A Simple Approach to Web-Site Usability Testing

Mikael B. Skov & Jan Stage

Department of Computer Science, Aalborg University, Fredrik Bajers Vej 7, DK-9220 Aalborg East, Denmark

Universal access requires dissemination of usability engineering competence to a wider audience. The multitude of qualifications involved and the strong limitations on development time make classical usability work virtually impossible. This article reports from an empirical study of web-site usability testing conducted by a diverse group of people with no formal education in software development or usability testing. 36 teams of four to eight first-year university students with an interest but no education in information technology were trained in a simple approach to web-site usability testing. The article describes how they applied this approach for planning, conducting, and interpreting a usability evaluation of the same web-site. The student teams gained competence in defining good task assignments and ability to express the problems they found. They were less successful when it came to interpretation and analytical skills. They found quite few problems, and they seemed to lack an understanding of the characteristics that makes a problem list applicable. It is concluded that basic usability testing skills can be developed in a week.

1. INTRODUCTION

Universal access relates to the fundamental relation between use and technology. We are on the move towards a society where the fusion of emerging information technologies will facilitate new or improved patterns of work and interaction. In this way, information technology may impose radical changes to our daily life in a variety of ways. These changes will create new demands that are summarized in the requirement for information technology that is available and accessible to anyone, anywhere, anytime.

The World Wide Web represents a significant move towards the ideal of universal access. Today, the Web makes substantial amounts of information available to anyone with a computer that is connected to the Internet. We are still far from the ideal because of limited distribution of computers and networks, lack of mobility, complexity of network connections, etc. But the Web has definitely brought us a step forward.

The Web has also imposed drawbacks. The present 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, cf. (Sullivan & Matson 2000). Web developers experience a strong push for speed because users of web sites rapidly change preferences and patterns of use, and this makes customers and management ask for development cycles that are considerably shorter than in traditional software development (Anderson 2000, Broadbent & Cara 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 et al. 2000). Moreover, the strong limitation in terms of price and development time effectively prohibits usability testing in the classical sense, conducted by experienced testers in sophisticated laboratories as illustrated in cf. (Fath et.al. 1994, Rohn 1994, Rubin 1994). Even if budgets and deadlines allowed for classical usability testing, the mere amount of web-site development is incompatible with the number of organizations that conduct professional usability testing.

There are a broad variety of suggestions that aim at overcoming the challenges and time pressure that are typical for web-site development. These include iterative testing to spur creativity and informed the design process (Braiterman et al. 2000), user-centered research to improve the designers’ understanding of the prospective users (Anderson 2000), and life stories as a means to surface implicit requirements (Broadbent & Cara 2000). Years of research within usability issues show that usability problems in computerized systems prohibit optimal use and access to the system, its functionality, and its information, cf. (Rubin 1994). Hence, the lack of focus on usability issues and lack of practical skills with usability testing constitutes a potential barrier for universal access of information on the Web, cf. (Sullivan & Matson 2000).
This article reports from an empirical study of the potential for supporting universal access through dissemination of fundamental usability engineering skills. The diversity and extent of Web development calls for systematic usability testing as an indispensable source to enhance universal access. Our aim is to teach a simple approach to usability testing to people with an interest in information technology but without formal education in software development or usability engineering, and to do it in less than a week. Many suggestions for new techniques are analytical activities that may be applied early on in a project, and they include no measures of the time to learn and grasp the technique. The usability testing technique presented here has been examined through an empirical study that was designed to assess the extent to which usability testing can be made accessible to a wider audience of people working with web-site development. 36 teams of five to eight first-year university students with an interest in information technology but no formal education were trained in a simple approach to usability testing. The article describes how they planned, conducted, and interpreted a usability evaluation of a web site. It is described what problems they faced and the quality of their evaluations is assessed.

2. USABILITY TESTING APPROACH

The simple approach to usability testing was developed through a course that was part of a curriculum for the first year at Aalborg University, Denmark. The title of the course was use of information technology, and the overall purpose was to teach and train students in fundamentals of computerized systems with a particular emphasis on usability issues. The course included ten class meetings, each lasting four hours that was divided between two hours of class lectures and two hours of exercises in smaller teams.

