User Participation Evaluation of A Scholarly Information Site

학술정보사이트의 이용자 참여형 평가

Minsoo Park(박민수)*

ABSTRACT

The purpose of this study was to develop a methodology of user participation evaluation of a scholarly information site in the field of science and technology and to enhance the site by applying a set of testing protocols. Experiments were conducted in a laboratory setting. Data from multiple sources, including eyetracking, search logs and post surveys, were collected and analyzed quantitatively. Based on the results of eyetracking, the contents and images were reorganized after removing unessential elements in the site. The resulting data from the search logs showed that the users were able to finish the tasks more quickly with the revised user interface. The results of the data analysis of post surveys indicated an overall improvement in the revised website compared to the original one.

초 록

본 연구의 목적은 과학기술분야 학술정보사이트의 이용자 참여형 평가 방법론을 개발 및 적용시킴으로써 평가사이트를 개선시키는 데 있다. 실험은 통제된 환경에서 이루어졌다. 아이트랙킹, 검색로그, 설문지를 통해 수집한 데이터는 정량적으로 분석하였다. 아이트랙킹 데이터 분석 결과를 토대로, 평가사이트상에서 불필요한 항목을 제거하였고 콘텐츠와 이미지들을 재배열하였다. 검색로그 데이터는 시스템의 효율성이 증가하였음을 보여주고 있다. 설문지 데이터 분석 결과 또한 시스템이 전반적으로 개선하였음을 보여준다.

Keywords: user participation evaluation, science and technology, user study, methodology, scholarly information site
이용자 참여형 평가, 과학기술, 이용자연구, 방법론, 학술정보사이트

* Senior Researcher, Knowledge Information Center, Korea Institute of Science and Technology Information(mspark7@gmail.com)

■ 논문접수일자 : 2011년 11월 13일 ■ 최초심사일자 : 2011년 11월 21일 ■ 게재확정일자 : 2011년 12월 6일
■ 정보관리학회지, 28(4): 85-97, 2011. 〔http://dx.doi.org/10.3743/KOSIM.2011.28.4.085〕

1. Introduction

In the paradigm of user-oriented designs, the information systems are designed with a focus on the optimization to the users who will actually use them. It is possible to obtain an insight on the series of actions within this process by observing the interactions between the users and the systems, employing surveys or interviews, or going over through search logs, as part of a number of different research approaches for this purpose. In this process, the possible issues in need of improvement or problems are identified in terms of the system's functions or designs, which will duly reflected in the development process to result in actual enhancement of the system.

Introduced in 2008, the National Discovery for Science Leaders (NDSL) is a national, integrated, one-stop information service for science and technology. It is being run by the Korea Institute of Science and Technology Information (KISTI), which is the national scientific and technical information center. The center provides various types of content and media in science and technology such as journal articles, patents, reports, trends, standards, and factual data.

Since its opening, however, the attempts of studies that evaluate the information system of science and technology from a user's perspective have been scarce. Successful design of an information system comes from understanding the people who interact with and use the system. The purposes of this study are to develop a methodology of user participation evaluation and enhance the information system of science and technology by applying the user testing protocol.

2. Related Studies

To design an information system that supports the users as they perform various tasks, it is necessary to understand the users and their needs (Monk et al. 1993). One way to gain perspectives on the user needs is to evaluate an information system from the user's vantage point. User-oriented evaluation focuses on learning how users interact with an information system and how information systems and user interfaces need to be designed to facilitate task performance.

There are varying definitions of usability in the human computer interaction/user interface literature. Lowgren (1993) defined usability as “the results of relevance, efficiency, attitude and learnability.” Preece et al. (1994) stated, “usability, a key concept in HCI, is concerned with making systems easy to learn and easy to use.” One area of human-computer interaction is user interface between a user and a system. Many theories and models in this area come from the fields of computer science, cognitive science, human factors, and others.

