Here is an example: If researchers want opinions on a software program that uses voice recognition to create a text document, they can lead a focus group in which they ask participants their views about such a software program. They can collect qualitative data from the focus group. The discussion might reveal that the program should be able to recognize more voices than it currently does or that it is too expensive relative to its capabilities. Of course, focus groups can lead to "group think," so data must be interpreted cautiously.

Collecting Quantitative Data

Quantitative collection methods are generally used to obtain performance measures. Researchers compare this data to pre-defined criteria for acceptable usability of a product.

Using Performance Measures

Performance measures can provide researchers with data in the form of time, success rates, and error rates of a participant's performance (Rubin, 1994). Researchers can use an existing performance objective or create their own (Rubin, 1994). Because researchers can create their own objectives to assess participant results, it is important for the measure to be highly structured to ensure that the performance data is valid.

An example of quantitative data researchers can obtain from performance measures are the numerical results of a task that asks participants to use a word processing program to add footnotes to three non-consecutive pages of an eight-page document. Performance objectives may be to complete the task in three minutes, to add footnotes on only the three specified pages, and to complete the task without making any errors. After researchers collect the participant data from the footnote creation task, they can record the results and find the average and median for all participants for later analysis.

The performance data researchers collect is important because it can point to the strengths and limitations of a product. Researchers can gain a better understanding of these strengths and deficiencies by looking at the tasks for which participants were most and least able to meet performance objectives.

Researchers can also generate statistics from performance data. For example, surveys can instruct participants to circle numbers indicating level of experience; the researcher can average the numbers and generate statistics. Likewise, thinking-aloud protocols could use time measures and error rates during the task to generate raw frequency counts.

Identifying Tasks That Did Not Meet Criteria

Before a usability test is administered, test developers define criteria that are considered acceptable for results. These predetermined criteria are then used to compare actual results to reveal problems with usability.

One measure Rubin (1994) suggests is the "70 percent success rate." If 70 percent of participants do not successfully complete a task, it can be labeled "difficult" or "problematic" (Rubin, 1994, p. 274). However, a researcher can use a different rate, depending on the usability issue.

Performance measures depend on the type of test. The measurement may be the percentage of correct results a participant must achieve to reflect a satisfactory knowledge of the product or a period of time in which a task should be accomplished. Regardless of the type of measurement, it must be defined before the test takes place. Then, in the analysis, researchers can focus on problematic tasks – those that did not meet the criteria – and base their recommendations on them. Tasks that do not meet criteria may represent limitations of the product and hence something with a high priority for being remedied (Rubin, 1994).

Identifying User Errors and Difficulties

After researchers collect results and determine which task(s) did or did not meet criteria, they can identify the errors that caused the unsatisfactory performance. Rubin says an error is "any divergence by a user from an expected behavior" (Rubin, 1994, p. 276). For example, if an e-mail program is designed to allow users to attach three files to an e-mail and a participant can only attach two files to the e-mail, then a researcher would identify that unexpected result as an error. Failing to meet the test objective (the user should be able to attach three documents to an e-mail) indicates that the user either committed an error, or that the e-mail program is difficult to use, or that there is an error in the documentation or program. Note that it is important to consider multiple interpretations when attempting to determine the source of a problem.

Analyzing Data

There are several valid methods for analyzing data. For quantitative data, researchers can conduct source-of-error analysis or generate inferential statistics. For qualitative data, researchers can categorize results to identify trends in data.

There are also methods of analyzing data where qualitative and quantitative data are used together to complement each other. Counting the frequency with which a behavior occurs and theme-based content analyses are two such methods that can be combined successfully to give a more robust interpretation of the data.

Researchers analyze qualitative data to separate participant responses into meaningful categories, thus, they usually code their data by creating words or phrases that are used as schemes to categorize responses. For example, if researchers were analyzing qualitative data on participants' reactions to cellular phones, they could create data codes such as comments about the reliability or ease-of-use of the phones.

One method researchers can use to analyze qualitative data is theme-based content analysis, a method by which researchers group categorized responses. As patterns in

the data emerge, researchers can classify results into meaningful groups of responses.

Conducting a Source of Error Analysis

Identifying the source of error "is a transition point from task orientation to product orientation" (Rubin, 1994, p. 276). When researchers conduct a source-of-error analysis, they make the distinction between the task they asked the participant to perform and the participant's actual performance. When analysis shifts to product orientation, researchers shift their focus to patterns they see in performance errors when using the product. A researcher must be clear about why user errors occurred, otherwise recommendations may not be accurate.

Using Inferential Statistics

If researchers use inferential statistics correctly, it means that they may be able to infer or apply the results of a usability study to a general user population. But in order to use inferential statistics effectively, researchers must generate statistically significant results. The results of a test are statistically significant if the researcher can assume that if the test were conducted again with different people of similar experience and background, the same results would occur.

- In usability testing, researchers usually do not use a large enough sample of participants to generalize and do not have the training in order to use and interpret statistics effectively. Rubin (1994)

Counting Frequency

Researchers can analyze both qualitative and quantitative data by performing simple frequency counts, such as the percentage of test participants who successfully perform a task and the percentage of participant errors.

Theme-Based Content Analysis

Theme-based content analysis (TBCA) helps researchers understand qualitative data by identifying the frequency of a theme in words, characters, written responses, behaviors and comments. In TBCA, researchers perform data collation on the data they have collected and then identify themes to classify the qualitative data.

Data Collation

According to usability researchers Neale and Nichols (2001), data collation involves several steps.

- Data is grouped according to the question or hypothesis it addresses.
- Data is recorded on a simple matrix, with rows to show raw data from individuals and columns for summarizing data.
- The matrix is retained throughout the iterative process of defining themes, allowing researchers to view and analyze the data individually and then collectively.
- Usability researchers read through the data a number of times until the responses of the participants become familiar (p. 172-173).

The following is an example of data collations: Researchers review focus group data and thinking-aloud protocol responses gathered during a test in which participants created several different types of graphs using a software program. Researchers categorize the results based on such distinctions as participant age, previous experience with similar software, and intuitive understanding of graphical representations of data.

