Evaluating the usability of web pages: a case study

M.A.E. Lautenbach, I.S. ter Schegget, A.M. Schoute & C.L.M. Witteman
Psychological Laboratory
Faculty of Social Sciences
Utrecht University
P.O. Box 80.140
3508 TC Utrecht
The Netherlands
Fax: +31-30-2534511
E-mail: C.Witteman@fss.uu.nl

Abstract
An evaluation of the Utrecht University website was carried out with 240 students. New criteria were drawn from the literature and operationalized for the study. These criteria are surveyability and findability. Web pages can be said to satisfy a usability criterion if their efficiency and effectiveness is satisfactory to the user. We operationalized efficiency and effectiveness as surveyability, that is, the users' satisfaction with the legibility and comprehensibility of the pages and findability, the users' ability to find information on the pages or the pages' ease of use, respectively. We conducted a case study in which subjects were observed while they performed a search task on the Internet and then answered questions about findability and surveyability. The surveyability and findability criteria seem to be an effective measure of the usability of web pages and give a reliable indication of the users' judgements of the pages' user-friendliness.

1. Introduction

In this paper we describe a study that we undertook out of puzzlement with the ease of use, or lack thereof, of the Internet. We questioned the usability of many of the web pages we visited. We wanted to find out whether this was a subjective opinion or one shared by others. Therefore, we set up a study to evaluate the usability, or experienced user-friendliness, of web pages, using those of our own University as a case in point. Consulting the literature, we found quite a few articles and books describing models, guidelines and standards for designers to ensure usability of human-computer interfaces. Although web pages are not always specifically mentioned, most of these guidelines are equally applicable to designing usable web sites.

Usability is defined in different terms by different authors. We will adopt the standard translation of usability as the efficiency and effectiveness of use of a system, to the users' satisfaction (ISO DIS 9241-11, 1993). Users will be satisfied with web pages when they are successful in finding the information they are looking for, without taking too many wrong turns and within a reasonable amount of time. This success will to a large extent depend on the layout of the pages. For example: are the links presented in such a manner that it is clear where they are leading, are there not so many links and pictures and so much text on the pages that the user gets confused, etc. In short: are the pages surveyable?

Remarkably, evaluations by users of the usability of web pages, whether or not the pages were designed following usability guidelines, are not often reported, not even in this Journal's 1997 Special Edition on Web Usability. We therefore constructed a method and designed a study to get user evaluations of the usability of web pages. First, in section 2, we will briefly summarize usability and introduce our method, then, in section 3, we will describe our case study.

2. Usability

Usability is a measure of the ease with which a system can be learned and used, its safety, effectiveness and efficiency and the attitude of its users towards it (Preece, 1994). Because users principally know a system by its interface, designers' efforts at improving usability are chiefly directed at improving interfaces. Norman and Draper (1986) have alerted designers to the importance of presenting users, through the interface, with a system image that helps them form a user model that is consistent with the design model and that enables them to understand and use the system. The primary task of a designer is thus to construct a proper system image or interface. For web pages this involves thinking about such aspects as displays, instructions, layout, colors and error-messages.

More specifically, designers are well advised to keep the following in mind (Den Buurman, Leebeek and Lenior, 1985). Confusion and frustration may be avoided by making sure that system reactions to user actions are logical responses, and that the system reacts appropriately when a user's command is not exactly the same as that specified by the designer. Also, system reactions as well as terminology used should be consistent and uniform across situations. For example: 'quit' should always close a

program and should not be used interchangeably with 'stop'. To avoid impatience, the time that elapses between a user's command and the system's response to that command should be as short as possible. If the system cannot respond within a few seconds, it has to make clear, for example by displaying a small hourglass, that it is indeed processing the command. Also, a page that is too difficult to survey may cause discomfort and thereby loss of concentration to the user (Downton, 1993). Because of the well-documented fact that people can only concentrate on a few items at a time, the amount of information on a web page should be limited so that many items do not compete for attention (Eberts, 1994). This is especially important for a homepage or for a page with many links.