In the area of human-computer interaction can be found numerous heuristics (Nielsen 1993; Nielsen 1994; Shneiderman & Plaisant 2009). Complete with a set of predefined criteria, this type of evaluation method focuses on identifying problems by

observing and analyzing interactions between users and information systems (Agerfalk et al. 2002; Cronholm & Goldhuhl 2002). Items in predefined criteria derive from specific perspectives necessary for evaluation (e.g. information, service interaction, system error, etc.). Throughout the evaluation, strengths and weaknesses are identified and analyzed in terms of cause and effect, focusing on perceived system functionality (Agerfalk et al. 2002; Beyer & Holtzblatt 1998) and Web quality (Barnes & Vidgen 2002).

WebQual (Barnes & Vidgen 2002) is a method for assessing site quality and consists of three areas: usability, information quality, and service interaction quality. Although WebQual is grounded in subjective assessment from website users, the data produce useful quantitative analyses and results.

Another way to evaluate websites is eyetracking, a process of measuring the motion of an eye or the flow of attention on an object. Eye movement reflects human thought processes so a user's thoughts may be followed from records of such eye movement (Yarbus 1967). It is easy to determine from these records which elements attract the user's eye, in what order, and how often (Yarbus 1967).

Eyetracking studies demonstrate that long pages with a large quantity of information are problematic because of cognitive limitations. People prefer sites that get to the point and let them get things done quickly and easily. These studies suggest that it is important to keep primary information within a page's initially viewable area, without further actions necessary to access it (Nielsen & Pernice 2009; Nielsen 2010).

For this study, the researcher adopts a user participation evaluation method to learn the ways real users interact with a scholarly information site and how they perceive it. Combined with a standard protocol and observations, problems and their causes are identified and examined.

3. Research Design

The current study was designed with a view to user participation evaluation of a scholarly information site, partially referring to ISOTR16982 (2002). For this purpose, twelve subjects were selected for participation in the test. Using eyetracking, search logs, and surveys, the website was tested by six subjects. Employing search logs and surveys, the revised website was then assessed using the other six subjects. The data from the experiments were analyzed quantitatively.

3.1 Subject Sample

The user sample consisted of twelve heavy users of the site, selected in consideration of diverse educational and occupational backgrounds. For the website in question, twelve heavy users were selected as participants based on their past experience of using the site. No demographic limits were set to minimize bias in the sampling process.

Typically, an assessment of a system like what we have in the current study does not require an

excessively large participant groups. Nielsen(1999) demonstrated that it is possible to address 80% of the system issues with only five participants. Also, Krug's study (2000) showed that they could discover more system vulnerabilities when conducting the tests twice with a different group of three participants than when they did so only once with eight participants.

3.2 The Tasks

Four information searching tasks were given to the users:

- Task 1: Find reliable information on green technology. Provide sources and full text, if available. (5 items maximum)
- Task 2: Find information to understand research trends on green technology in different time periods. Provide sources and full text, if available. (5 items maximum)
- Task 3: Find images of green technology. Submit images with sources. (5 items maximum)
- Task 4: Find information on your topic of interest. Submit sources and full text, if available. (5 items maximum)

Based on the data of log analysis from NDSL in 2009, the topic of the tasks was green technology. The results of log analysis demonstrated that green technology was a popular area in science and technology in 2009. The tasks focused on information reliability, time, type, and interest. The first task relates to information reliability and the second focused on temporal aspects of information. The third relates to information type and the fourth was created by the subjects, reflecting individual areas of interest.

The tasks described above were not direct and answer driven. Each task was designed to be generic to engage the subjects in cognitive and physical actions to solve the problems and further accomplish the goals the subjects might have.

3.3 Data Collection and Analysis

The scholarly information site was tested by employing multiple research techniques, including eyetracking technology, search logs, and post-task surveys. Eyetracking technology was applied only to the former site user interface, before revising the site. Data of search logs and surveys were collected both before and after making changes to the site. Eyetracking was used to test the overall design of the site. In recent years, advancement of eyetracking technologies produced a great deal of interest in the research field of web usability. Eyetracking data provide useful insights into features that are most attractive and unnoticed.

