Enhancing Usability Testing Skills of Novice Testers:
A Longitudinal Study

Mikael B. Skov   *Jan Stage*

Department of Computer Science   Department of Computer Science
Aalborg University, Denmark   Aalborg University, Denmark
dubois@cs.auc.dk   jans@cs.auc.dk

Abstract

Universal access in web-sites requires dissemination of usability engineering competence to a wide audience in order to make the web-sites useful. Many web-sites have severe usability problems that prohibit effective and successful use. Identification of web-site usability problems in dedicated usability tests is a difficult and challenging task for novice usability evaluators. In this paper, we try to address this issue by reporting from a longitudinal study where we try to train the analytical skills of novice usability testers. The result is that it is possible to enhance novice usability testers' ability to identify more usability problems in a test and to describe the identified problems better. Additionally, this kind of training seems to avoid that these novice usability testers will fail the task of identifying usability problems.

1   Introduction

The World Wide Web represents a major move towards the ideal of universal access where information is made available to anyone with a computer that is connected to the Internet. World Wide Web supports thus the requirement for information that is available and accessible to anyone, anywhere, anytime. However, we need to ensure e.g. usability in order to achieve this universal access. On the other hand, research studies indicate that more web-sites are difficult to use (Rubin, 1994) and people face difficulties in e.g. buying products on e-commerce web-sites.

Research indicates that work practices in web-site development seem to largely ignore the body of knowledge and experience that has been established in the disciplines of software engineering, human-computer interaction, and usability engineering (Sullivan & Matson, 2000). Many web sites are designed and implemented in fast-paced projects by multidisciplinary teams that involve such diverse professions as information architects, Web developers, graphic designers, brand and content strategists, etc. Such teams are usually not familiar with established knowledge on human-computer interaction (Braiterman, Verhage, and Choo, 2000). Hence, web-site development projects would typically include novice usability testers.

We need to offer novice usability testers means for creating usable web-sites (Skov and Stage, 2001). However, such novice usability testers face key difficulties in the conduction of usability evaluations, e.g. they encounter problems of identifying, describing, and classifying usability problems – a crucial activity in usability testing (ibid.). This paper outlines a study where we specifically want to train and stimulate the analytical skills of novice usability testers. By doing this, we wish to explore whether this will improve their capabilities of identifying, describing, and classifying usability problems.

2   Method

The training of the usability testing skills took place in courses at Aalborg University, Denmark in 2000 and 2001. The purpose of these courses was to teach and train students in fundamentals of computerized systems with a particular emphasis on usability issues. The course comprised ten class meetings. One class meeting would last four hours that are roughly divided into a two-hour lecture for the whole class and two hours of exercises that are carried out in smaller groups. Thus, the total time spent for each individual student on the course should be around 40 hours.

The content of the course was designed to cover the required general issues of fundamentals of computerized systems and usability issues. The two primary techniques were presented in the course. The first one was the think-aloud protocol, which is a technique where test subjects are encouraged to think aloud while solving a set of tasks by means of the system that is tested, cf. (Nielsen, 1993; Molich, 2000). The second technique is based on questionnaires that test subjects fill in after completing each task and after completion of the entire test, cf. (Spool, 1999). Additional techniques such as interviewing, heuristic inspection were also presented. The exercise time was allocated to experiments. In the first exercises, they were to try out the two taught usability evaluation methods on small examples, but as soon as possible that should start to work in the exercise time on the main experiment. The experiment was limited by the exercise time in order not to conflict with other activities on the semester. We chose the hotmail.com web site as the subject for this evaluation. From the functional point of view, we chose this site because it involves a high degree of interaction, and this interaction involves both data entry and display.

Based on the experiences from the first study in 2000 (Skov and Stage, 2001), we decided to change the course slightly in order to explicitly address issues of identifying, describing, and classifying usability problems. This was primarily done in two ways. First, we designed an entire lecture to the identification of usability problems and we gave examples of problems at different kinds of web-sites. These problems were furthermore discussed with the students. Secondly, smaller exercises were given during the lectures to promote and train their analytical skills by letting them work with specific problems in different kinds of web-sites. The topics of the additional individual lectures can be found in (Skov and Stage 2001).

2.1   Participants

The course was given in 2000 and 2001 by the same two instructors. In 2000, the course was followed by 234 students that were grouped into 36 teams. In 2001, the course was followed by 174 students that were grouped into 27 teams. Each team consisted of app. 6-7 students. Figure 1 summarizes key characteristics of the teams in 2000 and 2001.