It is evident that many aspects of design may add to, or lessen, the usability of a web page. There is no blueprint that only needs to be filled in to yield a usable page. Nielsen (1993, p. 26) states that ".. usability is not a single, one-dimensional property of a user interface". Measuring usability means checking the efficiency and effectiveness of use of the system, as well as the satisfaction of its users (compare Dumas and Redish, 1994).

On the Internet itself, a variety of guidelines for designing usable web pages may be found (to name the major sites: Ameritech, 1998; IBM, 1998; Sun, 1998; Yale, 1998). Design is not the major topic of our paper, so we will just summarize the two major points of those guidelines here. First, the user should always be able to see where she or he is on the Internet. Designers are advised to organize the buttons that are important for moving through the pages of a site in a navigation panel, which should be clear and consistent throughout the site. Our criterion of findability is proposed as a measure of the success designers have had in following this guideline. Second, designers should try to keep the interface as simple as possible, with a well-organized layout. White lines on a page give the user some “breathing room”. Also, one should not use superfluous visual elements, because pictures and frames considerably prolong the time it takes to download the pages, which may tax the user's patience. Further, it is preferable to present information in a familiar or self-evident framework in order to facilitate the user’s processing. This translates into the advice to use images that match reality and to design in correspondence with the user’s model. Put differently, it pays to use images with high affordance. That is, those whose use or function is immediately clear to users (compare Gibson, 1979; Gaver, 1991). We propose the criterion of surveyability as a measure of the design’s success in this respect. We propose that findability, that is, the observed ease of use, and surveyability, that is, user perceptions of satisfactory layout, together constitute usability.

We define surveyability in terms of, among other aspects, the consistency, simplicity, clear layout, effective use of colors and choice of text characters and graphics used in the web pages. We define findability as a measure of the users' ability to find the information they are looking for in a reasonable time. Our method therefore consists of asking users’ opinions about the surveyability of the web pages and measuring the users’ ability to find information on the Internet.

Our two research questions are, first, are the web pages surveyable according to their users? That is, can the users comprehend the information displayed, do they approve

of the quantity of information and the manner of presentation? Second, is information easy to find for users? That is, are users able to find the information they are looking for, and in a reasonable period of time and after following a reasonable number of links?
Shneiderman (1997), whose sustaining dedication to human-computer interaction issues is indicated by the numerous references to his work on the web, states that it will take a decade until we have had sufficient experience, experimentation and hypothesis testing to clarify interface design issues. Until that time, and to substantiate the existing guidelines, every design, including that of web sites, should be subjected to usability testing (Nielsen, 1993, 1995). How to set up such a test depends on the design one is interested in, and is a function of particular users performing particular tasks in a particular environment (Smith, Newman and Parks, 1997). Controlled experimental studies are less time-consuming and may be effective for narrow issues, while field studies, data logging and on-line surveys may be more informative when one has no clear ideas yet and wants to find out the opinions of a group of users who may have widely different backgrounds and interests. Heuristic evaluation, in which reviewers evaluate a system against high-level heuristics such as ‘be consistent’ or ‘prevent errors’, is used to check and improve the usability of interfaces (Nielsen, 1992; Dix, Finley, Abowd & Beale, 1998), but unfortunately this technique can only be properly used by expert evaluators, whereas web sites should also be evaluated by inexperienced users.
At this time, there is still a need for experimental studies. Thus, we conducted an evaluative study, in which users performed a search task on the Internet and after which we asked them for their opinion about the pages they had visited in their search. Our aims were both to evaluate the web pages of Utrecht University and to put our proposed evaluation method to the test.

3. The study

3.1 Methods

Design
Subjects performed three searches for information on the Internet. All information could be found within the Utrecht University website. Subjects were observed by one of the researchers, who noted the time needed to perform a search, the number of links used and whether the desired information was found. After each search, subjects completed a short questionnaire. The first seven questions of the questionnaire measured the users' ability to survey the page; the eighth, together with the researchers' observations, measured the users' ability to find the information.

The task
The subjects were randomly assigned one of four sets of search tasks. Each set consisted of three different searches. An example search set, translated from the Dutch, is given in Table 1.

