A Longitudinal Study of Usability in Health Care – Does Time Heal?

Jesper KJELDSKOV , Mikael B. SKOV and Jan STAGE
Aalborg University, Department of Computer Science, Denmark
{jesper, dubois, jans}@cs.aau.dk

Abstract. We report from a longitudinal laboratory-based usability evaluation of a health care information system. A usability evaluation was conducted with novice users when an electronic patient record system was being deployed in a large hospital. After the nurses had used the system in their daily work for 15 months, we repeated the evaluation. Our aim was to inquire into the nature of usability problems experienced by novice and expert users, and to see to what extend usability problems of a health care information system may or may not disappear over time, as the nurses get more familiar with it – if time heals poor design. On the basis of our study, we present findings on the usability of the electronic patient system as experienced by the nurses at these two different points in time and discuss implications for evaluating usability in health care.

Keywords. Electronic Patient Records, usability, longitudinal study, experts and novice users

Introduction

Usability evaluations are increasingly applied to assess the quality of interactive software systems. Usability has been defined as consisting of three aspects: efficiency, effectiveness and satisfaction and is often also measured on the basis of identified of usability problems [10] [14] [15]. Most mainstream approaches to usability evaluation involve "prospective users" thinking aloud while using the system [6] [15] [17]. According to mainstream guidelines, there is a considerable difference between involving so-called novice or expert users because these users may have different levels of experience with the system being evaluated. However, the consequence of involving novice or expert users as test subjects when evaluating a system's usability is still being debated (see for example [16]) and several comparative studies are being reported (see for example [3] [8] [18] [16]).

Inspired by Nielsen [13], the purpose of the study reported in this paper is to inquire into nurses' experience of a health care information system over time as they develop system expertise. The key question is how the nurses' experience of the system's usability changes when they transform from being novices to being experts. Do usability problems disappear when users get more familiar with a system? Does time heal poor design? Addressing these overall questions, we report from an experiment comparing the experienced usability of an electronic patient record system when it was introduced into a large hospital to the experienced usability after one year of extensive use. The results of this experiment are presented in detail and discussed as a basis for advising evaluators on selection of test subjects and design of task assignments when preparing a usability evaluation within the health care domain.

1. Evaluating with novice and expert users

The Human-Computer Interaction (HCI) literature generally discusses the importance of using "appropriate test subjects" when carrying out a usability evaluation. Typically, it is pointed out, that it is vital to choose participants that are representative of the intended target user community with respect to parameters such as their demographic profile (sex, age, education, profession etc.), and their level of experience (for example if they are novices or experts) [6] [15] [17]. In relation to the level of user expertise, Nielsen [14] propose that there are (at least) three different dimensions to consider:

1. The user's knowledge about the domain (ignorant versus knowledgeable)
2. The user's experience with computers in general (minimal versus extensive)
3. The user's experience with the system being evaluated (novices versus experts)

In relation to system experience, the discussion of when and why to choose test subjects with high or low level of experience is still ongoing. Some systems are only intended to be used infrequently by first-time users, such as many web-based systems, installation programs, etc, and should thus support novices by being quick and easy to learn. Other systems, such as airline booking systems, advanced industrial control systems, and many systems within the health care domain are designed for more frequent use and for highly experienced users. These may take longer time to learn to use but should, in the long run, support expert users by being highly effective. When evaluating such systems it is often intended to have test subjects that reflect the expected profile of the end users. However, in reality it is often difficult and sometimes not even possible to make such a simplistic differentiation between novice and expert users [14]. In real life, users often don't acquire expert skills in all parts of a system regardless of how much they use it because most systems are often very complex and offer a wide range of features that are not frequently used. Thus even highly experienced users of a system may still be novices in respect to some parts of it. Likewise, novice users of a system may have a high enough level of expertise with, for example, the use domain or computers in general to be able to understand and operate even very complex new systems if they are designed properly. Also, it is commonly known that test subjects may feel under considerable pressure during a usability evaluation because they feel that they are being assessed and not the system [15] [17]. For novice users, this feeling of insecurity may be higher than for experts because they are not familiar with the system, and more efforts may consequently be required for making the test subject feel comfortable with the situation [17]. On the other hand, when testing with experts, some usability problems may not appear because these users have developed workarounds to compensate for poor design. A final issue is access to test subjects. While it is typically not a problem to find novice users, it can sometimes be difficult to gain access to a large number of system experts, especially if the system is still under development or has not yet been deployed in the target organization.

Several experiments have inquired into the difference between novices and experts. In information retrieval, it has been observed that novice users often perform poorly [1]. An empirical study of information retrieval through search in a database compared the performance of novices and experts. Though there were no significant differences in the accuracy with which tasks were solved, the expert users performed significantly faster than the novices [5]. In a usability evaluation of a nursing assessment system, novices experienced severe usability problems that were not experienced by the experts. The novice users could not complete the tasks without going back to the patient for more information, and had difficulties locating where information should be entered into the system. The experts, on the other hand, could complete the tasks and had learned to use the system as a checklist for collecting the necessary information [4].