Theme Definition and Classification

In theme definition and classification, Neale and Nichols (2001) state that each researcher reviews data independently and groups data according to common raw data themes that emerge from the participants' responses. For example, if the

results of surveys and interviews regarding a software program show that 23-year-old male and female participants express the highest level of confidence in the application of all participants, the researcher can start making groupings by age. Or, a researcher could make user confidence a category, with specific comments recorded in this category.

After individual researchers independently identify raw data themes, they meet to discuss interrater reliability – a measure that describes how consistent multiple reviewers and analyzers of data were with rating the data – making adjustments as necessary if discrepancies have occurred.

To determine interrater reliability, researchers define the data themes they found individually. If researchers individually identified many of the same themes, interrater reliability is high, and thus, the analysis is more reliable; if themes varied greatly among researchers, the interrater reliability is low. Researchers need to report the low rating because common themes are necessary for study results to be considered valid.

After the collective analysis and an agreed-upon group of raw data themes, researchers then record the number of responses for each theme on a classification matrix. There are two key purposes for creating a matrix: First, it ensures that data are easily accessible for the interpretation process. Second, the matrix allows all of the important information to be displayed when development decisions are being made (Neale & Nichols, 2001). Researchers present the raw data, raw data themes, and theme classifications in a standard classification table format as shown in the table that follows. Columns summarize data with increasingly specific codes and frequency counts to indicate the commonality and popularity of a theme.

The table that follows contains participant data on visual behavior noted during a usability test.

Table 5. Sample TBCA table of participant behaviors

Participant Behavior	Type of Behavior (Raw Data Theme)	Behavior group (Higher order theme)
P1 Looked at the clock P5 Looked at the ground and sighed P3 Looked at other participants	Looked away (3)	Change in visual behavior (7)
P4 Squinted eyes for more than five seconds P7 Squinted once or twice	Squinted (2)	
P2 Shifted focal point more than three times during a task P6 Shifted focal point while reading instructions	Shifted focal point on screen (2)	

In the sample TBCA table, participants are identified by number: P1, P2, and so on. The data noted for individual participants represents a behavior recorded on videotape (Participant Behavior column). The Type of Behavior (Raw Data Theme) column lists the frequency count for each type of behavior. The numbers in parentheses represents the total number of participants who had responses that fit into that theme.

The Higher Order Theme column is a summary of all like behavior. Raw data themes are summarized under one label and all responses are totaled so that more general themes can be assigned to the responses. A higher order theme, according to Neale and Nichols, requires a higher level of inference than a raw data theme. There may be several levels of higher order themes.

The use of frequency counts to analyze qualitative data, as described above, is one example of how qualitative and quantitative methods of collection and analysis can complement each other.

Interpreting Data

After researchers analyze data, their next task is to interpret the data. Researchers interpret data to understand what findings mean, what problems a product might have, and whether those problems can be attributed to the test, the participants, or both. Researchers can interpret data by simply reporting what they found in a study; or, they can interpret data by considering whether analyzed data supports a hypothesis or problem statement (Zimmerman & Muraski, 1995, p. 9). If researchers have enough relevant data about a product, they may draw a conclusion. Researchers can usually present key findings about a usability issue in one or two brief statements that are reflected in and supported by the data (Zimmerman & Muraski, 1995, p. 9).

As an example of data interpretation in usability, consider a usability study to determine the accessibility of audio features on DVD players in computers. The hypothesis is that only participants aged 18-24 with technology backgrounds will be able to successfully complete the task, using Rubin's "70 percent" criteria.

If researchers analyze raw data and find that more than 90 percent of participants in all age groups and experience levels, and 82 percent of participants aged 18-24 with technology backgrounds, cannot successfully choose the best sound configuration (i.e. stereo, surround sound), the results do not support the hypothesis. Researchers could use this information to determine whether they need to investigate the accessibility of audio features further. They might consider conducting tests in other areas related to the findings to determine how to address the problem. However, researchers would probably not draw immediate conclusions about the accessibility of audio features based on the findings of just one study.

Even when the results of a study are statistically significant enough to support a hypothesis, they still do not necessarily prove the hypothesis true (Muraski &

Zimmerman, 1995, p. 9). The results of more studies may give further support to the hypothesis and increase its value. When data do not support a hypothesis, this information helps researchers rethink a particular perspective, gain new insights into a topic, or identify alternative explanations for occurrences.

The Importance of Data Analysis and Interpretation

Performing a thorough analysis, interpreting results, and writing precise result statements are not just important for use in future studies. Synthesizing comprehensive data and its implied meanings into effective result statements is also important to the people responsible for making recommendations and presenting the results. They need to have a clear understanding of the results to present them to members of the organization who will decide whether to implement the changes.

Making Recommendations

Based upon analyses of data, usability researchers make recommendations for modifications to a product. Each recommendation should be directly related to a test objective. For example, a technical communicator might have this test objective for a user guide: "The user should be able to locate instructions for tasks in the manual by using the table of contents." If test data indicates that users were not successful in completing that task, a test recommendation might be to change the wording of table of contents entries to make them more task-oriented.

But before making recommendations, one usability expert suggests that those involved with the usability testing set the product aside for a few days. Stepping away from the testing and analysis can foster an attitude of more creative thinking (Rubin, 1994, p. 284). The real reason that users may not have been able to locate the instructions in the table of contents may be because the title of the task was not familiar to the user.

Creative thinking about cause/effect relationships works better in groups. In fact, making recommendations should be a group effort. Therefore, all who are involved with a product – project managers, product developers, and others on the product team – should be involved in this process. Their diverse perspectives contribute to making informed judgments, although the test monitor(s), who have witnessed the testing and are closest to the test results, drive the recommendation effort. Also, having the entire team involved in and committed to the recommendations makes it less likely that any one member or group of members will later challenge or refuse a recommendation in the report (Rubin, 1994, p. 285).

The product team starts by making a list of possible recommendations before they use a product, decide what their priorities should be in considering recommendations and then rank the recommendations. Ranking recommendations benefits all those who are involved with a product: it benefits the company who owns the product because it ensures that the most significant usability problems are addressed first and that the company will release a quality product, and it benefits

users because users experience less frustration if the most critical usability problems have been resolved before they work with a product. Before ranking the recommendations, those involved decide what their priorities should be in considering recommendations.