In the four different sets of search tasks, different links had to be followed and different pages visited. All searches in all sets could be completed using ten or eleven links. Each set of search tasks was assigned to sixty subjects.

1. Find out how many students are currently studying at Utrecht University. 2. Find an overview of the full-time programs offered by Utrecht University. 3. Find out how to subscribe electronically to a computer course at the Academic Computer Center Utrecht.

Table 1. Sample search set

The questionnaire

After each search, subjects answered a short questionnaire with eight questions (see Table 2). The first six questions addressed the different aspects of surveyability of the pages visited during the search, the seventh question asked for an overall mark, from 0 to 10, for the surveyability of the pages visited in the search[1]. The eighth question asked for an overall mark, again from 0 to 10, for the ease with which subjects had been able to locate the information they had been looking for.

Please indicate your opinion by circling the appropriate number on the scale.					
For all pages you visited during your search,	bad ..... perfect				
what do you think of the quantity of information offered?	1	2	3	4	5
what do you think of the layout?	1	2	3	4	5
what do you think of the color combinations?	1	2	3	4	5
what do you think of the fonts?	1	2	3	4	5
what do you think of the background patterns?	1	2	3	4	5
what do you think of the pictures?	1	2	3	4	5
Please give an overall mark (from 0 to 10) for the surveyability of the web pages you visited during your search:					
Please give an overall mark (from 0 to 10) for the findability of the information you had to search for:					

Table 2. Questionnaire for user opinions about surveyability and findability

Subjects

The task was completed by 240 subjects, all students, 134 of whom were male and 106 female. Their ages ranged from 18 to 35, with an average of 21.8 years (SD = 2.8). Subjects were recruited from all fourteen faculties of Utrecht University except Medicine, through a personal request by one of the researchers in a hallway, cafeteria or computer room. Participation was voluntary and unpaid; about 60 % of the students who were invited to participate were willing to do so.

As can be seen in Table 3, the Social Sciences had the highest representation (n = 67), while Theology and Geophysics were represented by the lowest number of subjects (both n = 2).

[1] In the Netherlands it is customary to express judgements on a ten-point scale, for anything from children's schoolwork and student papers to a work of art.

Of the 240 subjects, 33 or 14 % had used the Internet less than twice. Of these inexperienced subjects, significantly more were female (27) than male (6) ($t = 4.55$, $p = .0005$). There were no other significant differences that depended on subject characteristics or on interactions of subject characteristics and scores for surveyability or findability.

	Experience with use of Internet		
Faculties of Utrecht University	Used no more than twice	Used more than twice	Total
Arts	7	17	24
Biology	3	19	22
Chemistry	-	6	6
Geographical Sciences	1	46	47
Geophysics	-	2	2
Law	2	11	13
Maths and Computer Science	-	7	7
Pharmacy	-	9	9
Philosophy	-	16	16
Physics and Astronomy	-	16	16
Social Sciences	15	52	67
Theology	-	2	2
Veterinary Medicine	5	4	9
Total	33	207	240

Table 3. Numbers of inexperienced and experienced subjects per faculty

Procedure

Pilot work showed that about twenty minutes were required to perform four searches. It was therefore decided to assign each subject only three searches so that the task, including answering the questions, could be performed within 15 minutes. The pilot work also showed that measuring the time in seconds was not very revealing, because extraneous factors, such as the computer used or jams on the Internet, varied. Therefore time was measured in minutes.

Subjects were tested in the computer room of their own faculty. Most of the computers were Pentium 200 or faster, linked to the University network by Ethernet. For each subject, one of the researchers restarted a computer with Netscape, displaying the University's homepage. Possible search histories were removed.

Most subjects needed little instruction before they started their task. Some had to be given a short explanation about Netscape and the Internet. All subjects were reassured that we were not interested in their performance, but that we were investigating the quality of the web pages and were interested in their personal judgement, with no correct or incorrect answers.

The assignment was presented on paper. Subjects were allowed to ask any question except questions about which links they should follow. They were not allowed to use any search engine.

