Usability Testing: Recipe for Success

By Deborah Abratt, Wayne Mallinson and Antonet Bekker

Prepared for StarWest (2003), San José, California

The awareness of the benefits of non-functional testing types is steadily growing. Software usability testing is an example of such a discipline. Usability testing evaluates how easy a system is to learn and use and it offers many benefits to both users and development companies. Benefits to users include increased pace of learning, decreased user task time and errors and increased job satisfaction. Development companies benefit through providing less technical support and training to users, improved management of risks associated with usability rework as well as improved user acceptance. Companies also develop software faster and more cheaply since only relevant functionality is produced and they can identify which designs are effective and which to avoid. Benefits of web usability include increased sales, increased traffic and improved user productivity and features usage. Evidence exists for each of the above benefits in case studies representing how companies have harnessed the power of usability testing to add quality to products and processes. One such success story comes from a usability test the authors' company recently performed in a major South African retailing corporation, which will potentially cause drastic improvements in efficiency and cut working hours almost in half for certain tasks.

1. Non-Functional Testing

When striving for software quality, it is important to take a holistic approach. Although non-functional system testing is as important as functional system testing, it is often overlooked and not well specified. The BS 7925-1 (1998) standard defines non-functional system testing as "testing of those requirements that do not relate to functionality." In other words, non-functional system testing analyses those aspects of a system that are measured on a scale and cannot be verified as valid or invalid. These "subjective aspects" of software quality (Roth, 1999) are more difficult, although not impossible, to define and measure and require specialist skills to test.

The following are examples of non-functional testing types (BS 7925-1, 1998):

- **Performance Testing** – testing conducted to evaluate the compliance of a system or component with specified performance requirements
- **Stress Testing**: – testing conducted to evaluate a system or component at or beyond the limits of its specified requirements
- **Security Testing** – testing whether the system meets its specified security objectives
- **Storage Testing** – testing whether the system meets its specified storage objectives
- **Usability Testing** – testing the ease with which users can learn and use a product

Testers should be aware of non-functional system attributes so that they can provide worthwhile feedback about them to the rest of the development team.

2. What is Software Usability?

According to ISO 9241-11 (1998), usability is the "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use." Important aspects of this definition include:

- **"specified users"**: It is important to note that when systems are being designed for usability, the first step should be identifying target user population. Usability is not an absolute term but, rather, a relative one. A system can only be usable relative to the user population it serves.
- **"specified goals"**: The functionality designed into a usable system will be relevant to its users. It is often the case that unnecessary or inappropriate functionality is incorporated into a system. This functionality can "clutter" the interface and make relevant functionality more difficult to access. On the other hand, a usable system presents its users with routes to achieve their goals in a clear fashion.
- **"effectiveness"**: Effectiveness is the accuracy and completeness with which users achieve specified goals. For example, a user is effective if they can complete tasks making a minimal amount of errors.

- **"efficiency"**: Efficiency is the resources expended in relation to the accuracy and completeness with which users achieve specified goals. For example, a user is efficient if they are able to achieve their goals quickly (saving time) or cheaply (saving money). Examples of ways that usability allows users to be efficient are minimizing the number of user keystrokes or giving users clear paths to achieving their goals.
- **"satisfaction"**: Satisfaction is the freedom from discomfort and positive attitudes towards the use of the product. For example, a user is satisfied if they can achieve their goals with a positive frame of mind. The user may also prefer one particular system to other systems.
- **"context of use"**: The context of use constitutes the broader framework in which a product is operated. It concerns the system's particular users, their tasks and the system's broader environment of use. An example of a context of use could, therefore, be "a novice user buying a book on the Web in his home" or "an expert air traffic controller directing planes from a darkened control room." It is very important to consider context of use when designing for usability because a usable system in one context of use may be completely unusable in another.

The discipline most concerned with the implementation of computer system usability is known as Human-Computer Interaction (HCI). As defined by ACM SIGCHI (Association for Computing Machinery Special Interest Group in Computer-Human Interaction) (1992), HCI is "the discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and the study of major phenomena surrounding them." HCI is a diverse multidisciplinary field, containing theoretical principles based in disciplines such as Ergonomics, Psychology and Computer Science. HCI's main focus area is the system-user interface – those aspects of the system with which the user comes into contact. The system-user interface is the one of the most critical areas of a system under development. As Jef Raskin, the creator of Macintosh comments, "As far as the customer is concerned, the interface is the product" (Raskin, 2000).