Categories for Ranking Recommendations

The following categories provide key ways for ranking usability recommendations:

- **Scope of the problem (global and local changes).** Global changes are those that affect many features of a product, have the most impact, and should be considered first. Local changes affect fewer features. Denoting whether a recommended change is global or local also determines, in part, the severity level of the problem and the need for recommendations (Dumas & Redish, 1993, pp. 333-334).

- **Severity level of a problem and the need for the recommended change**. Many usability researchers develop a scale for ranking recommendations. Sometimes the scale verbally denotes the criticality for the recommendation. Such a verbal list may consist of words like "most critical," "moderately critical," and "least critical." The criticality for the recommendation might also take the form of a scale of numbers, where 5 represents "most critical" and 1 represents "least critical."

  A numerical scale works well with large groups as a basis for negotiation. Group members can easily average numerically ranked recommendations and throw out any outliers. The numerical scale is advantageous because a number may encourage individuals to rank with precision.

- **Time needed to make the modification (short-term and long-term recommendations).** Short-term recommendations are those that a company can take care of right away. Long-term recommendations require a greater amount of time and may not be able to be addressed if the company has pressing deadlines for releasing a product (Rubin, 1994, p. 288). Long-term

recommendations are often incorporated in future versions of a product. Categorizing recommendations according to the time frame for the recommended modification allows designers and developers to coordinate modifications with product deadlines.

Possible Formats for Ranking Recommendations

The following tables are sample formats for ranking recommendations. These tables may be included in the explanation of findings and recommendations section of the final written report. The final written report is addressed in "Presenting the Results."

Table 6. Sample recommendation format. Explanations (rationale) follow each recommendation.

Company Web Page's Vertical Menu Bar
1. First recommendation: Increase the font size of the vertical menu bar from 8 pt. to 12 pt.
Explanation for first recommendation: Four test participants (80%) had to lean in and focus very carefully on the vertical menu bar to read its text. Two participants (40%) commented that they found the text too small to read on screen.
2. Second recommendation: Change the font of the vertical menu bar from a Script font to a more readable font, such as Arial or Helvetica.
Explanation for second recommendation: Four test participants (80%) had to trace with their fingers the shapes of the letters on the vertical menu bar in order to read the text. Four participants (80%) commented that they found the Script font looked like "little scribbles" on the screen and was particularly difficult to read

Source: Adapted from Rubin, 1994, p. 292.

Table 7. Sample recommendation ranking format. Table of findings, recommendations, and severity (indicated verbally).

Company Web Page's Vertical Menu Bar
1. Finding and Explanation: Test participants find it difficult to read the text on the vertical menu bar. Four participants (80%) had to lean in and focus very carefully on the vertical menu bar to read its text. Two of these participants (40%) commented that they found the text too small to read on screen.
Recommendation: Increase the font size of the vertical menu bar from 8 pt. to 12 pt.
2. Findings and Explanation: Test participants found it difficult to discern the letters of text in the vertical menu bar. Four test participants (80%) had to trace with their fingers the shapes of the letters on the vertical menu bar in order to read the text. The same four participants (80%) commented that they found the Script font, "little scribbles" on the screen, was particularly difficult to read.
Recommendation: Change the font of the vertical menu bar from Script font to a more readable font, such as Arial or Helvetica.

Source: Adapted from Rubin, 1994, p. 291.

Table 8. Sample recommendation ranking format. Table of findings, recommendations, and severity (indicated verbally).

Finding	Recommendation	Severity
Test participants are unable to read text on vertical menu bar without leaning into the screen.	Increase the font size of the menu bar's text from 8 pt. to 12 pt.	Most Critical

Source: Adapted from Information & Design (Information & Design, 1998, http://www.infodesign.com.au/usability/samplewebsiteevaluation.pdf).

Table 9. Sample recommendation ranking format. Table lists problem number (for easy reference), scope, severity level, and frequency.

Problem #18	Text on Vertical Menu Bar is Difficult to Read
Scope (global or local)	Local
Severity Level (the severity # and what it means)	5- Most Critical
Frequency	80% of test participants had difficulty reading the text on the vertical menu bar because of small font size and hard-to-read font.
Explanation	Test participants who had difficulty reading the vertical menu bar text leaned closer to the monitor, squinted to read the text, and tried to trace letters of text with their fingers. Two participants asked for assistance reading the text because they said the text "looks like little scribbles" on the screen.
Recommendations (a bulleted or numbered list)	Enlarge the text from 8 pt. to 12 pt. Change the font from Script to a more readable font, such as Arial or Helvetica.

Source: Adapted from Dumas & Redish, 1993, p. 350.

Table 10. Sample recommendation ranking format. Table of tasks participants were asked to perform, the source of participant error, severity of problem, recommendations, and business rationale.

Participant Task	Source of Error	Severity Ranking	Recommendation	Business Rationale
Find the shipping department's website on the company home page.	Link to shipping department's website is in the hard-to-read vertical menu bar on the home page.	5	Enlarge the font of vertical menu bar text from 8 pt. to 12 pt. Change the font from Script to a more readable font, such as Arial or Helvetica.	If we want to appeal to potential customers, we need a home page that looks professional and is easy to navigate. In order for current customers and employees to also be able to connect to our organization's various departments, they must be able to locate links to each department.

Source: Adapted from Rubin, 1994, p. 292.

Any of these formats for ranking recommendations may be used or adapted. The choice for how to display the findings and recommendations depends heavily on what purposes the information will serve for its audiences. Therefore, those who rank the recommendations should take into consideration how results will be presented.

Presenting the Results

The results of usability testing may be presented in several forms: a written report, an oral presentation, or a videotape. A written report is required, but selecting other formats for presenting the results will depend on the audience's needs and uses for the data. To choose the appropriate presentation format(s), the presenter should answer the following questions:

- Who is the audience?
- What is the audience's expertise? What does the audience already know?
- What does the audience want or need to know?
- How will the audience respond to the presentation?
- What will the audience do with the data?