In addition to eyetracking, each user's search activities while working on the four tasks were recorded and analyzed using Camtasia Studio. Log data were analyzed in terms of two aspects: (1) how successfully users found information and accomplished the tasks and (2) how fast users completed the tasks.

The survey used a scale of 1 to 10 to rate the replies. An evaluation list of the scholarly information site was developed by partially referring to Nielsen (1994, 1999), Barnes and Vidgen (2002), and DeLone and Mclean (1992). The list was not limited to usability; it covered evaluation items of information quality and service quality, including seven assessment dimensions: information, visualization, error, learnability, speed/access, service interaction, and personalization. Each dimension consisted of seventeen sub-evaluation categories and sixty-five evaluation items.

3.3.1 The Experiments

The study was conducted in a laboratory setting where the researcher conducted experiments with identical equipments, procedures, and standardized instruments to collect data from diverse sources. The procedure was as follows: First, six users participated in the experiment sessions of eyetracking to test the user interface of the former site. Mountable systems including MA06-monocular and BA06-minocular manufactured by Arrington Research were used to measure eye movements and eye positions. The primary purposes of using an eyetracking technique in this study were to ascertain:

① Which parts of the website attract a user's eyes (attention)?
② Which elements of the website are unnecessary?
③ Which parts of the website draw a user's attention, in what order, and how often?
④ Is the layout of the website appropriate?

The testing procedure of eyetracking was as follows: At the beginning of the testing process, each subject was interviewed to discover whether he/she wore eyeglasses. Once subjects felt comfortable wearing the equipment, they started performing tasks on the site. Data from the testing sessions were collected and the eye flow and fixations shown in the data were analyzed to evaluate the usability of the website. Each subject was asked to conduct four information searching tasks (three assigned and one non-assigned) using a PC with no time limit. All activities on the PC were logged using Camtasia Studio and analyzed later. After completing all information searching tasks, an evaluation form was provided to the users, who filled in and submitted for later analysis by the researcher. Data to a quantitative analysis included eyetracking, search logs, and surveys, which were tabulated, compiled, and analyzed using Excel software. The following section discusses the results of data analyses.

4. Results

4.1 Subject Information

Twelve users from diverse backgrounds participated in the study (See Table 1). Most of them majored in science and technology. Two studied Economics. One majored in Business. Eight were senior researchers. The subjects' ages ranged from twenty-four to forty-two.

〈Table 1〉 Subject Information

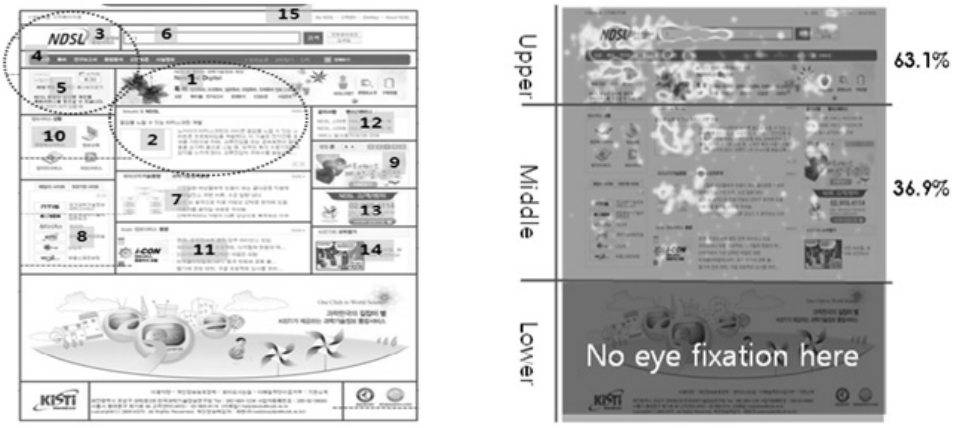
Subj	Age	Major	Occupation	Content Used
B1	40	Marine Engineering	Researcher	Trends
B2	35	Electronics	Researcher	Journals
B3	39	Food Science	Researcher	Patents
B4	24	Management Tech	Grad Student	Trends
B5	42	Computer Science	Patent Expert	Patents
B6	33	Economics	Info Consultant	Journals
A1	39	Marin Engineering	Researcher	Trends
A2	31	Business	Info Consultant	Journals
A3	35	Electronics	Researcher	Reports
A4	29	Computer Science	Researcher	Journals
A5	32	Computer Science	Researcher	Journals
A6	34	Economics	Researcher	Journals