3. What is Usability Testing?

As mentioned above, usability testing refers to evaluating the ease with which users can learn and use a product. The scope of this phrase, however, can mean different things to different people. Some people refer to usability testing as only the process that employs participants who are representative of the target population to evaluate the degree to which a product meets specific usability criteria (Barnum, 2002). Other practitioners use the term more globally to refer to any technique used to evaluate how easy a product or system is to use. The authors tend to concur with the latter definition because our research has shown that participant involvement is a desirable, but not essential, part of evaluating all aspects of system usability.

There are two main categories of usability testing: usability testing *with* users and usability testing *without* users.

3.1 Usability Testing with Users

This category of usability testing involves both usability testers and representative users. Users have specific characteristics that differ from person to person. Examples of these are their task and product skills, knowledge and experience, training level, input abilities as well as various physical and psychological characteristics. To take account of these specific user characteristics, one has to test usability with the direct involvement of the users themselves. Examples of usability testing techniques involving users are:

- **Performance Evaluation**: Performance Evaluation involves measuring users' performance while they perform tasks on a system or prototype. This technique often occurs in a formalized usability laboratory. Performance measurements (e.g. task time and error rate) serve as quantitative criteria with which to evaluate the system.
- **Card-Sorting**: Card-Sorting is a technique for exploring how users categorize and find information. It involves sessions where users sort cards containing names of topics of relevance to the system under development. This information helps develop systems with a maximized probability that users will understand its structure and find what they are looking for.
- **Field Studies:** Field Studies involve observing users using a product in its actual environment of use. Depending on the product, this could be in a user's home, office, factory or even store. Field Studies are most useful early on in the design process.

3.2 Usability Testing without Users

This category of usability testing involves usability testers alone, without the involvement of users. This is form of testing is important because, apart from users having specific characteristics that differ from person to person, they also have general characteristics that are common to most people. These are human performance factors that are independent of a user's age, gender, cultural background or level of expertise. Examples are the way we see the world around us, how much we can remember at once and certain norms of society.

According to Nielsen and Mack (1994), usability testing without users is widely used because:

- It is inexpensive
- It does not usually require special equipment
- It can be integrated easily into the product development lifecycle
- Problems and recommendations are generated without delay

Examples of usability testing techniques without users are:

- **Expert Review**: The Expert Review is a comprehensive usability guidelines audit. Important screens are sampled and inspected according to compliance with a list of thorough usability guidelines. This gives insights into how to enhance user productivity and satisfaction.
- **Competitive Analysis**: Competitive Analysis is an invaluable testing technique for products that operate in a highly competitive industry. It evaluates the usability, not only of a company's product, but also of equivalent products of key competitors.
- **Keystroke Level Model Analysis**: Keystroke Level Model Analysis is a methodology for evaluating users' data entry efficiency. This not only helps the development team identify specific aspects of a task that are slowing users down, it also can be used to compare different designs to ascertain which would allow the user to work more efficiently.

4. Cost-Benefit Value of Usability Testing

The basic assumption of a cost-benefit analysis of usability strategies is that a usable system will result in tangible, measurable benefits. There are many different areas where usability strategies can have cost-benefit value. Four of these are discussed in the forthcoming sections:

- **Development** – how usability results in optimization of the development process
- **Technical Support** – how usability results in technical support savings
- **Sales** – how usability increases product sales
- **Use** – how usability positively impacts user performance and satisfaction

4.1 Development Benefits

Usability techniques can impact positively on a company's development process in a number of ways:

- **Reduced development time**: A user-centered design approach ensures that the needs of users are built into the product development cycle. This, in turn, ensures that only relevant functionality is designed into a product – no more, no less. The user-centered design approach combats the common industry misconception that "the more features a product has, the better". An explosion of features in a system can be highly detrimental because extra features can clutter an interface and make it more difficult to find useful features. Features are also expensive to develop and, if they are not used, developing them becomes futile.
- **Reduced rework**: If usability testing is done early in the design life cycle, changes may be made to eliminate problems and achieve benefits much more cheaply because code has not been written. Thus, changing and fixing usability problems early in the development life cycle can substantially reduce rework costs later on.

Development Case Study – Israel Aircraft Industries (Bevan, 2001).

Company: Israel Aircraft Industries (IAI), a company that designs and builds aircraft and avionics equipment

Project Leader: Serco Usability Services led the project, as part of the Trial Usability Maturity Process (TRUMP) project (part-funded by the European Commission). Serco Usability Services is a London-based user-centered design consultancy.

Project: The project involved the development of a new Mission Planning Centre. The Mission Planning Centre enables a pilot to plan an airborne mission that is then loaded onto a cartridge and taken by the pilot to the aircraft. In the aircraft, the pilot loads the data into the aircraft's main mission computer.