The audience for the report usually includes project managers, product designers, and product developers. The audience may also include upper-level managers and external clients. Because of the audience's diversity of expertise and differences in the way they use or view the product, presenters sometimes use two or three different presentation formats, tailoring each to its appropriate audience. For example, busy individuals may prefer a quick synopsis of conclusions in an oral presentation in addition to a copy of the written report. Individuals who require the finer details need to view the entire contents of the detailed written report but may or may not prefer to hear an oral presentation.

Certainly, though, presenting in writing, presenting orally, or presenting through videotape can all be effective ways to communicate the results of usability testing to colleagues.

Presenting in Writing

Regardless of any other presentation form used, a written report is always necessary for capturing the results of usability tests (Rubin, 1994, p. 288). The written report

is completed after each usability test, which means that iterative testing requires multiple reports.

The purpose of the report is two-fold: to communicate the results of testing to project managers and product developers, and to serve as an historical record for the product. An historical record may alert future developers to problem areas as they develop subsequent versions of the product or entirely new products. The written report is most useful for product designers and developers, although all individuals who are involved in evaluating and making alterations to a product (e.g., product managers, external clients, etc.) should receive a copy as well.

Steps for Generating the Written Presentation

There are four steps to writing a usability report: planning, organizing, drafting, and designing.

- **Step One: Planning.** The first task in writing the report is to compile the results of usability analyses, which means that data analyses must be completed by the time writers begin to plan the report. After results are compiled, writers need to consider how they can make the findings accessible for the audience. A good way to begin to plan the report is to create a short, simple outline of the lists or tables of results that will be included in the report and fill in around the outline with text.

- **Step Two: Organizing.** Writers organize the usability report according to the ranked findings and recommendations. The report might put the most critical issues first, or might start with product fixes that are easier to accomplish before the product's release date. Another way to organize the report is by topic and group findings for designers and developers who will make the modifications to a product (Dumas & Redish, 1993, pp. 348-349). Grouping findings by topic helps designers and developers easily locate the modifications they need to make.

- **Steps Three and Four: Drafting and Designing.** The decision to draft or design first is based on the writer's preference. Writers also include visual aids, such as lists and tables, to draw readers' attention to important findings. Writers may choose to include participant comments to increase the impact of the report on the audience.

  Generally, presenters should apply the same visual design principles to the usability report that they apply to other technical documents to make information easily accessible.

General Format for the Written Presentation

Usability is a specific kind of research, so the general format for the written presentation looks much like that of a research report. This research format allows others to determine the validity and reliability of the results. The general format for a usability report usually includes five major sections.

Section One: Executive Summary

Dumas and Redish suggest that an executive summary include the following background information of the usability test (1993, p. 349).

- Product tested.
- Names of test monitor(s).
- Dates of the test.
- Objectives of the test.
- Participants (how many, characteristics).
- Major findings.
- Recommendations.
- The business rationale for fixing the problems.

Section Two: Objectives, Methods, and Procedures

The major section of all usability reports should always begin with the objectives that drove the testing. Each objective – performance, attitudinal, and document – should be tied to a method. Each objective should also be reported on in the results section.

Dumas and Redish advise writers to include only the special points of the test methods and procedures because a usability report is "not an academic paper" (1993, p. 349). Writers should include the following information in the methods section.

- Objectives.
- Data collection method(s).
- Tasks the participant was asked to perform (preferably a list).
- Participant demographics (a table is an effective format for presenting demographics).
- Plan for analyzing the data.

Section Three: Results

According to Dumas and Redish, the results section may contain a brief introduction followed by tables that summarize the data and results (1993, p. 349).

Section Four: Explanation of Findings and Recommendations

The explanation of findings and recommendations may take several different formats, but are generally presented in tables like those shown in "Making Recommendations."

Section Five: Appendices

The appendices include the test instruments, such as a sample consent form, scenarios, questionnaires, and data collected from the participants.

Common Industry Format from a Written Report

For a summative usability evaluation – "one that produces usability metrics that describe how usable a product is when used in a particular context of use" (Serco Ltd., 2002, http://www.usability.serco.com/prue/cif.htm) – a special Common Industry Format (CIF) – a report format developed and used by many major companies, such as Hewlett-Packard, IBM, Microsoft, Boeing, and Kodak, in collaboration with the National Institute of Standards in Technology (NIST), may be used. The purpose of the format is to ensure that the following aspects of testing have been met.

- Usability researchers adhered to good usability practice.
- Usability researchers have sufficient data to judge the validity of the results.
- The test would produce essentially the same results if the test were replicated on the basis of the information given in the CIF.
- Usability researchers use specific effectiveness and efficiency measures in their objectives.
- Usability researchers measure participants' satisfaction with the product.

By ensuring the inclusion of the items shown above, the CIF serves to reduce the confusion many companies have experienced in the past when evaluating differently formatted usability reports. (See "Appendix A: Outline of the Common Industry Format (CIF) for Usability Reports" for a detailed outline of the Common Industry Format).

Feedback from a Written Report

Many writers find it personally advantageous to solicit feedback after distributing the written report. Feedback from the report's audiences provides ideas to incorporate in future reports, such as how to make information more accessible to a particular audience. Feedback also informs writers how well they have analyzed their audiences and indicates the success of the report.

Soliciting feedback from the audience is essentially performing a usability test on the written report. The following are some specific questions to ask the audience that will provide insightful feedback.

- Were you able to get the information you needed?
- Was the type of information you needed all there?
- Was the format easy to read?

(Rubin, 1994, p. 293)

Presenting an Oral Report

An oral presentation of results can complement the written report but should never be the only medium used to communicate usability test results (Dumas & Redish, 1993, p. 341). Oral presentations are especially helpful for those who may not need or want to review the finer technical details of usability testing. It is helpful to schedule an oral presentation of the results before anyone sees the written report (even if some individuals already know the test findings). Presenting the oral presentation first can pave the way for acceptance of the details in a written report. Also, the presenter should record what is discussed and agreed upon at that meeting, so that the issues discussed there can be incorporated into the final written report.