The empirical studies mentioned above all share the characteristic that experiments with novices and experts are conducted at the same time. Thus these experiments rely on a classification of different people as experts and novices. Such a classification is not without problems [2]. Our aim with the study reported in this paper has been to examine the difference between novice and expert user performance within the health care domain but based on a longitudinal study involving the same users in both evaluations. We have focused on the following research questions:

- RQ1: To what extent is the effectiveness and efficiency of using an EPR system different from novices to experts and is this measure identical for different tasks?
- RQ2: Which usability problems of an EPR system are experienced by novices and by experts: which problems are the same, and is there a difference in the severity of the problems that are experienced by both novices and experts?

The first question reflects two of the fundamental aspects of usability. Although they may seem related, it has been shown empirically, that it is necessary to consider both, as they are not correlated [7]. The next question focuses on the usability problems experienced by novices and experts both in terms of the problems and their severity.

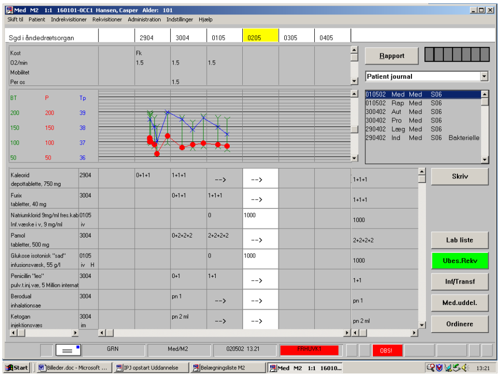
2. Electronic patient record usability

Between 2002 and 2003 we undertook a longitudinal empirical study of novice and expert users' experience of the usability of an electronic patient record (EPR) system for a large regional hospital in Denmark (IBM IPJ 2.3, figure 1). The basic design of the study was to conduct 2 usability evaluations of the same system with the same users. The first evaluation was conducted in May 2002 when the EPR system was being deployed at the hospital. The second evaluation was done in August 2003 when the users had used the system in their daily work for more than a year.

A key part of the system's use domain is the hospital wards. The nurses in each ward and the medical doctors use patient records to access and register information about their patients. They also use it to get an overview of the patients that are in a ward. Through the patient record, they can see the state, diagnosis, treatment, and medication of each individual patient. The nurses use the patient record in three different situations: 1) monitoring how the state of a patient develops, 2) daily treatment of a patient, and 3) emergency situations.

The monitoring typically involves measurement of values, for example blood pressure and temperature. These values are usually measured at the patient's bed and typed in later. The daily treatment of patients can be described as structured problem solving. A nurse will observe a problem with a patient, for example that the temperature is high. She will then make a note about this and propose an action to be taken. This action is subsequently evaluated after some time. All steps are documented in treatment notes. In addition, the patient record provides a basis for coordination between nurses. For example, a nurse coming on duty will look through the list of patients to get an overview of their status and to check the most recent treatment notes to see what treatment has been carried out and what treatment is pending.

Medical doctors and nurses have developed the traditional paper-based patient record as a manual document style over a long period of time. The aim of the electronic record is to computerize that manual document. An electronic patient record is confronted with all the classical problems of creating a database that is shared across a complex organization and designing an interface that is both easy and effective to use. In addition, a hospital has many different groups of employees who may record and interpret data differently. The advantages of electronic patient records are also classical. The primary one is that data will be accessible to all personnel at all times whereas paper-based patient records usually follow the patient physically and is only accessible at one physical location at a time. Electronic patient records also potentially make overall processing of information about large groups of patients much easier.



Sgd i åndedrætsorgan		2904	3004	0105	0205	0305	0405
Kost		Fk					
O2/min		1.5	1.5	1.5			
Mobilitet							
Per os			1.5				

010502	Med	Med	S06					
010502	Rap	Med	S06					
300402	Aut	Med	S06					
300402	Pro	Med	S06					
290402	Læg	Med	S06					
290402	Ind	Med	S06	Bakterielle				

			2904	3004	0105	0205	0305	0405	
Kaleorid depottablette, 750 mg	2904		0+1+1	1+1+1	-->	-->			1+1+1
Furix tabletter, 40 mg	3004			0+1+1	1+1+1	-->			1+1+1
Natriumklorid 9mg/ml fres.kab Inf.væske i v, 9 mg/ml	0105 iv				0	1000			1000
Pamol tabletter, 500 mg	3004			0+2+2+2	2+2+2+2	-->			2+2+2+2
Glukose isotonisk "sad" infusionsvæsk, 55 g/l	0105 iv H				0	1000			1000
Penicillin "leo" pulv.t.inj.væ, 5 Million internat	3004			0+1	1+1	-->			1+1
Berodual inhalationsae	3004			pn 1	-->	-->			pn 1
Ketogan injektionsvæs	3004 im			pn 2 ml	-->	-->			pn 2 ml