Project Objectives: The following were the company's main project objectives:

- Improve the operational requirements definition and evaluation process
- Increase usability of IAI products
- Increase customer satisfaction from IAI products

Techniques Used: The following are some of the key usability techniques that were used:

- **Usability maturity assessment:** The usability maturity of the organization was assessed using the Usability Maturity Scale (INUSE, 1998) to identify which user-centered methods were needed.
- **Context of use analysis**: A long checklist covering many aspects of the user's skills, tasks and the product's working environment was used.
- **Evaluation of the usability of existing system**: Four users evaluated the existing system. Each user was given a mission to prepare and commented as he went along. Comments were captured by the facilitators generating a detailed list of about fifty problems. The users filled out satisfaction questionnaires after the evaluation.
- **Setting usability requirements**: Goals for task time were agreed on and a list of potential user errors was compiled.
- **Testing usability against requirements**: The system was tested against timing requirements defined for two typical tasks. Eight pilots participated in the test and filled in the satisfaction questionnaires.

Outcome: IAI estimated that the methods used resulted in savings in development costs of between $5,000 and $70,000 for each method, with a total saving of $330,000. The cost of using the methods was only $22,000, giving a cost-benefit ratio of 1:15.

4.2 Technical Support Benefits

Usable systems decrease technical support costs in a number of ways:

- **Decreased user training**: Many organizations spend large amounts of money training users on new software systems. When usable systems are designed, they become more intuitive and users require less training on the system before can they use it.
- **Decreased help-desk calls and on-site technical support**: Usability decreases the burden on a company's help-desk and on-site technical support staff because users do not require assistance addressing usability problems.
- **Decreased unofficial technical support**: Users often choose to disturb non-technical colleagues to help them solve usability problems. When a system is more usable, these disturbances are less likely to occur.

Technical Support Case Study: Ford Motor Company (Bias and Mayhew, 1994)

Company: The Ford Motor Company, a leading consumer company for automotive products and services

Project Leader: The Ford Motor Company

Project: The Ford Motor Company's accounting system for their small car dealerships.

Project objectives: The main project objective was to improve the effectiveness, efficiency and satisfaction of users of Ford's accounting system.

Techniques Used: User testing was conducted in a usability lab. A usability lab usually consists of a main room where the user performs a list of tasks, an adjacent room where the testers are located, and a one-way mirror joining the two rooms. Testers can observe the user performing tasks through the mirror. There are also cameras recording the user's performance. The testers can observe what the cameras record on monitors in the testers' room. The user's level of performance is measured during the test, for example the number of errors made, the number of tasks completed and how long the user takes to perform tasks. The users are also often asked for their opinions regarding the usability of the interface.

Outcome: Through the usability study, the following problems were discovered:

- Car dealers needed to make an average of three calls to the help-line to be able to start using the system.
- The commands used to enter credits and debits were designed by the engineers without first consulting users to learn the commonly used abbreviations.

Ford changed the abbreviations as well as 90% of their accounting system. The new system was so easy to use that the calls to the help-line dropped to zero. It was estimated that this new version saved the company $100,000. Since the cost of the usability testing was $70,000, the company gained a cost-benefit ratio of 1:14.

4.3 Sales Benefits

Usability strategies have a positive effect on company sales revenues. This is due to a number of reasons:

- **Competitive edge of marketing usable products:** Usable software can accrue increased revenues from external customers due to increased marketability. An area where this is particularly critical is that of e-commerce. On the Web, instead of buying software first and then experiencing its level of usability later, customers encounter usability first and then decide whether to pay or not. In addition, in e-commerce, the company's website becomes the crucial primary interface to the customer.
- **More satisfied customers:** When users suffer frustration caused by unusable software, the company's credibility is tarnished. In contrast, satisfied customers have brand loyalty and are much more likely to buy the same brand in the future with less researching of the particular product.
- **Higher ratings for usability in the press**: Usability is one of the main criteria used to critically rate software packages in commercial publications. A wide variety of trade magazines evaluate new software releases that are sent to market. According to Nielsen (1993), approximately 15% of reviews by trade journals is devoted to analyzing the "user-friendliness" of new software products.

Sales Case Study: Website Redesign (Nielsen J, 2003).

Companies: 42 anonymous companies

Project Leader: The Nielsen-Norman Group, a leading US-based usability consultancy

Project: The Nielsen-Norman Group analyzed data from 42 cases where usability metrics were available for website designs.

