

## USER SATISFACTION EVALUATION OF AN EDUCATIONAL WEBSITE

Goknur Kaplan Akilli  
Middle East Technical University

Author note: Goknur Kaplan Akilli, Computer Education and Instructional Technology Department, Faculty of Education. Correspondence should be addressed to, Department of Computer Education and Instructional Technologies, Middle East Technical University, İnönü Bulvarı 06531, Ankara/Turkey.  
e-mail: akilli@metu.edu.tr.

### INTRODUCTION

World changes, so do people. From the time that man-computer symbiosis began, both of them have unrecognizably changed. Who could have guessed that the ideas that rooted in 1930s, 1940s and 1960s would blossom as today's technology. Who could have guessed that, when Nelson (Baecker. et al., 1995) first coined the term 'hypertext', it would be the key that opens up gates of the "Wide World of Wonders"? As one can predict, the bricks that are used to build the new fantastic places of this world are very important. That is why what was once shaped in the hands of the designer, is now sculptured according to users. That is why 'usability' is now recognized as a vital determining factor in the success of any new computer system or computer-based service (Carvalho, 2001).

Since building a website, whether for distribution over the Internet or over an intranet, can and should be viewed as a major software development effort and one of the factors that affect the acceptability of software is its usability, it is obvious that usability does matter. Moreover, educational researchers should not overlook usability testing, if they want to develop educational software that is efficient, effective and satisfactory for the user. For achieving such specific aims it is worthwhile to know about usability methods, techniques, evaluators, when to apply usability tests, and how to plan and conduct a test, as well as the usability itself.

However, this study focuses on one particular aspect of usability, namely, user satisfaction, for an educational website used as a supportive tool for various courses by employing only one specific usability testing technique, a questionnaire.

### DEFINITION OF TERMS

#### *Usability Definition(s)*

Human-Computer-Interaction (HCI) is the area where usability is planted. Several books or papers about HCI present a definition or characterization of usability. For instance, Hix and Hartson (1993) consider usability as it is related to the interface efficacy and efficiency and to user reaction to the interface. Nielsen (1993) asserts usability as one of the parameters associated with the acceptability of any system. He articulates the acceptability of a computer system as a combination of its social acceptability and its practical acceptability. If the system is socially acceptable, it is necessary to analyze its practical acceptability within categories such as cost, compatibility with existing systems, reliability, etc., as well as the category of usefulness and employs usefulness to define usability. He defines usefulness as the issue of whether the system can be used to achieve some desired goal and further claims that it can be divided in two categories as 'utility' (whether the functionality of the system can do what is needed or in an educational hypermedia students learn from using it) and 'usability' (how well users can use that functionality). He associates five attributes to usability: easy to learn (learnability), efficient to use (efficiency), easy to remember (memorability), the relevance of prevention of catastrophic errors for applications such as process control or medical applications (few errors), and pleasant to use (satisfaction).

Shackel (1990) refers to four aspects of interest in usability testing: effectiveness, learnability (ease of learn), flexibility, and attitude. Rubin (1994) accepts that usability includes one or more of the four factors: usefulness, effectiveness (ease of use), learnability, and attitude (likeability). For Smith and Mayes (1996) usability focuses on three aspects: easy to learn, easy to use and user satisfaction in using the system (cited in Carvalho, 2001).

In international standards, usability refers to effectiveness and efficiency to achieve specified goals and users satisfaction. According to Bevan (2001)'s article, "Usability: the extent to which a product can be used by specified users to achieve a specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (ISO 9241-11)(p.536). Moreover, since in the software engineering community the term usability has been more narrowly associated with user interface design, ISO/IEC 9126, developed separately as a software engineering standard, defined usability as one relatively independent contribution to software quality associated with the design and evaluation of the user interface and interaction: "Usability: a set of attributes that bear on the

effort needed for use, and on the individual assessment of such use, by a stated or implied set of users (Bevan, 2001, p.537).”

*Usability testing*

Methodologies for building usable systems have been introduced and refined over the past fifteen or so years under the discipline of Human-Computer Interaction (HCI). HCI principles include an early and consistent focus on end users and their tasks, empirical measurements of system usage, and iterative development. Much effort has been put into exploring cognitive models of human behavior as it relates to computer usage, and developing guidelines for screen layout and system dialogues. These are predictive endeavors whose purpose is to assist the software developer in the initial task analysis and system design.