The ten class meetings comprised the following topics: #1 introduction and computer networks; #2 usability issues: guidelines and measurement; #3 usability testing: think-aloud protocol; #4 usability testing: questionnaires; #5 computer architecture; #6 usability testing: data analysis; #7 usability issues: techniques and documentation; #8 web-site: usability; #9 web-site: orientation and navigation; and #10 web-site: visual design. All class meetings, except number one and five, addressed aspects of usability and web-sites. The purpose of the exercises was to practice selected techniques from the lectures. In the first four class meetings, the exercises made the students conduct smaller usability pilot tests in order to train and practice their practical skills. The exercises of the last six class meetings were devoted to conducting a more realistic usability test of a web-site.

A number of techniques for usability testing were presented in the course. The first of two primary techniques 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). 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, etc. were also presented to the students.

The tangible product of a usability evaluation was described as a usability report, that identifies usability problems the product, system, or web-site in question. It was suggested that a 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 discussion of methodology (1 page). It was also emphasized that the problems identified should be categorized, at least in terms of major and minor usability problems. In addition, a report should include all data material collected such as log-files, tasks for test subjects, questionnaires etc. A prototypical example of a usability report was given to the students for inspiration.

3. EMPIRICAL STUDY

We made an empirical study of the usability approach that was taught to the students. This study involved 36 teams of first year university students who used the approach to conduct a usability evaluation of the email services at the Hotmail web-site. The 36 teams consisted of students from such diverse educations as architecture and design, informatics, planning and environment and chartered surveyor. These studies are all part of a natural science or engineering program at Aalborg University. Figure 1 describes the 36 teams that participated in the study.

<table>
<thead>
<tr>
<th>Team size</th>
<th>Team size</th>
<th>No. of test subjects</th>
<th>No. of test subjects</th>
<th>Age of test subjects</th>
<th>Age of test subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>average</td>
<td>min / max</td>
<td>average</td>
<td>min / max</td>
<td>Average</td>
<td>min / max</td>
</tr>
</tbody>
</table>
Each student team was required to apply at least one of the two primary techniques, and they were allowed to supplement this with other techniques according to their own choice. With both techniques, each team should among themselves choose a test monitor and a number of loggers (they were recommended to include two loggers), who should examine the system, design task assignments for the test subjects, and prepare the test. The rest of each team acted as test subjects, and the web-site used for testing was kept secret to them. During the usability test sessions, the test monitor should instruct the test subject in matters related to the test whereas the loggers should write down comments and actions from the test subject as well as task completion times.

Each team was given a very specific two-page scenario stating that they should conduct a usability test of the Hotmail web-site (www.hotmail.com). The scenario included a list of features that emphasized the parts of Hotmail they were supposed to test. Each usability test session was planned to last approximately one hour. Due to the pedagogical approach of the university, each team has their own office. Most teams conducted the tests in this office, which 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 the usability report.

The usability reports were the primary source of data for our empirical study. All reports were analyzed, evaluated, and marked by both authors of this paper. First, they worked individually and marked each report in terms of 17 different factors. Second, all reports and evaluations were compared and a final evaluation on each factor was negotiated. In this article, we focus on the following five factors: #1 the planning and conduction of the evaluation, #2 the quality of the task assignments, #3 the number and relevance of the usability problems identified, #4 the clarity and quality of the problems listed in the report, and #5 the practical relevance of these problems.

The specific conditions of this study limit its validity in a number of ways. First, the environment in which the tests were conducted was in many cases not optimal for a usability test session. In some cases, the students were faced with slow Internet access that influenced the results. Second, motivation and stress factors could prove important in this study. None of the teams volunteered for the course (and the study) and none of them received any payment or other kind of compensation; all teams participated in the course because it was a mandatory part of their curriculum. This implies that students did not have the same kinds of incentives for conducting the usability test sessions as people in a professional usability laboratory. Finally, the demographics of the test subjects are not varied with respect to age and education. Most test subjects a female or a male of approximately 21 years of age with approximately the same school background and recently started on a design-oriented education. The main difference is the different curricula they follow.