Subjects gave their age, gender, faculty and experience with the Internet. They then performed the first search, observed by one of the researchers, and answered the eight questions about this search. The homepage was restarted and the second and third

search progressed in the same way. When subjects did not complete a search within four minutes, which had been found to be ample time, the researcher stopped them and showed them the correct path. Finally the subjects were thanked for their participation and presented with a lollipop.

Data analysis

The mean marks for surveyability per search and the mean overall mark for surveyability were internally reliable (Cronbach's alpha ranging from .86 to .95), as were the mean ratings for the five aspects of surveyability per search and the mean overall rate for surveyability (Cronbach's alpha ranging from .88 to .96). To check whether the different aspects of surveyability were valid indications of the subjects' judgements, the ratings (on the five-point scale) for each of the six aspects (quantity of information, spacing, colors, fonts, background pattern and pictures) were added separately and averaged, and compared to the mean overall mark (on the ten-point scale) for surveyability given in answer to the seventh question. Pearson correlations between the ratings for the individual aspects and the marks given in question seven, were above .30 and significant at $p = .01$, two-tailed, for all aspects except font. Because this latter aspect did not contribute significantly to the average total score, it was dropped in further analyses.

The observed data on findability, time, number of links and success, had to be recoded on a five-point scale before they could be compared to the subjects' marks for findability. Time, in combination with success, was recoded as follows. When a subject had not found the information and had used the maximum of four minutes, her or his time was coded as zero. A code of two was given for subjects who had found the information in four minutes, a code of three for subjects who had found the information in three minutes, a code of four for subjects who had found the information in two minutes and a code of five for subjects who had found the information in one minute. The numbers of links used were, in combination with success, recoded as follows. When a subject had not found the information, a score of zero was assigned irrespective of the number of links followed. When a subject had found the information, a score of one was assigned when she or he had followed nine links or more above the minimum number of links necessary, a score of two for five, seven or eight more links than necessary, a score of three for three or four extra links, a score of four for one or two extra links and a score of five when the information was found with the minimum number of links.

The mean overall score for findability (mean scores for recoded time together with mean scores for recoded number of links) was internally reliable (Cronbach's alpha ranging from .48 to .84, with lower values caused by the complexity of the score). We also established the internal reliability of the average marks given by the subjects for findability for each search and the overall average mark for findability (Cronbach's alpha ranging from .81 to .89). Pearson correlations between the average scores of findability and the average marks given to findability by the subjects were significant (.58, $p = .01$), which means that these recoded scores, or data on the subjects' actual

performance, were valid indications for subjects' judgements about their ability to find information.
A final score for usability of the web pages was calculated by taking the average of the overall score for surveyability and the overall score for findability.

3.2 Results

Surveyability

All five aspects of surveyability that were found to be reliable indications of the subjects' judgements, that is, the quantity of information presented on a web page, spacing of text on the pages, the colors used, background patterns and pictures, were rated between tolerable and fair (between 3 and 4 on the five-point scale) for all searches in all search sets (see Table 4).
The averages of the subjects' ratings on the five aspects together show that the subjects judged all of the web pages encountered during their searches to be sufficiently surveyable. In Table 4 these averages are presented, multiplied by two to allow comparison with the marks for question seven, which were given on a 10-point scale (see Table 5). Averaging over all rates for all searches in all search sets gives an overall rate for surveyability of 6.41.