Figure 1. The status window of the EPR system

3. Method

The first usability evaluation involved 7 trained nurses from the same hospital. Prior to this evaluation, they had all attended a course on the IPJ system, and they were just starting to use the system in their daily work. All 7 nurses were women, aged between 31 and

54 years, their experience as nurses varied between 2 and 31 years. Before the first evaluation they had received between 14 and 30 hours of training in the EPR system. They characterized themselves as novices in relation to the EPR system and IT in general. The purpose of the second evaluation was to facilitate a longitudinal study of the usability of the system after one year of use. In order to avoid the source of error that originates from individual differences between randomly selected test subjects we used the same 7 participants in both evaluations. Before the second evaluation, all the nurses had used the system in their daily work for about 15 months. They indicated that they on average used the system 10 to 20 times a day, amounting to a total time of use of about 2 hours per day. Therefore, we now characterized them as experts.

In preparation for the evaluations, we visited the hospital and had a number of meetings and discussions with the two nurses who trained the personnel in the EPR system and dealt with the deployment of it. The purpose was to understand the work at the hospital wards related to patient record use and to get an overview of the system. Based on this we made a number of scenarios of the use of the system in collaboration with the nurses who were responsible for the deployment of the system.

The purpose of the usability evaluations was to inquire into the usability of the EPR system for supporting nurses in solving typical work tasks. Based on our scenarios, we designed 7 tasks, including a number of subtasks, centered on the core purpose of the system such as retrieving information about patients, registering information about treatments, making notes, and entering measurements. The tasks were developed in collaboration with the two nurses dealing with the implementation of the EPR system at the hospital. The exact same tasks were used in both evaluations.

The test sessions were based on the "think-aloud" protocol as described by Rubin [16] and Nielsen [13] where the test subjects solve a series of tasks while thinking-out loud, describing their actions, how they perceive the system etc. In both evaluations, the 7 test sessions were conducted over two days. The order of the nurses was random. Each nurse used the system to solve the 7 tasks. This lasted approximately 45 minutes. If a test subject had problems with a task and could not continue on her own, the test monitor provided her with help to find a solution. If a test subject was completely unable to solve a task, the test monitor asked her to go on to the next one. One of the authors acted as test monitor throughout all 14 test sessions. All test sessions were conducted in a dedicated state-of-the-art usability laboratory at Aalborg University, Denmark with a desktop PC setup matching the hardware used at the hospital.

All 14 test sessions were recorded on digital video. The video recording contained the PC screen with a small image of the test subject and test monitor inserted in the corner. The time spent on solving each task was measured from the video recordings. This measure is relevant for addressing RQ1.

The data analysis reported in this paper was conducted in August 2004, one year after the second evaluation. The 2 authors who did not serve as test monitor analysed all 14 videos. Each video was given a code that prevented the evaluator from identifying the year and test subject. The videos were assigned to the evaluators in a random and different order. The evaluators produced two individual lists of usability problems with a precise description. A usability problem was defined as a specific characteristic of the system that prevents task solving, frustrates the user, or is not understood by the user, as defined by Molich [12] and Nielsen [14]. Each evaluator also made a severity assessment for instance of a usability problem. The typical practice with severity is to make one general severity assessment for each problem expressed on a three-point scale, e.g. cosmetic, serious, and critical [12]. Yet this general severity assessment introduces a fundamental data analysis problem. Two users may experience the same problem very differently, and it is rarely clear how individual differences influence the general assessment. Moreover, we wanted to understand to what extent the severity changed from novices to experts. Therefore, we rated severity based on the extent to which it impacted the work process of each individual user. The severity ratings were necessary for addressing RQ2.

The individual problem lists from the 2 evaluators were merged into one overall list of usability problems. This was done in a negotiation process where the problems were considered one at a time until consensus had been reached. Out of the total number of 103 usability problems, 64 were identified by both evaluators, 17 only by evaluator 1, and 22 only by evaluator 2. The overlap between problems identified by the 2 evaluators suggests a low presence of the evaluator effect [9] and thus a high reliability of the merged list of problems. The resulting problem list was the basis for addressing RQ2. The evaluators also produced a 2-4 page log file for each of the 14 test sessions containing the exact times and descriptions of the users' interactions with the EPR system. The log file also describes whether the user solves each task, and to what extent the test monitor provides assistance. The extent to which each task was solved and the test monitor interference was necessary for addressing RQ1.