The oral report will generally have the same elements as the written report, discussed in "Presenting in Writing." There are, however, a few format, style, and preparation features of an oral presentation that differ from those of a written report. For example, visual aids are a necessary feature for oral presentations. A question-and-answer session is another feature of oral presentations that allows presenters to solicit feedback from the audience.

Format

When creating an oral presentation of usability results, presenters tailor format features specifically to the oral presentation. The format for an oral presentation can

generally be broken out into four major sections: title, introduction, main body, and conclusions and recommendations.

- Titles for oral presentations are often shorter and more concise than they are for written reports. Attendees usually read the title of the presentation in advance and may base their decision to attend or not attend the meeting on the title, so the title should inform the potential attendees whether or not their attendance is necessary.
- In the introduction to the presentation, presenters may address the following items in this order: outline of the presentation, purpose of the test, background information, and references.
- The main body of the oral presentation describes the usability tests to the audience. The description of testing will include participant demographics and a brief overview of the methodology. Presenters discuss only the top three to five findings of the usability testing. An oral presentation is generally considered an inappropriate venue for presenting all of the details included in a written report.
- Conclusions and recommendations in the oral presentation summarize only the top three to five findings presented in the main body and offer recommendations pertinent to those findings. Presenters may point out the business rationale for specific recommendations or for further testing and analyses. Presenters may also briefly mention a few other significant findings and recommendations that the audience may want to investigate further.

Effective Visual Aids

Visual aids are absolutely necessary in an oral presentation. Using effective visual aid(s) gives presenters the power to make a substantial impact on an audience and can build presenters' credibility with colleagues.

Selecting the appropriate visual aid(s) requires that factors like audience size, frequency of updating the presentation, and audience's use(s) for the results presented are considered. The following visual aids can be used effectively in different group settings.

- Flip charts and posters are good visual aids for interaction with a small audience. If the audience will make recommendations during a meeting, flip charts are an easy and inexpensive way of recording audience comments. Flip charts are also easy to update as the meeting progresses. The flip chart or poster may serve as a display during the entire meeting.
- Handouts are a good medium for disseminating information so that the audience members can take the information with them when they leave. But handouts can also draw audience members' attention to the paper and away from the presenter. Presenters should weigh the potential loss of attention against the benefits of the handout to determine if a handout is an appropriate visual aid. If they choose to use handouts, effective presenters make sure that handouts have been given to the audience before the meeting so that listeners can follow the handout's information as it is discussed.
- For small audiences, a model of the product can communicate findings more effectively than an oral description. Models appeal to audience members' tactile and/or auditory senses as well their visual senses and may provide a more authentic look at usability test results (Honolulu Community College, n.d.).
- Slides developed using presentation software work well for larger audiences. Slides are a professional way of displaying usability test data and results in the form of tables, graphs, and figures, which can be used to compare data, such as frequency of test participant errors. Another advantage of using slides as visual aids is versatility. Presenters can easily animate slides to create a presentation with visual impact.
- Transparencies can be effective with large or small groups. Transparencies are easy to create, transport, and update and are often used in technical presentations for these reasons.
- If video was used to monitor usability testing, video clips can be used during an oral presentation. Preparing video clips sometimes requires extensive searching of the videotapes in order to select just the right comments. However, presenting participants' humorous and insightful comments usually makes a more powerful

impact on the audience than presenter's descriptions of usability issues (Dumas & Redish, 1993, p. 360).

Throughout the presentation, presenters point to and explain visual aids. Effective presenters are careful not to address a particular visual aid until they are ready to explain its significance.

Feedback from an Oral Presentation

Presenters find it very advantageous to solicit feedback from the audience after an oral presentation. An oral presentation has the advantage of the questions-and-answers session following the presentation, which allows the audience to resolve points of confusion and usability issues of immediate concern. This discussion gives the presenter the opportunity to ensure that the audience understands the test results. The question-and-answer session also gives the presenter the opportunity to gauge how topics might be effectively presented in the written report.

Presenting a Videotape

Videotapes are a useful form of presentation for companies who are unable to conference with off-site audiences. Dumas and Redish suggest starting the videotape by accentuating the design strengths followed by three to five usability problems and recommendations (1993, p. 357). Although it requires more time and expertise to prepare, an effective videotape allows presenters to make a powerful impact relative to the short amount of time in a video.

Responsibilities and Rewards of Presenting Results

Presenting the results of usability testing is a significant responsibility. Effective presenters know each of their audience's needs and expertise and adapt the form of presentation to suit them individually.

Presenting the results is also rewarding. Those who present the usability results to other professionals have the opportunity to play a powerful role in recommending modifications to products and, in the long run, to enhance customer satisfaction.

Conclusion

Usability testing can be a valuable resource for those who endorse it. It can be a major factor in determining the effectiveness, efficiency, and usefulness of many different types of products, including communication products. Technical communicators and other professionals who advocate usability testing value the results obtained because they almost always form the basis for recommendations for improvements.

Like anything worthwhile, usability testing requires time, resources, and money. Professionals cannot ignore the probability of product delays and the possibility of inconclusive test results, but they must consider the long-term value of not just more highly usable products, but the value of the brand loyalty and repeat business that is directly related to high usability.

When preparing for usability testing, product developers must set objectives for the test, and should select the appropriate type of test, considering the time and resources available. Test developers can select quantitative or qualitative data collection methods, including surveys and questionnaires, narratives, participant observation and others. Test types and data collection methods are important to help prepare for testing, to successfully administer usability testing, and for evaluating results.

The selection of test participants is a critical element of usability testing. Care must be taken to ensure that participants reflect the product's target audience and that they represent a broad range of experience. Organizations must be cognizant of participants' rights and the principles and regulations that protect them, and must conduct testing in an ethical manner.

Selecting appropriate test monitors can have serious consequences that may affect the validity of the test results. Properly monitored usability tests increase the potential for conclusive and valuable data. The test environment is equally

important. Selecting a neutral test environment that is free of distractions is most conducive to proficient task performance.

After the test data are collected and compiled, they must be analyzed and interpreted. Usability researchers then rank recommendations and present them to product developers and others involved in the product development process in the form of a recommendation report.