		Search 1	Search 2	Search 3
Set 1	Quantity of info	3.73 (.73)	3.67 (.88)	3.60 (.85)
	Layout	3.20 (1.07)	3.12 (1.01)	3.32 (.98)
	Color	3.18 (.93)	3.33 (.82)	3.12 (.83)
	Background	3.35 (1.07)	3.25 (.97)	3.43 (.93)
	Pictures	2.77 (.95)	2.77 (.93)	3.00 (.96)
	Overall rates	**6.49 (1.27)**	**5.63 (1.34)**	**5.73 (1.26)**
Set 2	Quantity of info	3.58 (.56)	3.62 (.61)	2.98 (.93)
	Layout	3.23 (1.00)	3.10 (1.00)	3.15 (1.09)
	Color	3.10 (.99)	3.22 (.92)	3.00 (1.04)
	Background	3.38 (1.03)	3.20 (1.04)	3.17 (.98)
	Pictures	2.80 (.95)	2.98 (1.02)	3.08 (1.01)
	Overall rates	**6.44 (1.11)**	**5.55 (1.23)**	**5.43 (1.51)**
Set 3	Quantity of info	3.30 (.85)	3.48 (.75)	3.43 (.85)
	Layout	3.23 (1.01)	3.08 (.93)	3.18 (.93)
	Color	2.75 (.93)	3.22 (.76)	3.02 (.89)
	Background	3.28 (.83)	3.38 (.64)	3.23 (.85)
	Pictures	2.67 (.97)	3.48 (1.00)	2.78 (.96)
	Overall rates	**6.09 (1.02)**	**5.20 (1.00)**	**5.50 (1.14)**
Set 4	Quantity of info	3.38 (.74)	3.23 (.81)	3.57 (.79)
	Layout	3.23 (1.00)	3.27 (1.06)	3.20 (.88)
	Color	3.17 (1.03)	3.12 (.88)	3.30 (.79)
	Background	3.15 (1.01)	3.25 (.79)	3.28 (.90)
	Pictures	2.93 (1.02)	3.22 (1.01)	2.92 (.94)
	Overall rates	**6.35 (1.12)**	**5.65 (1.01)**	**5.67 (1.11)**

Table 4. Mean ratings (and SD) on a five-point scale for each search in each set of searches for the five aspects of surveyability and doubled overall mean ratings (and SD) for surveyability, with n = 60 in each cell

The overall mark given to the surveyability of the web pages in answer to the seventh question after each search, was slightly higher: 6.48 (averaging the marks for all three searches in all four search sets) (see Table 5 for details).

	Search 1	Search 2	Search 3
Set 1	6.67 (1.20)	6.70 (1.27)	6.65 (1.20)
Set 2	6.68 (1.00)	6.75 (1.05)	6.45 (1.37)
Set 3	5.38 (1.29)	6.47 (1.20)	5.90 (1.43)
Set 4	6.70 (1.27)	6.23 (1.39)	6.72 (1.46)

Table 5. Mean marks (and SD) for surveyability on a ten-point scale for each search in each set of searches with n = 60 in each cell

Pearson's correlation coefficient between the overall rate and the overall mark for surveyability was significant (.565, $p = .01$).

Findability

The average time to complete a search, computed over all searches in all four sets for all successfully completed searches was 2.19 minutes (see Table 6 for more details).

	Search 1	Search 2	Search 3
Set 1	1.65 (.88)(n=52)	1.51 (.76)(n=57)	1.80 (.98) (n=49)
Set 2	1.88 (1.01)(n=56)	1.33 (.66)(n=60)	1.74 (.88)(n=39)
Set 3	3.36 (.95)(n=25)	1.30 (.70)(n=60)	1.77 (1.04)(n=43)
Set 4	1.85 (1.02)(n=49)	2.42 (.92)(n=31)	1.56 (.96)(n=54)

Table 6. Average time in minutes (and SD) taken to successfully complete the searches for those subjects (maximum = 60 per cell) who found the information within four minutes

Overall, twenty percent of the searches were not completed successfully within the four minutes allowed. Some searches were completed successfully by all subjects, while one search in particular was experienced as very difficult to complete (35 of the subjects, or 58 % percent of those attempting it, failed).

The mean number of links followed beyond the minimum number necessary to find the information, was 2.5 when subjects were successful and 5.8 when they were not successful (see Table 7 for more details).