But, just as comprehensive functional requirements and a detailed design document do not by themselves guarantee that a programmer's final product will be correct, so up-front usability guidelines do not by themselves guarantee a usable end product. In both cases a distinct validation process is required.

Usability testing is the process by which the human-computer interaction characteristics of a system are measured, and weaknesses are identified for correction. Such testing can range from rigorously structured to highly informal, from quite expensive to virtually free, and from time-consuming to quick. While the amount of improvement is related to the effort invested in usability testing, all of these approaches lead to better systems.

As mentioned above, there are various methods and techniques that are used to test and measure usability. Preece (1993) articulates four usability evaluation methods that imply different types of evaluators, different number of users, and different types of data to be collected. These are expert evaluation (also known as heuristic evaluation), observational evaluation, survey evaluation and experimental evaluation. Table 1 shows the method, techniques and above-mentioned issues:

‘Expert evaluation’, also known as heuristic evaluation, is normally carried out by experienced people in interface design and human factors research who are asked to describe the potential problems they foresee for less experienced users.

‘Observational evaluation’ implies collecting data that provide information about what users do when interacting with educational software. Several data collection techniques may be used.

‘Surveys’ are employed to know users' opinions or to understand their preferences about an existing or potential product through the use of interviews or questionnaires.

In ‘experimental evaluation’ an evaluator can manipulate a number of factors associated with the interface and study their effect on user performance.

Table 1  
*Usability Testing Methods, Techniques and Evaluators (Preece, 1993)*

Method	Techniques	Type of Evaluator
Expert / Heuristic	Walk-through Questionnaires	Experts
Observation	Direct Observation Video recording Software logging Verbal protocols (Think aloud)	Experts / Users
Survey	Interviews Questionnaires	Experts / Users
Experimental	Software logging Questionnaires Interviews	Experts / Users

Other methods can also be applied such as: focus group, walk-through, paper-and pencil evaluations, usability audit, field studies, and follow-up studies (Rubin, 1994).

There are two important points here: Firstly, the researcher should always keep in his or her mind that the selection of a method has to take into account the appropriate techniques for data collection. Secondly, virtually

any kind of usability test, whatever method(s) and technique(s) are utilized, will improve the product, as long as its results are fed back to the development group and acted on (Levi & Conrad, 2001). Moreover, the researcher believes that usability testing, like most methodological process improvements, will gain attention and devotees as its benefits emerge through use.

#### *User Satisfaction*

As can be seen from Table 1, the observational, survey and experimental methods imply the presence of users. In addition, users' individual characteristics and differences are important issues for usability. 'User satisfaction' is mentioned as preference data represent measures of participant opinion or thought process, whereas user's 'performance data' correspond to measures of participant behavior, focusing on aspects such as 'efficiency and efficacy of use.' User satisfaction includes participant rankings, answers to questions, and so forth. Rubin (1994) points out some aspects to measure, for example, usefulness of the product, how well product matched expectations, ease of use overall, ease of learning overall, ease of set up and installation, ease of accessibility, usefulness of the index, table of contents, help, graphics, and so on. User satisfaction can also be measured through a comparison between two products or two versions of the same product. There are several tests for evaluating the user satisfaction. Examples of these are SUMI (Software Usability Measurement Inventory) and QUIS (Questionnaire for User Interface Satisfaction) (Kirakowski, 1996). More recently and due to the rapidly changing patterns and technology of computing today, two new questionnaires are being developed, MUMMS (Measuring the Usability of Multi-Media) to assess multimedia software and WAMMI (Website Analysis and Measurement Inventory) to assess web sites (Levi & Conrad, 2001).

### **PURPOSE OF THE STUDY**

The researcher aimed to find out whether eighth semester undergraduate students of Computer Education and Instructional Technologies (CEIT) Department at the Middle East Technical University (METU) Ankara, Turkey, are satisfied with the website that is used as a supportive tool for a traditional classroom. Based on the findings from this study researcher hopes to provide web interface designers with some empirical support, especially about the powerful and weak attributes, in case of designing a website with similar facilities and properties.

#### *Research Questions and Subquestions*

The study addressed the following research questions related to students' use of website of the course as a supportive tool.

1. How are the overall reaction of users towards the website?
  - 1.1 To what extend are they impressed by the website?
  - 1.2 To what extend are they satisfied with the website?
  - 1.3 To what extend are they stimulated by the website?
  - 1.4 Is the website easy to use for them?
  - 1.5 Do they perceive website as 'powerful'?
  - 1.6 Do they find the website flexible?
  - 1.7 Which of the duples of the above-mentioned overall reaction issues are users more concerned with?
  - 1.8 Are there any relationships among these properties of the website?