Cost and time requirements, potential delays in product releases, and the possibility of inconclusive data may prevent some organizations from adopting usability testing. But to many professionals and organizations, iterative usability testing throughout the product development cycle is an investment in product usability. *Testing early and testing often* means saving time and money in the long term.

A Message From the Authors

As graduate students in technical communication we do not claim to be experts on usability testing. We are, however, professionals committed to product users and user satisfaction. We believe that part of our role as technical communicators is to work to influence product developers about the benefits of usability testing. We can do that, in part, by conducting usability testing to improve our own products, whether they are policies, procedures, manuals, online help systems, computer-based training modules or any other type of communication we are asked to produce.

We hope this information on usability and usability testing is useful to you. We invite you to contact us at lutzja@.muohio.edu with your questions and comments.

Glossary

Assessment test – a usability test in which data is collected for the purpose of establishing the lower-level design details.

Automated data loggers – computer programs that record the actions the test participants perform on the computer.

Cameraless videotaping – the recording of the onscreen events on a test participant's computer by connecting the computer's graphics card to a video recorder.

Code – a word or phrases that are used as schemes for categorizing responses in data analysis.

Comparative test – a test in which two different designs are compared for their ease-of-use.

Computer-human interface design – see *User-centered design.*

Confidentiality – the principle that all access to data collected from participants must be restricted.

Constructive interaction – a variation of the thinking-aloud method where two test participants interact and talk with each other while performing an assigned task during the usability test.

Data collation – a procedure in theme-based content analysis in which researchers group data according to the question or hypothesis it addresses.

Debriefing – a period at the end of testing in which test participants are asked to review their difficulties using a product with the test monitor.

Efficiency – a measure of how productive experienced users can be if they use the product.

Empirical data – data that results from an experiment.

Ergonomics – see *User-centered design.*

Error tolerability – a product characteristic that describes how easy the product is for users to recover from errors they make in using the product.

Exploratory test – a usability test used early in the product development process to establish the validity of high-level design concepts of a product.

Focus groups – a group of at least six users formed for the purpose of discussing a product and its usability issues.

Front-end Analyses – the early, and often most important, stages of usability testing, consisting of task analysis and participatory design.

Greeked text – dummy text used as a placeholder for text content in a website.

High-level design – a preliminary product design early in the product development life cycle that shows only the product's most salient features.

Human factors design – see *User-centered design.*

Informed consent – the principle that agreement to participate in testing may occur only after test participants have received information about the test's purpose, procedures, their right of free will, and possible risks.

Interrater reliability – a measure that describes how consistent multiple reviewers and analyzers of data were with rating the data.

Iterative testing – the use of many small usability tests throughout the product development life cycle.

Learnability – a product characteristic that describes how easy the product is for users to learn.

Likeability – a product characteristic that describes the amount of customer satisfaction with the product.

Location task – a task in the usability test that measures how quickly and effectively the information in a document, computer program, online help, or website can be found by test participants.

Mean – the average.

Median – the midpoint in a range of numbers.

Memorability – a product characteristic that describes how easy the product is for users to recall after long periods of absence from the product.

Minimal risk – the principle that test participants should not be exposed to any harm greater than they would experience in their daily lives.

Non-disclosure – a confidentiality agreement; usually a form test participants are required to sign before they begin participating in testing.

On-site observation – a task analysis technique in which developers go to users' homes and/or places of work to observe the user using a product in its typical environment.

Orientation script – the planned dialogue of test monitors as they introduce the usability test and participants' rights to the test participants.

Participatory design – the inclusion of users in the discussion of the product's design and development.

Performance task – a task that test participants are asked to perform to measure how easy it is for participants to use the product to perform the assigned task.

Pilot test – a small-scale test performed prior to administering the usability test to a large number of participants.

Post-task questionnaires – questionnaires given out during a usability test to obtain immediate feedback from test participants as they perform their tasks.

Post-test questionnaires – questionnaires given out after the usability test to obtain information from test participants about their experiences while using the product.

Pre-test questionnaires – questionnaires administered before the usability test that collects the participants' background information and initial impressions about the product.

Product development life cycle – the process by which a product is developed starting with user needs analysis and concluding with the product's release.

Protocol analysis – observing users solely to determine the tasks they perform with the product and the expectations they have about the product.

Prototype – a mockup or model of the original product.

Qualitative data – any type of data that yields results that are not numerical, but may be able to be counted.

Quantitative data – any type of data that is numerical and can be counted.

Questionnaires – a data collection tool that poses questions about users' tasks and objectives for a product.

Raw data themes – the first level of categories researchers who analyze data created based on responses that emerge as common themes from the participants' responses.

Reliability – a characteristic of a usability test that describes how likely it is that researchers performing the same research under the same conditions will attain the same results as each other.

Retrospective testing – a variation of the thinking-aloud method where participants are asked to review a videotape of themselves during a usability test and comment on their actions.

Scenario building – an early usability testing technique that asks test participants to envision possible uses for a product that has yet to be fully developed.

Scenarios – the presentation of a usability task within a short story that relates the typical context of product use.

Scope – a characteristic of usability findings that describes how widespread the findings are in the product.

Self-reporting – a data collection method in which users are asked to answer questionnaires about the task they perform in the usability test.

Severity – a characteristic of usability findings that describes how critical a usability problem is to correct.

Source-of-error analysis – a type of data analysis in which researchers look at task-error correlations to determine what features of a product were responsible for errors participants made during testing.

Task analysis – a process developers use to determine the objectives users have about their product and the tasks users expect to perform.

Technical communicators – professionals who specialize in communicating technical information to diverse audiences.

Test monitor – the individual who observes and records data from the usability test.

Test plan – a formal document containing the purposes, objectives, methodology, materials, and final report format of the usability test.

Testing objectives – clear statements of what is to be tested. For example, if a user should be able to complete a task in a specified period of time, the objective states "the user can accomplish (name of the task) in (number of) minutes."

Theme-based Content Analysis (TBCA) – a method of data analysis in which researchers compile qualitative and quantitative data by identifying the frequency of

words, characters, or themes in written responses and comments of test participants.

Thinking-aloud protocol – a data collection method in which test participants talk about what they are doing and thinking as they perform assigned tasks.