4.2 Eyetracking

The main page and search results page of the website were tested using eyetracking. Data were collected and analyzed for the eye flows and fixations.

4.2.1 Analysis of Eye Fixations on Main page

As shown in Figure 1 on the left, the middle



〈Figure 1〉 Main Page - Flow (Left) and Percentage (Right) of Eye Fixations

1. Quick Menus
2. Issues in Science & Technology
3. Logo
4. Global Navigation Bar (GNB)
5. Login
6. Search Box
7. Trends in Science and Technology
8. Links to External Sites
9. Rolling Banner
10. Request for Service
11. Trends in Information Services
12. Notices
13. User Support (Help Desk)
14. Scent of Science
15. Header

section including the quick menus and issues in science and technology of the main page first attracted subjects' eyes; this area was followed by the upper section including the logo, GNB, and login areas. Users tended to move attention quickly in the areas containing images and Flash animations. It is interesting to note that user attention focused on the middle first but was held in the upper area longer. The image on the right in Figure 1 shows that the upper section of the main page drew subjects' attention most (63.1%), followed by the middle section (36.9%). The lower part of the page did not attract subjects' eyes at all.

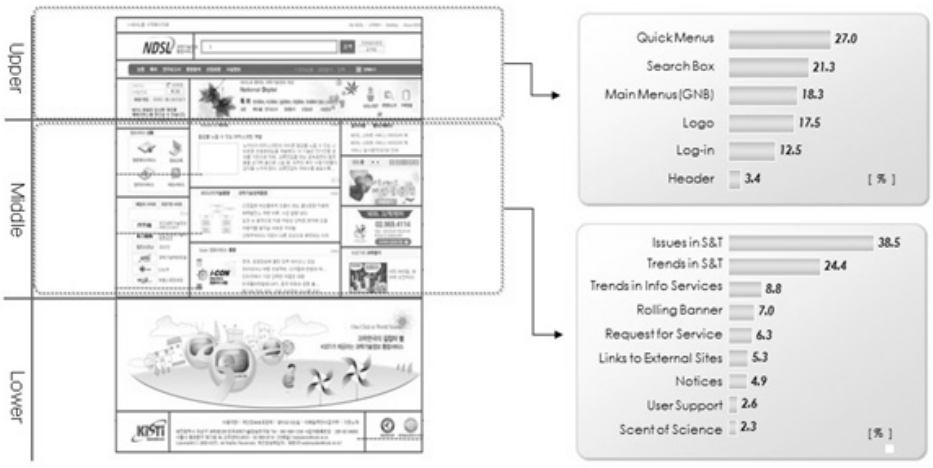
4.2.2 Further Analysis of Eye Fixations on Main Page

In the upper area of the homepage shown in Figure 2, users focused vision primarily on quick menus (27%), the search box (21.3%), main menus (18.3%), the logo (17.5%), the login (12.5%), and the header (3.4%). In the middle area of the page, eye fixations occurred mostly in issues in science and technology (38.5%), trends in science and technology (24.4%), global trends in information services (8.8%), rolling banner (7.0%), request for service (6.3%), links to external sites (5.3%), notices (4.9%), user support (2.6%), and Scent of Science (2.3%).