### **METHOD**

#### *Procedure*

Students enrolled in "CEIT 419 Internet for Teachers" undergraduate course in the Computer Education and Instructional Technologies (CEIT) Department at the Middle East Technical University (METU), Ankara, Turkey, were invited to participate in a study designed to understand the user satisfaction levels of a website used as a supportive tool for a course in a traditional classroom. The researcher administered questionnaire during two hours on the ninth week of the semester due to the nature of the questionnaire, since the questionnaire is typically offered to users after they have completed a session of work with a particular system or program. Students were informed verbally and briefly on the research topic and the questionnaire. Participation of the students was voluntary since confidentiality was guaranteed (i.e., students did not place their name on any of the materials in the study), and by returning the survey they were giving their informed consent to allow the researcher to use their data as part of the study.

#### *Participants*

Participants consisted of 33 out of 37 (30% female, 70% male) students enrolled in "CEIT 419 Internet for Teachers" undergraduate course of CEIT department at METU. Ages of the participants ranged from 20 to 24 with a mean age of 22 (SD =.92). Table 2 illustrates the participants' profile, that includes their experience with

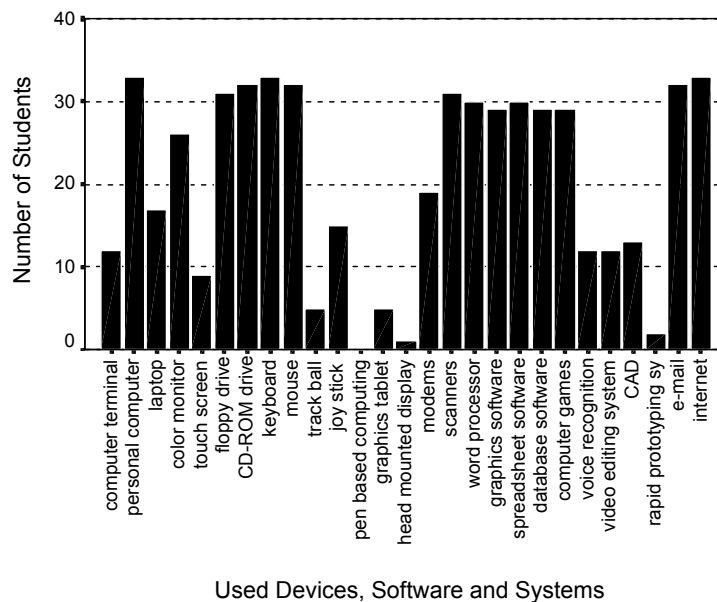
the website, such as duration of time they are working, average time that they spend working; and their past experiences, such as number of operating systems that they worked with. Figure 1 shows the frequencies of various devices, software and systems that participants have used or been familiar with.

*Materials*

Researcher employed the Questionnaire for User Interaction Satisfaction (QUIS) based on OAI (Object-Action Interface) model, developed by Shneiderman in the Human-Computer Interaction Laboratory at the University of Maryland and refined by Norman and Chin (Schneiderman, 1998). Since the evaluation of a system's accuracy is fairly straightforward, the assessment of the user's satisfaction with the human-computer interface is a subjective and complex question, the Questionnaire for User Interaction Satisfaction (QUIS) was created to gauge the satisfaction aspect of software usability in a standard, reliable, and valid way. The QUIS was first implemented as a standard paper and pencil form using a nine point Likert scale (Chin, Diehl, & Norman, 1988).

Table 2  
*Participants' profiles*

System Experience		
Duration of time they are working	Frequency	Percentile
1 hour to less than 1 day	1	3,0
1 hour to less than 1 day	1	3,0
1 day to less than 1 week	3	9,1
1 week to less than 1 month	1	3,0
1 month to less than 6 months	24	72,7
6 months to less than 1 year	1	3,0
3 years or more	3	9,1
Average time spent on the system per week		
less than one hour	4	12,5
one to less than 4 hours	23	71,9
1 day to less than 1 week	5	15,6
Past Experience		
Number of Previously Worked Operating Systems		
1	6	18,2
2	13	39,4
3-4	11	33,3
5-6	2	6,1



*Figure 1.* Number of participants that are familiar with various devices, software and systems.