Understandability task – a task given to test participants in the usability test to determine how well a written document is understood by the participants.

Usability – the characteristic of being easy to use.

Usability engineering – see *User-centered design.*

Usability testing – the use of experiments to determine the ease of use (usability) of a product. Experimenters observe users from the intended target audience while users perform representative tasks using a product or prototype.

User-centered design – an approach in which a product's goals, features, and the tasks it supports are customized according to the needs and desires of its users.

User-centered product – a product that has been customized for the user.

Utility – a product's ability to carry out its intended function.

Validation test – a test used late in the usability process to confirm a product's usability.

Validity – a characteristic of a usability test that describes how well a test's protocol measures the outcomes the protocol was designed to measure.

Verification test – see *Validation test.*

Workaround – the use of an alternative process to a desired end because the normal or accepted route is difficult to use.

Appendix A: Outline of the Common Industry Format (CIF) for Usability Reports

Title Page

- Identify the report as a Common Industry Format (CIF)
- Name the product and the version that was tested
- State when the test was conducted
- Provide contact information for questions about the test

Executive Summary

- Identify and describe the product
- State the reason for and nature of the test
- Summarize the method of the test, including the number of and type of participants and their tasks
- Show results expressed as mean scores
- Tell the probability that the difference did not occur by chance if differences between values or products are claimed

Product Description

- State the formal product name and release or version
- Describe briefly the environment in which the product should be used
- Tell what parts of the product were evaluated
- Describe the user population for which the product is intended
- Tell the type of user work that is supported by the product

Test Objectives

- State the objectives for the test
- Describe any areas of specific interest

Participants

- State the total number of participants tested
- Describe the segmentation of user groups tested (if more than one user group was tested)
- Describe the key characteristics and capabilities expected of the user groups being evaluated
- Tell how participants were selected and whether they had the essential characteristics and capabilities
- Describe differences between the participant sample and the user population
- Describe groups with special needs

Context of Product Use in the Test

- Describe the task scenarios for testing
- Tell why the tasks were selected
- Explain any task data given to the participants
- Describe any completion or performance criteria established for each task
- Describe the type of space in which the evaluation was conducted
- Tell what computer hardware model, operating system versions, etc. were used
- Tell if the product uses a web browser, and if so, the browser type and version
- Tell if the product has a screen-based visual interface and state the screen size and monitor resolution
- Tell if the product has an audio interface, relevant settings or values for the audio bits, volume, etc.

- Describe test administrator tools, such as questionnaires or data logging systems (a copy of tools, such as a questionnaire may be included in the appendix)

Design of the Test

- Explain the logical design of the test
- Define briefly any independent variables and control variables
- Describe briefly the measures for which data were recorded for each set of conditions
- Explain any policies and procedures for training, coaching, assistance, interventions or responding to questions
- Summarize briefly the task instructions (put the exact task instructions in an appendix)

Usability Metrics

- State the percentage of participants who completely and correctly achieved each task goal
- Tell if it is necessary to provide participants with assists (efficiency and effectiveness metrics must be determined for both unassisted and assisted conditions)
- State the mean time taken to complete each task, together with the range and standard deviation of times across participants
- Summarize other usability results

Data Analysis

- Show data scoring
- Show data reduction
- Provide the statistical analyses

Presentation of the Results

- Show the performance results per task or group in tables
- Include summary table(s) of all tasks across user groups
- Present performance results in a graph

Satisfaction Results

- Show satisfaction results in a table
- Include summary table(s) of satisfaction results
- Present satisfaction results in a graph

Appendices

- Include any custom questionnaires
- Provide participant general instructions
- Provide participant task instructions. (The National Institute for Standards and Technology, 1999).

Sources Consulted

Anderson, P.V. (1998). *Technical communication: A reader-centered approach* (4th ed.). Fort Worth, TX: Harcourt Brace College Publishers.

Andre, T.S., Belz, S.M., Hartson, H.R. & McCreary, F.A. (2001). The user action framework: A reliable foundation for usability engineering tools. *International Journal of Human-Computer Studies, 54,* 107-137.

Anschuetz, Lori, Hinderer, Deborah and Rohn, Janice Ann. (1998.) *When the Field is Far Afield: Multiple-Country Observations of Complex System Use.* Reprint of paper from UPA 98 Proceedings (Washington, DC), published by Usability Professionals Association Inc. Retrieved October 6, 2003 from http://www.teced.com/c_and_p.html.

Barker, T. (1998). *Writing software documentation.* Boston: Allyn & Bacon.

Barker, T. (2003). *Writing software documentation.* Boston: Allyn & Bacon.

Diamond Bullet Design, Inc. (2001). *Usability first glossary*. Retrieved September 22, 2002, from http://www.usabilityfirst.com/glossary.

Benefits of Usability – Cost justifying usability. Usability by Design. Retrieved September 8, 2003 from http://www.usability.uk.com/benefits_costjust.htm.

Berry, D. C. & Broadbent, D. E. (1990). The role of instruction and verbalization in improving performance on complex search tasks. *Behavior and Information Technology, 9,* 175-190.

Bevan, N. (1999) C*ommon industry format usability tests*. Retrieved October 4, 2002, from http://www.usability.serco.com/papers/cifus.pdf.

Burmeister, Oliver K. (2001) Usability testing: Revisiting informed consent procedures for testing Internet sites [Electronic version]. *Conferences in Research and Practice in Information Technology, 1*, 3-10. Retrieved September 14, 2002, from http://www.jmir.org/jmir-pdf/burmeister.pdf.

Callahan, M.R. (Ed.). (1985). Be a better task analyst. *INFO-LINE: Practical Guidelines for Training and Development Professionals*, 8503.

Common visual aids. (n.d.). Retrieved October 4, 2002, from the Honolulu Community College Faculty Development website: http://www.hcc.hawaii.edu/intranet/committees/FacDevCom/guidebk/teachtip/visuals.htm.

Communication Studies Department, University of Kansas. (n.d.) *Virtual presentation assistant: Visual aids*. Retrieved October 4, 2002, from http://www.ku.edu/cwis/ units/coms2/vpa/vpa7.htm.