Project Objective: The main project objective was to quantify the cost-benefit value of usability redesign on websites.

Techniques Used: For this analysis, the Nielsen-Norman Group used projects that collected the same metric both before and after the redesign. This allowed them to accurately compare them and estimate the percentage improvement in usability. Each of the websites was redesigned according to well-known usability design guidelines.

Outcome: The improvement in usability metrics differed depending on the metric, as the following table (Table I) shows:

Metric	Average Improvement Across Web Projects
Sales / conversion rate	100%
Traffic / visitor count	150%
User performance / productivity	161%
Use of specific (target) features	202%

Table I. Results of website redesign study by the Nielsen-Norman Group (www.useit.com)

The results indicate a remarkable increase in number of sales or conversion rate (a site's ability to "convert" visitors to buyers) of 100%.

4.4 Use Benefits

The last area where usability strategies can have a major financial impact is in terms of user performance and satisfaction. Examples of these are:

- **Reduced user task time:** Optimizing user efficiency relates to cutting down unnecessary work-hours resulting in potentially large savings to companies. The benefits due to improved user efficiency are particularly evident for high-volume, repetitive tasks, such as data capturing.
- **Reduced user errors:** Users will not make as many errors on a usable system because they have a better understanding of the system. Typical error-prone situations can be "designed-out". This, of course, relates closely to the previous benefit of improved user task time.
- **Improved user satisfaction:** This relates to higher levels of staff morale, which have been shown to reduce staff turnover and increase job motivation and company loyalty.

Use Case Study: The Pick 'n Pay Group

Company: The Pick 'n Pay Group, one of Africa's largest and most consistently successful retailers of food, clothing and general merchandise for the past three decades

Project Leader: Test and Data Services, a leading South African software testing company

Project: The system tested was Pick 'n Pay's core Enterprise Resource Planning application.

Project Objectives: The main project objectives were:

- to estimate current user data entry efficiency
- to make design recommendations to increase user data entry efficiency

Techniques Used: Tasks were reviewed in terms of data entry efficiency using the Keystroke Level Model (Card *et al.*, 1980). This method is used to predict the time an expert user would take to complete a task on a user interface. Time estimates are attained by breaking down tasks into sub-units called operators. There are seven main types of operators possible when a user performs any task (see Table II). These predictions are based on a study involving extensive tests with users.

Operator	Definition	Time (seconds)
Keying	Time to press a key	0.12
Pointing	Time to move the input device to a target.	1.10
Homing	Time to change between mouse and keyboard.	0.40
Drawing	Time to draw a line using an input device.	1.06
Clicking a button	Time to click the input device.	0.20
Mental processing	Time to perform mental processing.	1.35
System response	Time the system takes to respond.	?

Table II. The operators and estimated times used in the Keystroke Level Model (Card *et al.*, 1980)

In this project, two frequently performed tasks were analyzed. The results from the analysis of one of these tasks ("Capturing an Order") are detailed below. This task is performed in a manual receiving environment after goods have been received at the delivery area and a hard copy of the order with its invoice has been manually written. The capturing clerk then captures the order onto the system and specifies, among other details, what goods were received, in what quantity they were received and when they were received.

Outcomes

Data Entry Method 1: Current design – using mainly the mouse

This analysis estimates how long it would take an expert user to capture an order while using the current system design which mainly uses the mouse for data entry. Table III shows the summarized calculation. The total estimated task execution time is 71.81 seconds.

Operator	Number of times	Time (seconds)
Keying	63	63 X 0.12 = 7.56
Pointing	23	23 X 1.10 = 25.30
Homing	28	28 X 0.40 = 11.20
Clicking a button	24	24 X 0.20 = 4.80
Mental Processing	17	17 X 1.35 = 22.95
	TOTAL TIME	**71.81**

Table III. The operators and estimated time for capturing an order using the current interface design and mainly the mouse for data entry

Data Entry Method 2: Improved design – using only the keyboard

The second analysis estimates how long it would take an expert user to capture an order while using an improved system design that uses solely the keyboard for data entry. From the summarized results in Table IV, it can be seen that the total task time is 33.15 seconds. This is less than half of the time it takes to perform the identical task using mainly the mouse (71.81 seconds).

Operator	Number of times	Time (seconds)
Keying	85	85 X 0.12 = 10.20
Pointing	0	-
Homing	0	-
Clicking Button	0	-
Mental Processing	17	17 X 1.35 = 22.95
	TOTAL TIME	**33.15**

Table IV. The operators and estimated time for capturing an order using an improved interface design requiring use of only the keyboard

To illustrate the organizational impact of improving task efficiency, one can estimate the amount of working time taken when performing different forms of data entry. According to in-house company statistics, an average store captures about 450 orders per store per month. Figure 2 shows the estimates for the amount of hours per month the different types of data entry would take. From these results, one can see that by performing a few simple usability design improvements to the interface stores have the potential to improve order capturing efficiency by 56%.