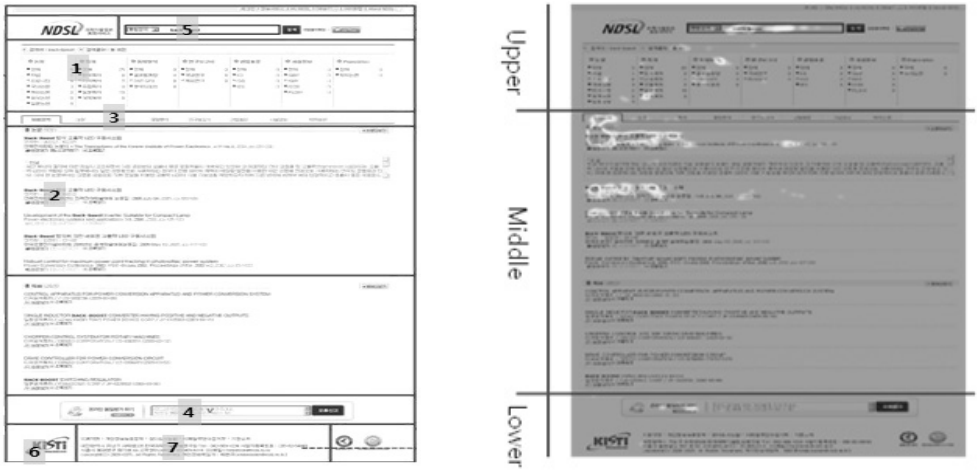
The lower section containing a promotional image, logo, and footer was ignored.

4.2.3 Analysis of Eye Fixations on Search Results Page

As shown in the left image of Figure 3, it was noticed that subject attention traveled first to the top display area, which provided search results by content categories such as journals, patents, trends analyses/monitoring, reports, standards, and factual data. The eye fixations then moved quickly to the middle area of the search results. In the



〈Figure 2〉 Main Page - Further Analysis of Eye Fixations



〈Figure 3〉 Search Results Page - Flow(Left) and Rate(Right) of Eye Fixations

1. Categorical display of search results
2. Search results
3. Tab display of search results by categories of contents
4. Error report
5. Search box
6. Logo
7. Footer

image on the right half of Figure 3, major eye movements occurred in the middle (search results, 67.3%) and upper (search box, 25.9%) areas of the page. Participants tended to check the first search result and did not browse through the results to the end of the page. Top search results, not all of them, drew user attention.

4.3 Task Performances

Task performance was measured in two aspects, which were: (1) successful completion rates (effectiveness) and (2) time on task (efficiency). There was no difference in the effectiveness of the site in terms of successful completion rates; users finished all tasks successfully with the former and revised user interfaces.

This section discusses the results of task performance for the temporal aspect. Results of the data analysis show that users spent different amounts of time on the tasks. The average time duration of a task was different depending on the user interface, former versus revised. Subjects finished the tasks more quickly under the new user interface.

As shown in Table 2, subjects spent less time on average to complete each task on the new website: 950 seconds for the first task, 920 seconds for the second, 730 seconds for the third, and 620 seconds for the last; it took a total of 3,220 seconds on average for the participants to complete all tasks using the new user interface. Subject spent the shortest time (620 seconds) on the last task (topic of interest) and the longest (950 seconds) on the

〈Table 2〉 Time Spent on Tasks (sec.): Comparison of Former and New Site

	Task 1		Task 2		Task 3		Task 4		Avg. Time	
Subj #	B	A	B	A	B	A	B	A	B	A
B1/ A1	1151	840	832	900	518	720	752	480	813	735
B2/ A2	1312	600	744	780	1184	1320	835	780	1019	870
B3/ A3	1610	900	1319	1020	1046	660	1058	540	1258	780
B4/ A4	1084	1380	1108	900	833	840	600	600	906	930
B5/ A5	1020	1380	1555	1260	1147	540	646	540	1092	930
B6/ A6	1081	600	2858	660	776	300	481	780	1299	585
Avg. Time	1210	950	1403	920	917	730	729	620		

Note: B and A stand for before and after, respectively

first task (reliable information). It appears that there may be a correlation between task performance and interest level.

