Web Usability and Evaluation: Issues and Concerns

S. Batra[2] and R.R. Bishu[1]

[1] Department of Industrial Engineering, University of Nebraska-Lincoln, Lincoln, NE
[2] Cognitive Ergonomist, Enterprise Rental Inc., St. Louis, USA

Abstract. This paper presents a summary of usability work done at the Usability Laboratory at University of Nebraska in the last few years. The main objective of the first study was to compare the efficiency and effectiveness between user testing and heuristic analysis in evaluating four different commercial websites. The results showed that both user testing and heuristic analysis addressed very different usability problems and both methods are equally efficient and effective. In the second study. the primary purpose was to compare the performance between remote usability testing and traditional usability testing. The results indicate that remote usability testing is no different from traditional usability testing. The third study attempted to look at cultural differences in web usability. The results indicated that cultural dimensions have significant effects on user's web preferences. The primary objective of final study was to determine if user's surfing behavior could be predicted through their cognitive style. Results show that cognitive span scores are not strong enough to form association rule with individual difference clusters of web surfing behavior. The results are discussed with respect to all perspectives of Web.

1 Introduction

The Last three decades have seen the communication field go through a technological revolution. Internet has changed professional and domestic life completely. Web based applications have become a standard, cross-platform, nonproprietary means for businesses to communicate with each other and with consumers. This in turn has made the design of a website and a web interface very important and an integral part of contemporary commerce. There are a number of issues with global development of web based commerce. The most important issues are evaluation methods and cultural differences among users. Over the years a number of modular studies have been done at the Department of Industrial Engineering, University of Nebraska. This presents an overview of some of the studies that have some ramification for web usability.

2 Comparison of Heuristics Evaluation and Usability Testing

Usability plays a crucial role as user experience is emphasized above anything else. Thus, it is essential to create a well-designed Website that is highly usable. The question is one of deciding what constitutes a well-designed site and how to evaluate the same? Different usability evaluation techniques have been developed and incorporated

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 243–249, 2007.

into the design and development of Websites. Among these techniques user testing and heuristic analysis are perhaps two of the most popular ones.

The main objectives of this study were:

1. Evaluate four commercial websites with both user testing and heuristic analysis,
2. Compare the efficiency and effectiveness of both the methods

A total of four commercial websites were evaluated for this study, the first two websites were considered to have average number of usability problems while the other two were considered to have high number of usability problems. Bad websites represented the user interface in the early stages of the development process that have abundant amount of usability problems. On the other hand, good websites represented the user interface in the later stage of the development process with lesser usability problems compared to the bad websites.

A total of 5 scenarios that represented typical site usage situations that might be encountered in real life were given to both the users and evaluators. The scenarios for both user testing and heuristic analysis were essentially the same but varied in terms of the degree of detail. The scenarios for user testing were more detailed while the scenarios for heuristic analysis were more open ended. A total of 12 users for user testing, and 9 evaluators for heuristic analysis participated in this study.

The analysis revealed several important findings as detailed below:

1. First of all, the proportions of common problems were small for both good and
2. bad websites, which were approximately 7% and 11%,
3. Second of all, proportion of the common problems found increased from good to
4. bad websites, about 4%.
5. Third of all, heuristic analysis found the most problems in both types of websites,
6. approximately 58%-61%.
7. Fourth of all, heuristic analysis is equally effective in both good and bad websites
8. environment.
9. Finally, user testing found less problems in bad websites compared to good
10. websites. These conclusions were consistent throughout all four websites.

The main premise for this study was that both user testing and heuristic analysis should be used as both an evaluation tool and as a method for guiding design improvements. The study confirmed the premise. The results indicated that both methods are needed as both methods find very few common problems. It is known that both user testing and heuristic analysis are based on very different fundamentals. User testing relies mainly on the experience and comments of the users and is usually conducted in a scenario-based environment. As a result user testing would usually evaluate according to what already exists, rather than to what is possible. On the other hand heuristic analysis relies mainly on the expertise and knowledge of human-factors engineer that would evaluate the web site based on a set off heuristics. Therefore both methods find different types of problems.

In summary, both user testing and heuristic analysis are needed in a usability study. In order to reap the optimal benefits, both user testing and heuristic analysis should, preferably, be used in different stages of the user interface design process.

3 Comparison of Remote Testing Versus Direct Testing

Usability testing continues to be the primary method of evaluation of web sites. Though time consuming and resource intensive, they are the most predominantly used. Usability studies are typically performed under direct supervision of the experimenter, under a controlled environment, which often poses a constraint. Further, the down side of running the evaluation in a controlled environment is that it does not recreate the Internet browsing environment from the subject's own computer. A good way to solve this problem is to have the subject evaluate the website on his/her own computer and in his/her own environment. Scenarios are mailed to the subject and the evaluator can monitor the evaluation process from his/her lab through remote access. *The working hypotheses was that there would be no differences between direct and remote usability.*

Two groups of student subjects (Asians and Americans) participated in this experiment. Two types of usability testing (direct and remote) were performed on two sites. In the remote testing, the subjects were in their own room separated from the moderator. The subjects answered the user profile questionnaire, evaluated the websites through 5 scenarios, and did a post evaluation questionnaire. As the subjects were doing the evaluation, PCAnywhere captured the subjects' computer screens in the form of video and the video was saved in the moderator's computer. In the direct testing, the evaluation took place at the Department of Industrial Engineering Ergonomics Lab. The subjects and the moderator were physically in the same room. The procedure was identical to the remote testing procedure with one difference, when the subjects got stuck in any of those scenarios, the moderator was physically there to give guidance to the subjects.