	No. of students	No. of teams	Team average size	Min/Max
2000	234	36	6.5	4/8
2001	174	27	6.4	4/7

Figure 1. Characteristics of the participating student teams.

2.2   Procedure

Each student team appointed one test monitor and 1-2 loggers among their own team members. The test monitors and loggers of all teams would then receive an email stating a two-page scenario

that described the web site and included a list of the web site features that they were supposed to evaluate. The test monitor and loggers were to examine the system, design task assignments for the test subjects, and generally prepare the test. The usability test sessions were planned to last approximately one hour for each test subject. Due to the pedagogical approach of the university, each team has their own office that was used for the test. The office was equipped with a personal computer and Internet access. After the test, the entire team worked together on the analysis and identification of usability problems and produced a usability report. The purpose of the report is to identify usability problems on the web site that was evaluated and the usability report should consist of an executive summary (1 page), description of the approach applied (2 pages), results of the evaluation (5-6 pages), and a combined conclusion and discussion of methodology (1 page).

2.3 Data Analysis

The usability reports were the primary source of data for our empirical study. All reports were evaluated and marked by both authors of this paper. Through a number of iterations, we designed an evaluation sheet with 17 factors categorised within quality of test, results, and report. When the list and description of factors was complete, both authors marked each of the 36 reports independently of each other. For each report, we marked each of the 17 factors on a scale of 1 (worst) to 5 (best). The overall grade was based on a scale of 1 (worst) to 10 (best). Then we compared all the markings and negotiated each of them in order to reach an agreed evaluation of each usability report. One factor was the number of usability problems that each group found. This factor was treated differently from the others. We went through the reports and noted all problems that were found by the team, first independently and then together. This produced an absolute number of problems found by each team. In this paper, we focus specifically on that factor for the usability reports produced in 2000 and 2001.

3   Results

The results of the study show that some significant changes in the results produced by the student teams in 2000 and 2001 exist.
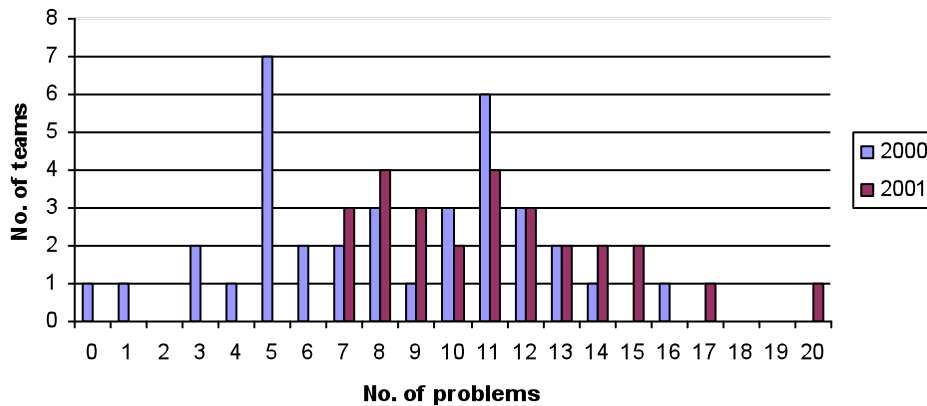

Figure 2. Number of teams identifying specific number of problems in 2000 and 2001

In course given in 2000, the student teams on average were able to identify 7.9 usability problems where the performing team identified 16 unique usability problems and the poorest performing team identified zero problems. In course given in 2001, the student teams on average were able to identify 11.1 usability problems where the best performing team identified 20 usability problems and the poorest performing team identified 7 usability problems. This indicates that overall the student teams in 2001 performed significantly better than the teams in 2000. In addition, the standard deviation of the 2000 teams is 1.68 whereas the standard deviation of the 2001 teams is 1.30 which also indicates that the 2001 results are more homogeneous that the results of 2000. Compared the professional laboratories illustrated in (Skov and Stage, 2001) that on average identified 23.0 usability problems, the teams of 2001 still is far from an ideal situation, but practical training seems to have impact on this issue.