There was no time limit to perform each task. Instead, the subjects were observed to see how quickly they finished the tasks with the system. Even though the subjects were asked to find 5 items maximum for each task, not all of them tried to find exactly 5 items for each task. Interestingly, they did stop performing tasks, when they thought they found enough.

4.4 Surveys

The scholarly information site was evaluated for seven aspects: (1) Information, (2) Visualization, (3) Error, (4) Learnability, (5) Speed/Access, (6) Service Interaction, and (7) Personalization. Primary categories of evaluation were composed of seventeen sub evaluation categories and sixty-five evaluation items. The first category of evaluation was Information, which included Recency, Clarity, and Accuracy/Relevancy. The second category of evaluation was related to visual design such as Navigation, Menu, Layout, Color, etc. The third and fourth groups of assessment were User Error and Learnability, which included Error Prevention and User Awareness. The fifth category was about Speed/Access, which covered evaluation items of Efficiency and Security. The sixth and seventh groups were Service Interaction and Personalization. Service Interaction consisted of two sub evaluation categories, Feedback and Help/User Support.

As shown in Table 3, total average testing scores were 6.8 out of 10 for the former site and 7.7 out of 10 for the new one. The testing scores of main categories are described further in Table 4.

For the revised site, the average scores of the main categories were increased generally from those with the original site, except for Speed/Access. Evaluation items related to Speed did not get high scores (e.g. “the site can be accessed fast” and “the webpages are loading quickly”). The reason for this may be because testing was conducted immediately after revising the site (i.e. before the system became stable).

〈Table 3〉 Evaluation Average Scores

Evaluation Categories		Evaluation Sub-categories	Scores (out of 10)	
			Before	After
Information	1	Recency	7.1	8.3
	2	Clarity	6.4	7.3
	3	Accuracy/Relevancy	7.6	8.4
Visualization	4	Navigation	6.6	7.3
	5	User Control	5.8	6.9
	6	Menu	6.7	8.1
	7	Labeling/Image	6.8	7.9
	8	Layout	6.8	8.1
	9	Icon	6.5	8.3
	10	Typography	7.3	8.0
	11	Color	7.3	8.3
Error	12	Error Prevention	6.0	7.8
Learnability	13	User Awareness	6.4	7.9
Speed/Access	14	Efficiency/Security	7.5	7.2
Service Interaction	15	Feedback	7.3	6.5
	16	Help/User Support	6.4	7.5
Personalization	17	Personalization	7.1	7.8
Total Average Scores			6.8	7.7

〈Table 4〉 Average Testing Scores of Main Categories

Evaluation Categories	Scores (out of 10)		
	Previous	Revised	+/-
Information	7.0	8.0	+1.0
Visualization	6.7	7.9	+1.2
Error	6.0	7.8	+1.8
Learnability	6.4	7.9	+1.5
Speed / Access	7.5	7.2	-0.3
Service Interaction	6.9	7.0	+0.1
Personalization	7.1	7.8	+0.7

5. Conclusion and Discussion

This study had two goals: develop a methodology of user participation evaluation and enhance a scholarly information site by applying the user testing protocol. To evaluate the website, data from eyetracking, search logs, and surveys were collected and analyzed quantitatively. The eyetracking

technique used in this study made it possible to diagnose which elements drew the user's attention most and least, and in what order. Based on the results of eyetracking, unnecessary elements of the site were removed (e.g. a promotional image); content and images were also reorganized and displayed (see Figure 4).

Task performance was measured in terms of successful completion rates and time on task. The users successfully accomplished the tasks both on the former site and the new one. With the revised site, users finished tasks more quickly. A survey was developed and conducted before and after the renewal process and data from it suggest that the new website was improved generally, except the speed of the system, which might be attributable to the evaluation being conducted before the system was stabilized.

User-oriented designs for information systems aim to optimize the systems around how users interact with the system; users should not be required to adapt behaviors while using systems. User-centered approaches lead to usable websites and result in high satisfaction with information systems.