The QUIS focuses on the user's perception of interface usability by the evaluation of specific aspects of the interface (i.e., overall reaction to the system, screen factors, terminology and system feedback, learning factors, system capabilities).

The QUIS 7.0 is an updated and expanded version of the previously validated QUIS 5.5. The Questionnaire for Interaction Satisfaction (QUIS) version 7.0 is arranged in a hierarchical format and contains: (1) a demographic questionnaire, (2) six scales that measure overall reaction ratings of the system, (3) four measures of specific interface factors: screen factors, terminology and system feedback, learning factors, system capabilities, and (4) optional sections to evaluate specific components of the system: technical manuals and on-line help, on-line tutorials, multimedia, Internet access and software installation. Each of the specific interface factors and optional sections has a main component question followed by related sub-component questions. Each item is rated on a scale from 1 to 9 with positive adjectives anchoring the right end and negative anchoring the left. In addition, "not applicable" is listed as a choice. Additional space, which allows the rater to make comments, is also included within the questionnaire. The comment space is headed by a statement that prompts the rater to comment on each of the specific interface factors (Harper, et al., 1990).

Moreover, it can be used as a whole or in parts and with addition of domain specific items (Schneiderman, 1998). Although statistical reliability, cross-correlations, and benchmarking have not, to researcher's knowledge, been achieved or independently assessed for the current version (Version 7.0) of QUIS, Kirakowski (1996) reported the reliability of the QUIS Version 5.5 as .94.

#### *Design*

This study is planned as a survey research by employing the QUIS to collect the data. However, researcher selected to use only the demographic part of the questionnaire and six scales that measure overall reaction ratings of the system, results of some sections were appeared to be unsound and not meaningful and some parts of the questionnaire were not applicable to the website. Moreover, the open-ended questions are also excluded from the selected parts, since there was only one participant that write some comments about the website.

Since the QUIS has proven to have high reliability with low variability, convenience sampling method is used for sample selection. This choice of the researcher is also appropriate for the theoretical population (Turkish undergraduate students who take web-supported courses in traditional classroom environments) and target population (Turkish undergraduate students who take web-supported courses in traditional classroom environments utilizing the mentioned website as a supportive tool) of this study. Accordingly, the sample of this study is Turkish undergraduate students of CEIT department of METU that take a specific web supported course in a traditional classroom environment utilizing the mentioned supportive tool.

The study has various dependent variables. For analysis, six scales that measure overall reaction ratings of the system, are assigned as dependent variables. Moreover, before the statistical analysis was conducted by employing the Statistical Package for the Social Sciences (SPSS), the researcher utilized SPSS to have the missing values completed.

## **DISCUSSION**

### *Results and Analysis of Results*

Due to the nature of this study less emphasis will be placed upon inferential statistics, as there is no system to which the current system is being compared.

Simple error bar charts were created to display a confidence interval around each item mean related to 'overall reaction rating part of the QUIS in order to determine its reliability, since the statistical reliability, cross-correlations, and benchmarking have not, to researcher's knowledge, been achieved or independently assessed for the current version (Version 7.0) of QUIS. Moreover, these bar charts also indicated whether the mean of an item is significantly above or below the criterion, selected as the overall mean of the related part. Paired samples *t* tests were conducted for items that measure users' overall reaction to evaluate the degree of the users' concern about impressiveness of the site, satisfaction, the feeling of being stimulated, ease of use, perceived powerfulness and the flexibility of the website.

*Overall Reaction Ratings.* Two of the six scales that measure overall reaction to the system were rated lower than the mean response ( $M = 6.17$ ). These factors were website's stimulating attributes and flexibility indicating that these areas are subject to additional scrutiny. The other four overall ratings, namely, impressiveness, satisfaction, ease of use and perceived powerfulness of the website were not less than the user response level.

Depending on the above mentioned results the researcher concluded that users found the website somewhat rigid and lack of stimuli. The most outstanding property of the system was the ease of use with the highest mean ( $M = 6.52$ ). Table 3 presents the means and the standard deviations of each item in overall reaction rating part.