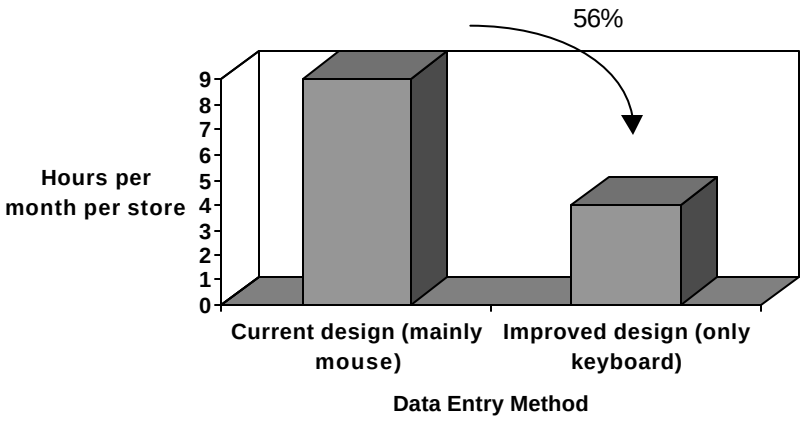

Figure 2. Estimated hours per month per store spent capturing orders using different data entry methods

5. Usability in Developing Countries

In a recent article in *Interactions*, Dray *et al.* (2003) comment, "Technology can knit the world together or it can create a schism". This is a very pertinent statement, particularly for developing countries such as South Africa (where the authors are based). In terms of a developing country, information technology can have a positive or negative contribution to economic development. On the negative side, the proliferation of technology in developing countries has been blamed for widening the gap between the rich and the poor. This takes the form of a "digital divide" between those that do and do not have access to technology.

On the positive side, developing countries can take advantage of information technology to achieve economic gains. Technology can assist in supplying basic education, improved health services and communication networks. It can also allow local businesses to market their products online to a global market. This has occurred in countries such as South Korea, Taiwan and Singapore where investments in technology have allowed these countries to establish themselves as technologically self-sufficient. These countries now export home-grown technology to other nations. HCI is one of the most important disciplines in making sure that information technology is used in a positive manner within developing countries. By designing systems to suit individuals' unique economic, political, social and cultural identity, technology can be a powerful catalyst towards national economic empowerment.

6. Conclusion

Investment into software usability testing can offer significant returns. Examples of these are in the areas of development costs, technical support costs, sales and user productivity. These benefits can only be realized through rigorous usability testing of software throughout the development process, whether it is by testers alone or with user involvement. It is inevitable that as the maturity of the software testing industry grows, non-functional testing types will be given higher priority. This holistic approach is essential in delivering products that are not only functionally correct, but that also support their users.

7. References

ACM SIGCHI (1992). **Curricula for Human-Computer Interaction**.

Barnum C (2002). **Usability Testing and Research**. New York: Pearson Education.

Bevan N (2001). **Introducing User Centered Design methods at IAI** from www.usability.serco.com.

Bias R and Mayhew D (1994) (eds). **Cost-Justifying Usability**. San Diego: Academic Press

BS 7925-1 (1998). **Software Testing. Vocabulary.**

Card SK, Moran TP and Newell A (1980). **The Keystroke-Level Model for User Performance Time with Interactive Systems**. Communications of the ACM, 23(7):396-410.

Dray S, Siegel A and Kotzé P (2003). **Indra's Net: HCI in the Developing World**. Interactions, 10(2):28-37

INUSE (1998). **Usability Maturity Model: Human-Centeredness Scale**, Lloyd's Register project IE2016 INUSE Deliverable D5.1.4s.

ISO 9241-11 (1998). **Ergonomics requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability**.

Nielsen J (2003). **Return on Investment for Usability** from www.useit.com

Nielsen J (1993). **Usability Engineering**. Boston: Academic Press.

Nielsen J and Mack R (1994). **Usability Inspection Methods**. New York: John Wiley & Sons.

Raskin J (2000). **The Humane Interface**. Boston: Addison-Wesley

Roth A (1999). **Using the Iced T Model to Test Subjective Software Qualities.** Proceedings of the 1999 International Conference on Software Testing, Analysis and Review held in San José, California.