4. USABILITY TEST RESULTS

The student reports were all marked on the five factors described above. The markings were made on a scale of 1 to 5, with 5 being the best. The results reveal factors where markings are distributed across the whole scale, reflecting varied competence among the students. The primary results of the study are illustrated in figure 2 and 3 where factors with similar and different distributions are shown.
The first of these factors is the task factor where the distribution of markings is illustrated on the left hand side of figure 2. This factor measures the relevance of the tasks, the number of tasks, and the extent to which they cover the areas specified in the scenario. The student teams cover all five elements of the scale with a few at the top and at the bottom. The average is 3.3 for all student teams, which is a little above the middle. The professional laboratories score almost the same result, with distribution across the whole scale and an average of 3.5. This is by no means impressive for the professionals; thus the comparable result produced by the students is rather due to a general low quality of the tasks defined by the professional.

The second factor with a broad distribution is the clarity of the problem list, cf. the right hand side of figure 2. This factor measures how well each problem is described, explained, and illustrated and how easy it is to gain an overview of the complete list of problems. Here, the student teams are distributed mainly around the middle of the scale with a few at the top and at the bottom. Their average is 2.9, just below the middle. Compared to the professional laboratories, they are also doing quite good. The professionals are distributed from 2 to 5 with an average of 3.5. Again, this is not impressive for professional laboratories.

The results also reveal factors with a more biased distribution. The first of these is the conduction of tests, which reflects how well the tests were planned, organized, and carried out. The majority of student teams score 4, which indicates well-conducted tests with a couple of problematic characteristics. The average on 3.7 also reflects the general quality of the test processes. The professional laboratories score an average of 4.6 on this factor, and 6 out of 8 score the top mark. This is, as it should be expected because experience will tend to raise this factor.

A key aim in usability testing is to uncover and identify usability problems, and this is the second factor in this study with differences. The student teams were on average able to find 7.9 problems. They find between 1 and 19 problems with half of the teams finding between 6 and 10 problems. Thus the distribution seems to be reasonable, but compared to the professionals there is a clear difference. The average for the professionals is 23.0 problems identified; only one of them scores in the same group as a considerable number of student teams that is between 11 and 15 problems. Only one student team identified a number of problems that is comparable to the professional laboratories.

The third factor that exhibits a difference is the practical relevance of the problem list. The student teams are almost evenly distributed on the five marks of the scale, and their average is 3.2. Yet when we compare these to the professional laboratories, there is a clear difference. The professionals’ score an average of 4.6 where 6 out of 8 laboratories score the top mark. This difference can partly be explained from the experience of the professionals in expressing problems in a way that make them relevant to their customers. Another source may be that the course has focussed too little on discussing the nature of a problem; it has not been treated specifically with examples of relevant and irrelevant problems.

5. CONCLUSION

Significant usability problems in web site interfaces could be a barrier for universal access. Though, low level of skills in usability engineering is likely to prohibit moves towards the ideal of universal access. This article has described a simple approach to usability testing that aims at supporting the ideal of universal access. The fundamental question behind this approach is whether we are able to teach fundamental usability skills to people without any formal education in software development and usability engineering. The answer given above is based
on a large empirical study where 36 student teams have both learned and applied the approach with time spent that amounts to 40 hours or a week’s work.

The strengths reflect that the student teams that learned the approach gained competence in two important areas. First, they were able to define good tasks for the test subjects by addressing most of the relevant features described in the specification for the test. Second, they were able to express the problems they found in a clear and straightforward manner. Overall, these strengths reflect competence in planning and writing.

The weaknesses reflect general limitations with interpretation and analytical skills. The students were less successful when it came to the identification of problems, which is the main purpose of a usability test. Most of the teams found quite few problems, and the number of problems found by the professional laboratories we compared was three times higher. It was also difficult for the students to express the problems found in a manner that would be relevant to a practicing software developer. They seem to lack an understanding of the characteristics that makes a problem list applicable.

The simple approach to usability testing did provide the students with some fundamental skills in usability engineering. Thus approaches like the one presented here will facilitate usability engineering conducted by people with primary occupations and competencies that are far away from software development and usability engineering. We see the approach as a relevant alternative to classical usability testing conducted in a formalized manner in advanced laboratories by highly specialized experts. Thereby it provides a valuable contribution to universal access as emphasized here: “Organizations and individuals stuck in the hierarchies and rigidity of the past will not foster what it takes to be successful in the age of creativity, the age of the user, and the age of the Internet economy”. (Anderson 2000).

REFERENCES