All of the results and analyses in this study suggested that remote usability testing provided the same result as traditional usability testing regardless of ethnicity. Usability testing has become one of the important tools in evaluating a website and many engineers are looking at ways to cut costs without compromising the effectiveness of the method. By using the method of remotely conducting the usability testing, costs can be cut down tremendously. There was no interaction between sites and method, and between ethnicity and method, suggesting the robustness of this method.

4 Cultural Differences in Web Usability

As per the name, World Wide Web has evolved as a medium for international communication, participation and transaction for a multi-cultured environment. Usability studies in past have been mainly limited to language impact on global market. Site visitors must be able to navigate freely, confidently and comfortably through a site in order to find, enjoy and make use of its contents. This requires the web designers to consider the cultural background and behavioral pattern of the users.

Culture is a learned phenomenon, which derives from one's social environment. It is the collective programming of the mind which distinguishes the members of one group or category of people from another. A dimension is an aspect of a culture that can be measured relative to another culture. Geert Hofstede in his classic study of cultures proposes that each national culture has five major dimensions which can be

its identification mark. They are Power Distance, Collectivism vs. individualism, Femininity vs. masculinity, Uncertainty avoidance and Long vs. short term orientation.

The objectives of this study revolved around cultural dimensions effecting web usability.

1. To measure the quantitative attributes for different cultures based on their cultural dimension scores.
2. To find whether the web performance of people differs based on their cultural background in terms of time taken to accomplish the given task and number of pages visited during the task.
3. To find whether people have different web preferences in accordance with their cultural dimensions

To measure the cultural dimensions, Value Survey Model questionnaire, developed by Geert Hofstede, was used. The questionnaire contains 25 questions which measures 5 dimensions for each culture. The experiment contained total of 4 tests to be performed by each of 20 participants from 3 nationalities. Subjects were first asked to solve the Value Survey Model (VSM) questionnaire. It was followed by Web Stereotype questionnaire. Then they were asked to surf the sites identified for the study and to complete a task following the simple guidelines. This task was monitored by Web Logger software. Finally a card sorting test was performed to determine ethnic differences, if any, in developing information architecture.

This study was based on the hypothesis that culture does have an important effect on web usability. The quantitative analysis was carried out to find the significance in total task time, number of pages visited and ethnicity. The descriptive analysis was performed to establish the relation between various cultural dimensions and user's web preferences. By analyzing VSM scores for each ethnicity and by using Marcus guidelines, we could interpret the web behavior for each country. The regression results between task time, number of clicks and ethnicity were quite promising. The ethnicity showed significant relation with total task time and number of clicks. Results concluded that ethnicity does have an important role to consider in web usability. Web surfing performance has shown significant difference in total task time according to user's cultural background. But the study has failed to show a distinct difference in web stereotype questionnaire and so user's web preferences. It also lacks some more research on card sorting exercise where it is unable to indicate the subtle differences in spatial arrangement of the information. As many users are not the web designers, they tend to arrange the card sorting data according to their impression of the other popular sites.

In summary, this study has contributed towards progress of web usability field by adding the information that cultural dimensions should be considered while designing the web sites for international users to capture global attention and increasing the profitability of the business.

5 Web Personalization Study

Web personalization can be described, as any action that makes the Web experience of a user personalized to the user's taste. This experience can be something as casual as browsing the Web or as economically significant as trading stocks or leasing an

apartment. User satisfaction is the ultimate aim of personalization. The primary objective of this study was to determine if user's surfing behavior could be predicted through their cognitive style. Stated formally, the objectives were:

Hypothesis I: Cognitive styles affect web surfing styles
Hypothesis II: Presence of discrete relation between individual cognitive styles and web surfing behavior.

A number of methodological details had to be sorted out before data collection. They are:

1. A general web surfing task had to be designed as to capture all complexities of web surfing behavior,
2. A method of capturing objective surfing data had to be designed, and
3. A battery of independent tasks that would be a valid predictor of surfing behavior had to be designed

A task was to be designed with following features in its way of execution. It should give subject sufficient sense of control with reasonable complexity so as to capture reasonable amount of data and there should be no time limit which might limit subjects natural response to task. Task was designed within Cognitive design model prescribed by Norman (1991). This study wanted to compare cognitive spans with cognitive processing in web surfing, which needs web-browsing data to be in comparable format to span scores from psychometric tests. Weblogger software developed by XEROX PARC was found with such capability. Cognitive spans were selected as mental measures for individual differences. Three spans, i.e., counting spans, operation span and reading span were used. Individual differences that are expected in these dimensions of working memory capacity primarily reflect differences in capability for controlled processing. Total 50 subjects participated in the experiment.