		Search 1	Search 2	Search 3
Set 1	Information found	1.90 (2.19)(n=52)	1.04 (2.18)(n=57)	1.55 (2.03)(n=49)
	Information not found	2.00 (3.38)(n=8)	4.00 (2.00)(n=3)	5.27 (4.73)(n=11)
Set 2	Information found	4.11 (4.25)(n=56)	1.23 (2.34)(n=60)	2.08 (2.82)(n=39)
	Information not found	7.75 (3.77)(n=4)	-	7.76 (4.05)(n=21)
Set 3	Information found	5.52 (2.92)(n=25)	0.95 (6.75)(n=60)	2.09 (3.99)(n=43)
	Information not found	6.46 (4.31)(n=35)	-	7.18 (3.50)(n=17)
Set 4	Information found	1.73 (2.64)(n=49)	3.58 (3.64)(n=31)	2.00 (2.95)(n=54)
	Information not found	6.09 (3.02)(n=11)	5.31 (3.24)(n=29)	6.50 (3.51)(n=6)

Table 7. Mean number of extra links (and SD), above the ten or eleven links necessary, used by subjects who did and subjects who did not find the information, respectively

The three measures, time, number of links and success, taken together as described above, gave distinctly different scores for different searches. For example, the first search in the third set of searches scored very low and the second search in this same set very high (see Table 8 for more details). As can be seen in this same Table 8, the marks given by the subjects for findability in answer to the eighth question after each search, were slightly lower than the scores.

		Search 1	Search 2	Search 3
Set 1	Score	7.27 (3.37)	8.52 (2.45)	6.77 (3.64)
	Mark	6.90 (1.74)	7.03 (1.23)	6.67 (1.47)
Set 2	Score	6.92 (2.71)	9.03 (1.65)	5.43 (4.26)
	Mark	6.77 (1.51)	7.07 (1.10)	5.60 (2.32)
Set 3	Score	2.12 (2.72)	9.22 (1.47)	5.88 (4.20)
	Mark	4.12 (1.89)	6.80 (1.54)	5.78 (1.81)
Set 4	Score	6.70 (3.69)	3.52 (3.71)	7.52 (3.48)
	Mark	6.97 (1.63)	5.27 (2.02)	7.07 (1.69)

Table 8. Average computed scores (and SD) and average marks (and SD) given for findability, for all searches in all search sets, with n = 60 in each cell

The overall score for findability over all searches in all search sets was 6.57 (on a ten-point scale). The overall mark for findability, averaging the marks for all three searches in all four sets, was 6.34.
Pearson's correlation, two-tailed, between the subjects' marks and the observation scores for findability was significant (.591, $p = .01$).

Usability
The subjects' marks for surveyability correlated significantly (.781, $p = .01$) with their marks for findability. The subjects' ratings of the five different aspects of surveyability also correlated significantly (.212, $p = .01$) with the scores for the three different aspects of findability. This correlation is low, because a comparison is made between different types of data. The surveyability ratings represented opinions, while the findability scores were observed. Comparison is therefore somewhat forced and the correlation low, yet sufficient for a valid association.
The overall averages for surveyability and for findability of information in our study correlated significantly (.510, $p = .01$). This correlation is inflated by the lower correlation between the ratings for surveyability and the scores for findability (compare above). Although these correlations could admittedly be stronger, still they are significant and one could express usability as the average of surveyability and findability. For the usability of the web pages of Utrecht University that were visited during by our subjects that would mean a score of 6.45.

4. Conclusion

A quarter of the subjects in our experiment were Social Sciences students, who are predominantly female and who do not, in general, frequently use computers. This explains why we had significantly more inexperienced female than male subjects. But because we found no significant correlation between experience and any of the scores or between gender and the scores, this phenomenon does not decrease the generalizability of our results. It may be that the restriction that no search engines could be used neutralized the factor of experience.

The overall score for the different aspects of surveyability was 6.41 and the overall mark for surveyability was 6.48. Because these two measures of surveyability were highly correlated, it seems safe to conclude that they give a valid indication of the surveyability of the pages of a web site.

The overall score for the different aspects of findability was 6.57 and the overall mark for findability was 6.34. These two measures of findability were highly correlated, therefore we may conclude that they are valid measures for the findability of information on web sites.

Finally, since the correlations between the scores for surveyability and for findability were high, we conclude that these two aspects together may give a valid measure for the usability of web sites. In the case of the web sites of Utrecht University, this usability measure is 6.45.