Table 3  
*Means and Standard Deviations for Items in Overall Reaction*

3. Overall Reaction	<i>M</i>	<i>SD</i>
Item 3.1. Impression	6,21	1,19
Item 3.2. Satisfaction	6,29	1,33
Item 3.3. Being stimulated	6,03	1,40
Item 3.4. Ease of use	6,52	1,97
Item 3.5. Perceived 'powerfulness'	6,26	1,82
Item 3.6. Flexibility	5,68	1,72

A simple error bar chart was created to determine the reliability of the items in overall reaction rating part. The plotted 95% confidence interval that included the overall mean of 6.17 within its boundaries indicated that the means of each particular item was not significantly different from 6.17 at the .05 level of significance (Figure 2).

Distinct paired samples *t* tests were conducted to evaluate the degree of users' concern for each duple of impressiveness of the website, satisfaction of the users, the feeling of being stimulated, ease of use, perceived powerfulness and the flexibility of the website. The results indicated that the mean concern for satisfaction ( $M = 6.29$ ,  $SD = 1.33$ ), mean concern for ease of use ( $M = 6.52$ ,  $SD = 1.97$ ), and mean concern for perceived powerfulness ( $M = 6.26$ ,  $SD = 1.82$ ) were significantly greater than the mean concern for flexibility ( $M = 5.68$ ,  $SD = 1.72$ ),  $t(32) = 2.11$ ,  $p = .04$ ;  $t(32) = 2.62$ ,  $p = .01$ ;  $t(32) = 2.49$ ,  $p = .02$  respectively. The standardized effect size indexes (*d*) were .37, .46 and .43, respectively, indicating medium values of effect size. The mean difference was .61 between the two 9 point Likert ratings for satisfaction and flexibility; .83 points between the two 9 point Likert ratings for ease of use and flexibility; and .58 points between the two 9 point Likert ratings for perceived powerfulness and flexibility. Let alone considerable overlapping, the distributions of ease of use and perceived powerfulness encompassed the distribution of flexibility, whereas vice versa was true for the distributions of satisfaction and flexibility, as shown in Figure 3.

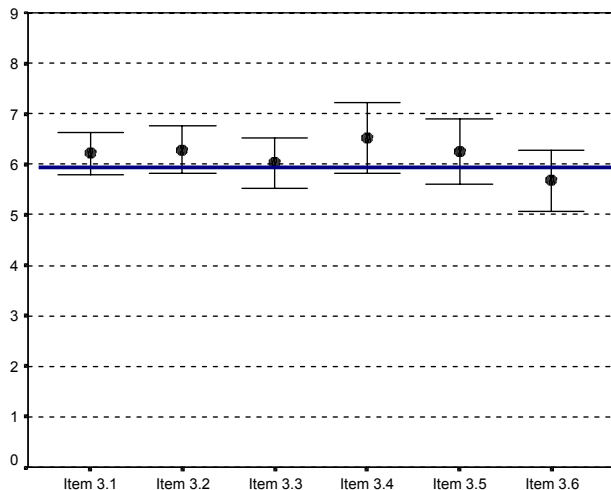


Figure 2. Distributions of six scales that measure overall reaction ratings of the system in a 95% confidence interval.

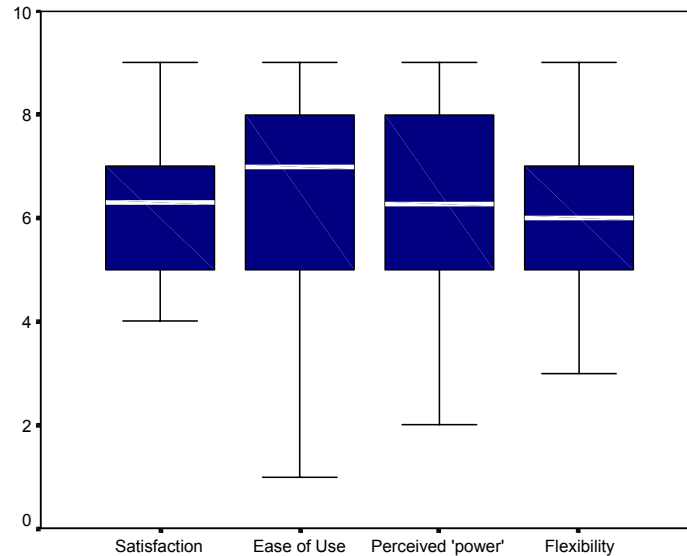


Figure 3. Boxplots of satisfaction, ease of use, perceived powerfulness and the flexibility ratings.