Figure 2 illustrates the distribution of identified usability problems in 2000 and 2001. Here, it is noticeable that more student teams in 2000 performed really poor, e.g. five of the 36 teams identified less than five usability problems at the web-site of hotmail.com. In (Skov and Stage, 2003), we calculate that overall number of different usability problems identified by all 36 student teams in 2000 is 123 usability problems. This implies that these five teams identify less than 5% of the total number of problems. Also, it is remarkable that no student team in 2001 identify less than seven usability problems whereas 12 of the 36 teams (33%) in 2000 identified less than seven problems. This indicates that the training seems to have an impact on the teams that would else wise perform rather poor.

We discovered another problem for student teams in 2000, namely the ability to fully describe the identified problem in a way that would be useful for designers when trying to fix the problem. We denote this issue as the practical relevance of the problem list. This is illustrated in figure 3 where the x axis signify a rating given by the authors of this paper to each produced usability report where 1 is poor, 3 is average, and 5 is very good. In 2000, the scores for the teams are fairly even distributed on the scale. Some of the teams produced problem lists where it would be impossible or very difficult to actually redesign the web-site whereas other teams produced really good problem lists. Most teams (33%) produced a problem list with an average clarity. In 2001, there is slight change since no teams produced a poor problem list, and 14 of the 27 teams (52%) produces a problem list rated good with respect to clarity.
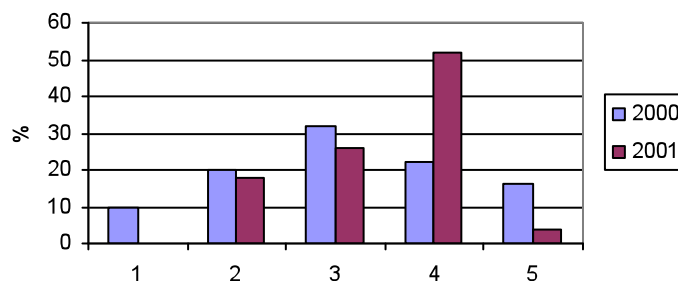

Figure 3. Practical relevance of problem list in 2000 and 2001

4   Conclusion

Universal access is perhaps unachievable if severe usability problems exist in the interface at web-sites. We need to ensure that people involved in web-site development have sufficient skills within

usability testing to be able to design and construct usable web-sites. This paper has described a study on the training of analytical skills of novice usability testers. The study is based on two courses that have taken place at Aalborg University, Denmark in 2000 and 2001 where the curriculum of the course was slightly altered in 2001 to reflect the enhanced focus on usability problem identification.

The study reveals that it is possible to improve the skills of novice usability testers when training their skills in identification and description of usability problems. The student teams in 2001 performed on average better than the teams in 2000 by the number of identified problems and the clarity of the problem list. Additionally, fewer student teams perform really poor in the 2001 study where no team identified less than seven usability problems. This is a remarkable improvement compared to the teams in 2000 where 14 of 36 teams (39%) identified less than seven problems.

References

Braiterman, J., Verhage, S., and Choo, R. (2000). Designing with Users in Internet Time. *interactions*, 7, 5 (September–October), pp. 23-27.

Molich, R. (2000) *User-Friendly Web Design* (in Danish). Teknisk Forlag, Copenhagen

Nielsen, J. (1993) *Usability Engineering*. Morgan Kaufmann Pub., Inc., San Francisco, California

Rubin, J. (1994) *Handbook of Usability Testing – How to Plan, Design, and Conduct Effective Tests*. John Wiley & Sons, Inc., New York

Skov, M. B. and Stage, J. (2001) A Simple Approach to Web-Site Usability Testing. In Proceedings of 1st International Conference on Universal Access in Human-Computer Interaction, Lawrence-Erlbaum, pp. 737 – 741

Skov, M. B. and Stage, J. (2003) Supporting Novices in Detecting Web Site Usability Problems: A Comparison of the Think-Aloud and Questionnaire Methods. In Proceedings of 10th International Conference on Human-Computer Interaction, Lawrence-Erlbaum

Spool, J. M., Scanlon, T., Schroeder, W., Snyder, C., and DeAngelo T. (1999) *Web Site Usability – A Designer's Guide*. Morgan Kaufmann Publishers, Inc., San Francisco, California

Sullivan, T. and Matson, R. (2000). Barriers to Use: Usability and Content Accessibility on the Web's Most Popular Sites. In *Proceedings of Conference on Universal Usability*, November 16-17, Washington, ACM, pp. 139-144.