User participation evaluation is an essential process in designing information systems. It is possible to discover from user evaluations which parts of the system are in need of improvement and which are satisfactory from a user's perspective. This study is part of a broader research initiative to design a user-oriented scholarly information site in the field of science and technology by developing evaluation instruments and applying advanced technologies. Understanding user experiences and behaviors is fundamental to assessing and improving current information-related services and to plan and develop new ones.

The researcher is currently conducting research with a larger user sample group to gain deep insights on user behaviors on the Web, employing think-

〈Figure 4〉 The Original Main Page(Left) and Revised Main Page(Right)

aloud technique. Think-aloud data are a collection of verbal expressions of what was going on in one's mind as he/she performs a given task. If one say the quantitative analysis of search log data covers an analysis on actions, the qualitative analysis of the think-aloud data can help understand the reasons behind such actions. By a qualitative analysis of this think-aloud data, it is expected to provide different sets of results from those available in quantitative analysis. However, this is not to say the quantitative method is vulnerable. Rather, it signifies that an even more comprehensive, yet deeper insight can be achieved when these two distinguished approaches are joined together.

References

Agerfalk, P.J., J. Sjostrom, E. Eliasson, S. Cronholm and G. Goldkuhl. 2002. "Setting the Scene for Actability Evaluation: Understanding Information Systems in Context." *Proceedings of ECITE Conference*, 2002: 15-16.

Barnes, S.J. and R.T. Vidgen. 2002. "An interactive approach to the assessment of e-commerce Quality." *Journal of Electronic Commerce Research*, 3(3): 114-127.

Beyer, H. and K. Holtzblatt. 1998. *Contextual Design: Defining Customer-Centered Systems*. San Diego: Academic Press.

Cronholm, S. and G. Goldhuhl. 2002. "Actable Information Systems-Quality Ideals Put into Practice." *Proceedings of the Eleventh Conference on Information Systems IST*, 2002: 12-14.

DeLone, W. H. and R.E. Mclean. 1992. "Information Systems Success: The Quest for the Dependent Variable." *Information Systems Research*, 3(1): 60-95.

The International Organization for Standardization. 2002. *ISO/TR 16982: Ergonomics of Human-System Interaction-Usability Methods Supporting Human-Centred Design*. 1st ed. Geneva, Switzerland.

Krug, S. 2000. *Don't Make Me Think: A Common Sense Approach to Web Usability*. Indianapolis: New Riders.

Lowgren, J. 1993. *Human-Computer Interaction - What Every System Developer Should Know*. Lund, Sweden, Studentlitteratur.

Monk, A., P. Wright, J. Haber and L. Davenport. 1993. *Improving Your Human-Computer Interface: A Practical Technique*. New York: Prentice Hall.

Nielsen, J. 1993. *Usability Engineering*. San Diego: Academic Press.

Nielsen, J. 1994. "How to Conduct A Heuristic Evaluation." *Jakob Nielsen's Alertbox*. <www.useit.com/papers/heuristic/heuristic_evaluation.html> (12 October 2011).

Nielsen, J. 1999. *Designing Web Usability: The Practice*

of Simplicity. Indianapolis: New Riders.

Nielsen, J. 2010. "Scrolling and Attention." *Jakob Nielsen's Alertbox*. <www.useit.com/alertbox/scrolling-attention.html> (12 October 2011).

Nielsen, J. and K. Pernice. 2009. *Eyetracking Web Usability*. Berkeley: New Riders Press.

Preece, J., R. Yvonne, H. Sharp, D. Benyon, S. Holland and T. Carey. 1994. *Human-Computer Interaction*. London: Addison-Wesley.

Shneiderman, B. and C. Plaisant. 2009. *Designing the User Interface: Strategies for Effective User Interface Interaction*. 5th ed. Reading, Mass.: Addison-Wesley.

Yarbus, A. L. 1967. *Eye Movements and Vision*. New York: Plenum Press.