#### LIMITATIONS AND DELIMITATIONS OF THE STUDY

The use of convenience sampling method and homogeneous structure of the sample made the obtained results difficult to generalize to a larger population. Additionally, participants were familiar to various kinds of researches, which might give birth to threads to internal validity of the study due to 'subject characteristics' and 'location.' Nevertheless, conducting the analysis of data two weeks after the collection and avoiding leading instructions or questions kept threads of data collector characteristics and data collector bias away.

Another limitation was the duration and the course of the study. Time was the biggest limitation to conduct efficiency and effectiveness tasks to complete the usability evaluation of the website.

One delimitation of the study was the familiarity of researcher to participants. It would have been better to utilize administrators trained for this purpose, but again due to lack of time, this could not have been possible.

#### SUGGESTIONS FOR FURTHER RESEARCH

Some suggestions for extending this study might be utilizing the same user satisfaction questionnaire with additional tasks for efficiency and effectiveness to complete the puzzle of the designated website's usability evaluation. Moreover, a comparative study of the designated website and another educational website, the usability of which was evaluated, might be conducted to diagnose lacking parts of the former.

The same study or the extended version may be conducted with larger sample, different groups of users or interfaces designed for different courses.

Eventually, another study might be conducted that covers some special challenges of the web, such as wide disparity in connectivity speed, deployment environment which blurs the distinction between the site content and the browser used to access the content, etc. (Levi & Conrad, 2001) to clarify the usability picture of the websites.

#### CONCLUSION

The results of the study indicated that the users were initially impressed and satisfied with the website. Additionally, they found the website easy to use and powerful, in spite of the lack of flexibility and stimulating attributes of the website. Moreover, experience of the researcher showed that usability testing is time consuming and demands a meticulous planning.

The researcher recognized from the results of this study that there are still many questions, which are unanswered and open to further investigation by researchers and careful consideration by website designers. However, achieved results compensate greatly!

#### REFERENCES

- Baecker, R. et al. (1995). *A Historical and Intellectual Perspective*. In Baecker, R., Grudin, J. Buxton, W. and Greenberg, S. (eds.), *Readings in Human-Computer Interaction, Toward the Year 2000*, NY: Morgan-Kaufman, 35-47.
- Bevan, N. (2001). *International Standards for HCI and Usability*. *International Journal of Human-Computer Studies*, 55, p. 533-552.
- Carvalho, A. A. A. *Usability Testing of Educational Software: Methods Techniques and Evaluators*. Retrieved on December 29 2002 from <http://www.esev.ipv.pt/3siie/actas/actas/doc15.pdf>
- Chin, J.P., Diehl, V.A., & Norman, K.L.(1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *CHI '88 Conference Proceedings: Human Factors in Computing Systems* (New York, 1988), ACM Press, pp. 213-218.
- Harper, B., Slaughter L. & Norman K. (1990). *Questionnaire administration via the WWW: A validation & reliability study for a user satisfaction questionnaire*. Retrieved on December 22 2002 from <http://www.lap.umd.edu/webnet/paper.html>
- Hix, D. & Hartson, H.R. (1993). *Developing User Interfaces: Ensuring Usability Through Product and Process*. New York: John Wiley & Sons.
- Human Computer Interaction Laboratories of University of Maryland. *New System Evaluates Human-Computer Interface*. Retrieved on December 22 2002 from <http://lap.umd.edu/quis/publications/newspaper1.html>
- Kirakowski, J. (1996) *The Use of Questionnaire Methods for Usability Assesment*. Retrieved on December 30 2002 from <http://www.ucc.ie/hfrg/questionnaires/sumi/sumipapp.htm>
- Levi M. D. & Conrad F. G. (2001). *Usability Testing of World Wide Web Sites*. Retrieved on January 2 2002 from [http://www.bls.gov/ore/htm\\_papers/st9601050.htm](http://www.bls.gov/ore/htm_papers/st9601050.htm)
- Nielsen, J. (1993). *What is Usability?* In 'Usability Engineering', New Jersey. Academic Press.
- Preece, J. (1993). *A Guide to Usability: Human Factors in Computing*. Addison Wesley, the Open University.
- Rubin, J. (1994). *Handbook of Usability Testing*. New York: John Wiley and Sons.
- Schneiderman, B. (1998). *Designing the User Interface*. 3rd Edition. Addison Wesley Inc., California.
- Shackel, B. (1991). *Usability - Context, Framework, Definition, Design and Evaluation*. In Shackel, B. and S. Richardson (eds.), 'Human Factors for Informatics Usability,' Cambridge: Cambridge University Press.