The analysis was based on four variables: three span scores and time. At first, descriptive analysis was carried out to check the suitability of data. Regression of span scores against total task time highlighted their interaction. The next phase of analysis was divided into task study, between page analysis and within page analysis. Task study consisted of analysis to develop predictive relation cognitive spans and total task time. Between pages analysis studied the individual web page and its interaction with neighboring pages in URL maps. Time span for each web page was analyzed with mean time analysis across the subjects' cognitive span scores. Finally, within page analysis was consisted of analysis of subject's performance for individual web page.

This study was carried out with intention of finding predictive capacity of cognitive style of individuals for their web surfing behavior. It was based on hypothesis that website surfing is dependent on users cognitive span (attention span). This study had more interest in finding whether the user coming to site falls in one of the groups of spans, which can help predict further course for such user, basically adapting the site to the individual. Results concluded cognitive span do predict surfing behavior of individual to some extent. For high cognitive ability individuals, web surfing performance was very much improved. This study failed to clearly define low cognition individuals who had their performance overlapped by medium span individuals. Psychometric tests were used to measure and quantify this natural performance in web

surfing. But the tests failed to capture the natural performance in there scores. Psychometric testing needs improvement in its scope to capture human cognition free of outside influences. Personalization can go to much higher extent in helping the user navigate efficiently and achieve the goal. But the designer still needs to answer whether in spite of all the efforts, will it be economically feasible to retain and maintain personalization soft wares and tools?

6 Discussion

What do all these mean for a web designer and web user? Web has come to stay. Ever since real time interaction on the web became a reality, growth in World Wide Web has been mind boggling. It has rendered commerce and all aspects of life truly global. With regards to web, there are four perspectives: that of a user, designer, developer and main organization which wants web presence. From a user view point two issues stand out: a) the experience of using the web (a particular site for example) has to be a pleasant one, and b) the interaction (transaction) has to be natural. The web site, for the organization seeking web presence, has to be a) simple, b) exhaustive, c) ensure customer retention and d) provide a pleasurable user experience. The designer has to have a good process in place for designing the site. The designers should always remember that competition is just one click away. Finally for the developer and the maintainer the web has to be easy to maintain and robust. Our studies, those summarized here and others, deal more with designers' and users' perspectives.

It is clear the both heuristic evaluation and usability testing are valid testing. They are *complementary* methods of evaluation and not *competing* methods. They identify different sets of problems, all of which have to be corrected before the final release of the web site. Both methods are equally efficient and effective in addressing different categories of usability problems.

It is also apparent that there are no differences between remote usability testing and traditional usability testing. This has two ramifications, reduced cost and better fidelity. The site can be tested with actual potential users rather than with trained usability subjects. From the next study reported here, it is clear that significant cultural differences exist in the manner the subjects perform the tasks on the websites. The results also indicated that cultural dimensions have significant effects on user's web preferences. Web designers need to consider the cultural background of the target users while designing the websites. Finally, the study on personalization suggests cognitive span scores are not strong enough to form association rule with individual difference clusters of web surfing behavior.

Bibliography

1. Deshmulkh, N.: Personalization of Webs sites, A Master's theses, Department of Industrial Engineering, University of Nebraska (2002)
2. Dixit, A.: Cultural Differences in web usability, A Master's theses, Department of Industrial Engineering, University of Nebraska (2003)
3. Liew, W.-C.: UsabilityTesting: Remote Vs. Direct. A Master's theses, Department of Industrial Engineering, University of Nebraska (2002)

4. Tan, W.S.: Comparison of heuristic evaluation and usability testing, A Master's theses, Department of Industrial Engineering, University of Nebraska (2002)
5. Deshmuk, N., Dixit, A., Bishu, R.R.: Web Personalization: Study Of Effect of Cognitive Style. In: The Proceedings of the Fourth International Conference on Work with Computing System, pp. 553–557 Kualalampur (2004)
6. Dixit, A., Bishu, R.R.: Cultural Differences in Web Usability. In: The Proceedings of IEA Congress, Seol, Korea. (Page numbers not provided) (2003)
7. Deshmukh, N.B, Dixit, A, Bishu, R.R.: Web personalization: Study of effect of cognitive style on web surfing, Presented at the 48th Annual Meeting of the Human Factors and Ergonomic Society, Denver (2003)
8. Liew, W.-C., Bishu, R.: Web Usability: Remote versus Direct Testing. In: The Proceedings of IEA Congress, Seol, Korea. (Page numbers not provided) (2003)
9. Tan, W.S., Bishu, R.R.: Which Is a Better Method of Web Evaluation? A Comparison of User Testing and Heuristic Evaluation. In: The Proceedings of the 46th Annual Conference of Human Factors and Ergonomics Society, pp. 1256–1260, Baltimore (2002)