Communication Technologies Branch, National Cancer Institute's (NCI's) Office of Communication. (n.d.). *Usability basics*. Retrieved September 16, 2002, from http://www.usability.gov/basics/index.html.

Covey, D. T., (2002). *Usage and usability assessment: Library practices and concerns*. Washington, D.C.: Digital Library Federation.

Dieli, M. (1989). The usability process: Working with iterative design principles. *IEEE Transactions on Professional Communication*, *32*.

Dumas, J.S., & Redish, J.C. (1993). *A practical guide to usability testing*. Norwood, NJ: Ablex Publishing Company.

Farson, R. (1997). *Management of the absurd*. New York: Simon & Schuster.

Fuccella, J. & Pizzolato, J. (1999, June). *Separating content and visuals for rapid results*. Retrieved September 25, 2002, from http://www-106.ibm.com/developerworks/ library/wireframe/wireframe.html.

Gaffney, G. (2002). *Writing usability reports*. Retrieved October 4, 2002, from http://www.infodesign.com.au/usability/writingusabilityreports.html.

Guth, Robert A. '*Worm' woes could force software-writing changes*. August 29, 2003. The Cincinnati Enquirer.

Information & Design, Pty Ltd. (1998, October). *website evaluation conducted for (the client) by Information & Design*. Retrieved October 4, 2002, from http://www.infodesign.com.au/usability/samplewebsiteevaluation.pdf.

Information and Design. (2002, May). *What is a participatory design workshop?* Retrieved October 8, 2002, from http://www.infodesign.com.au/usability/participatorydesign.html.

Johnson, R. (1998). *User-centered technology: A rhetorical theory of computers and other mundane artifact*. Albany, New York: State University of New York Press.

King, M. et al. (1995, September). *Evaluation of natural language processing systems: Final report*. (Expert Advisory Group on Language Engineering Standards [EAGLES] document EAG-EWG-PR.2). Retrieved October 5, 2002, from http://www.issco.unige.ch/ewg95/.

Law, C., Barnicle. K., & Henry. S. (2000, October). *Usability screening techniques: Evaluating for a wider range of environments, circumstances and abilities*. Retrieved October 5, 2002, from http://trace.wisc.edu/docs/usability_ screen_ techs_UPA2000/usability_screening_techs.htm.

Microsoft Corporation. (2000, September). Retrieved September 9, 2002, from the Microsoft Usability Research website: http://www.microsoft.com/usability/gratuity.htm.

Morgan, G. (1997). *Images of organization* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Muraski, M.L. & Zimmerman, D.E. (1995). *The elements of information gathering*. Phoenix, AZ: The Oryx Press.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979, April). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. Retrieved October 5, 2002, from http://ohsr.od.nih.gov/mpa/belmont.php3.

National Institute of Standards and Technology. (1999, October). *Common industry format for usability test reports version 1.1*. Retrieved September 27, 2002, from http://zing.ncsl.nist.gov/iusr/documents/cifv1.1b.htm.

National Institute of Standards and Technology. (1999, August). *The IUSR project: Industry usability report*. Retrieved October 4, 2002, from http://zing.ncsl .nist.gov/iusr/documents/WhitePaper.html.

Neale, H., & Nichols, S. (2001). Theme-based content analysis: a flexible method for virtual environment evaluation. *International Journal of Human-Computer Studies, 55,* 167-189.

Nielsen, J. (1993). *Usability engineering*. New York: Academic Press, Inc.

Nielsen, J. (2000, March 19). Why you only need to test with 5 users. *Jakob Nielsen's Alertbox: Current Issues in Web Usability*. Retrieved September 22, 2002, from http://www.useit.com/alertbox/20000319.html.

Nielsen. J. (2002, July 22). Becoming a Usability Professional. *Jakob Nielsen's Alertbox: Current Issues in Web Usability*. Retrieved October 12, 2002 from http://www.useit.com/alertbox/20020722.html.

Nielsen. J. (2003, June 2). Usability for $200. *Jakob Nielsen's Alertbox: Current Issues in Web Usability*. Retrieved September 17, 2003 from http://www.useit.com/alertbox/20030602.html.

Oreskovic, A. (2001, March 5). Testing 123. *The Industry Standard*. Retrieved October 10, 2002, from http://www.thestandard.com/article/0,1902,22338,00 .html?body_page=2.

Participatory Design History. (1998, June). Retrieved September 9, 2002, from http://www.cpsr.org/conferences/pdc98/history.html.

Rosenbaum, Stephanie (October 13, 1999). *Iterative Usability Methods: Why Testing Isn't Enough*. Retrieved September 17, 2003 from http:/teced.com/pdfs/mochi99.pdf.

Rubin, J. (1994). *Handbook of usability testing: How to plan, design, and conduct effective tests*. New York: John Wiley and Sons, Inc.

Schriver, K. A. (1997). *Dynamics in document design*. New York: John Wiley & Sons, Inc.

Serco Ltd. (2002). *What is the common industry format?* Retrieved October 4, 2002, from http://www.usability.serco.com/prue/cif.htm.

Silva, M., & Breuleux, A. (1994). The use of participatory design in the implementation of internet-based collaborative learning activities in K-12 classrooms. *Interpersonal Computing and Technology, 2*, 99-128. Retrieved September 15, 2002, from http://www.helsinki.fi/science/optek/1994/n3/silva.txt.

Smart, Karl L. (April, 2002). Quality and the Customer Experience: Methods of Collecting Data. *Intercom*, Volume 49, Issue 4, 36-37. [Electronic version]

Sullivan, Kent (1996). The Windows® 95 User Interface: A Case Study in Usability Engineering. Association for Computing Machinery, Inc. (ACM), CHI 96 Electronic Proceedings. Retrieved September 17, 2003 from http://turing.acm.org/sigs/sigchi/chi96/proceedings/desbrief/Sullivan/kds_txt.htm.

Sullivan, P. (1989). Beyond a narrow conception of usability testing. *IEEE Transactions on Professional Communication*, *32*, 260.

The CoHousing Company. (n.d.). *The CoHousing Company*. Retrieved September 15, 2002, from http://www.cohousingco.com/index.htm.