5. Discussion

Our study gave us two types of results. The first concerns the research questions we had about the usability of the web pages of Utrecht University and will only be of interest to the designers of these pages. In summary, we found the web pages to be surveyable and the findability of information on the pages to be sufficient, but neither measure was very high. On a ten-point scale, at least as it is used and understood in the Netherlands, a score of 6 means barely sufficient. Only scores of 8 or more reflect good performance. The most often recurring remarks of the subjects were that the web pages were boring, that hey contained too much information and that is was often unclear where the links pointed.

The second type of result concerns our method. Setting subjects a search task, observing them while they performed this task and asking their judgements about aspects of findability and surveyability afterwards gives a reliable indication of the usability of the web pages. Application of the method was time-consuming in our case, because we wanted enough subjects to allow statistical analyses. Because we employed an untried method, we needed to establish its worth. Having done so, further application may be done with fewer subjects. Furthermore, the high correlations we found in our study suggest another possible simplification. Because the observed measures (time, number of links and success) and the subjects' opinions are so strongly correlated, researchers could decide to measure only one of these two. One could choose only to observe subjects, at the loss, admittedly, of more specific data about which aspect of surveyability causes the low or high findability of information. Alternatively, one could choose to only present users with a

questionnaire, but then one would have to do without the more objective measure of findability.
Evaluation is always more time-consuming than designers would wish, and acting upon the results even more so. However, one cannot do without if one accepts the legitimacy of users' desires for usable web pages, or indeed wishes to ascertain the benefit of following usability guidelines while designing web pages.
We have presented our conclusions to the persons responsible for the sites of Utrecht University. We strongly hope that a study as reported here may be replicated next year, with much more favorable results.

References

Ameritech (1998). *Ameritech web page user interface standards and design guidelines*. http://www.ameritech.com:1080/corporate/standard/web_guidelines/

Den Buurman, R., Leebeek, H. J. & Lenior, T. M. J. (1985). *Beeldscherm ergonomie [Visual display Ergonomics]* Amsterdam: Nederlandse Vereniging voor Ergonomie [Dutch Ergonomics Society].

Dix, A.J., Finley, J.E., Abowd, G.D. & Beale, R. (1998). *Human-Computer Interaction (2nd edition)* New York: Prentice Hall.

Downton, A. (Ed.) (1993). *Engineering the human-computer interface*. London: McGraw-Hill.

Dumas, J.S. & Redish, J.C. (1994). *A practical guide to usability testing*. Norwood: Ablex Publishing Corporation.

Eberts, R.E. (1994). *User Interface Design.* Englewlood Cliffs, NJ: Prentice-Hall.

Gaver, W. W. (1991). Technology affordances. In *Proceedings of CHI,* ACM, New York, pp. 79-84.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.

IBM Web Guidelines (1998). http://www.ibm.com/IBM/HCI/guidelines/web.

ISO DIS 9241-11 (1993). *Ergonomic requirements for office work with visual display terminals. Part II: Guidance on specifying and measuring usability*. International Organization for Standardization.

Nielsen, J. (1992). Finding usability problems through heuristic evaluation. In P. Bauersfield, J. Bennett & G. Lynch, Eds. *Human factors in computing systems CHI'92 Conference Proceedings* pp. 373-380. New York: ACM Press.

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

Nielsen, J. (1995). *Multimedia and hypertext: the Internet and beyond*. London: Academic Press.

Norman, D.A. & Draper, S. (Eds.) (1986). *User-centred system design.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Preece, J. (1994). *Human-Computer Interaction.* Harlow: Addison-Wesley.

Shneiderman, B. (1997). Designing information-abundant websites: issues and recommendations. *International Journal of Human-Computer studies*, **1**, 5-29.

Smith, P.A., Newman, I.A., & Parks, L.M. (1997). Virtual hierarchies and virtual networks: some lessons from hypermedia usability applied to the World Wide Web. *International Journal Human-Computer Studies*, **1**, 67-96.

Sun (1998). *Sun on the net: guide to web style*. http://www.sun.com/styleguide/

Yale (1998). *Yale C/aim web style guide*. http://mirrored.ukoln.ac.uk/web-authoring/caim/caim/