4. Findings

4.1. Effectiveness and efficiency (RQ 1)

Effectiveness reflects the accuracy and completeness of the subjects achieving certain goals and this includes indicators of quality of solution and error rates. In this experiment, we distinguish between completely and partially solved tasks. The mean numbers of solved tasks for the expert subjects were 6.29 (SD=1.11) tasks and for the novice subjects 3.57 (SD=1.27) tasks and a Wilcoxon signed rank test shows significant difference z=2.116, p=0.034. Thus, we found that the test subjects solved significantly more tasks as expert subjects than as novice subjects. The calculated standard deviations indicate high variance for the novice subjects; in fact the novice subjects on numbers of solved tasks ranged from 3 to 6 whereas the expert subjects ranged from 5 to 7. All expert subjects solved all 7 tasks either completely or partially while only two novice subjects solved all tasks and this difference is strong significant according to a Chi-square test ^{2}[1]=6.667, p=0.0098. Considering only completely solved tasks, four expert subjects failed to solve all 7 tasks within the given time frame while all 7 novice subjects failed to solve all tasks completely, but this difference is not significant ^{2}[1]=3.000, p=0.0833.

In conclusion, the expert users were more effective than the novices. The experts solved significantly more tasks and there was less variation than among the novices.

Efficiency reflects to the relation between the accuracy and completeness of the subjects achieving certain goals and resources spent in achieving them. Indicators often include task completion time, which we use in this experiment. Despite the significant higher number of solved tasks, we found no significant differences in mean values for the total task completion times z=1.402, p=0.161. The assignments enfold important variances and the two simple data entry tasks were solved faster by the experts, but we found no significant differences for any of the individual tasks.

In conclusion, the experts were faster for simple data entry tasks, though not significantly faster, and on more complex tasks there were no major differences.

4.2. Usability problems and severity (RQ 2)

We identified a total number of 103 usability problems. These top most of these were related to the three overall themes of 1) complexity of information, 2) poor relation to work activities, and 3) lack of support for mobility [11]. The novices experienced 83 of these 103 usability problems whereas the expert subjects experienced 63 of the 103 usability problems (this is shown in table 1). Attributing severity to the identified usability problems, the highest experienced severity for each problem is used. We found that the novices experienced 93% of the critical problems (25 of 27 problems) while the experts experienced 70% (19 of 27 problems). Similar distributions were identified for the serious problems where the novices experienced 80% of the identified problems compared 61% for the experts. Finally, minor differences were found for cosmetic problems: 65% for novices against 50% for experts.

Table 1. Total numbers of identified usability problems for the novices and experts.

	Novice (N=7)	Expert (N=7)	Total (N=14)
Critical	25	19	27
Serious	45	34	56
Cosmetic	13	10	20
All	83	63	103

Table 2 outlines key results on mean numbers of identified problems for the novices and experts. We found that the novice subjects experienced significantly more problems than the experts according to a Wilcoxon signed rank test z=2.159, p=0.031. However, this difference is mainly a result of more serious problems z=2.159, p=0.031, whereas we found no significant differences for the critical problems z=1.420, p=0.156 or the cosmetic problems z=1.876, p=0.061.

Table 2. Mean numbers of identified usability problems for the two setups.

	Novice (N=7)	Expert (N=7)	z	p
Critical	5.29 (1.50)	3.29 (1.98)	1.420	0.156
Serious	17.29 (3.09)	9.14 (2.97)	2.159	0.031
Cosmetic	8.86 (2.41)	11.43 (2.76)	-1.876	0.061
All	31.43 (4.93)	23.86 (4.49)	2.159	0.031

Figure 2 outlines problems unique to the novice subjects, problems unique to the expert subjects, and problems experienced by both novices and experts. 40 of the 103 identified problems were experienced by the novice subjects only and most of these problems concerned simple data entry tasks such as typing in values for patients. 43 of the 103 identified problems were experienced by both novice and expert subjects and they typically concerned advanced data entry or solving judgment questions. 20 problems were identified for experts only. These mainly concern functionality and services that the novices did not use for solving the same tasks, for example work task lists, because they were not familiar with those parts of the system.

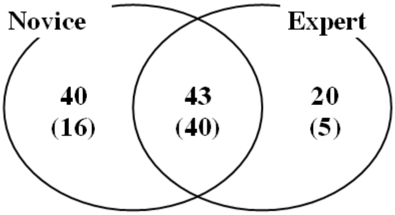


Figure 2. Distribution of the identified problems for the novices and experts. Numbers in parentheses show total numbers of problems subtracted unique problems.

Discarding problems from the distribution only experienced by 1 test subject, we see that most of the usability problems (40 of the 61) were identified in both the novice sessions and expert sessions. Further, the experts experienced 5 non-unique problems not experienced by any novice subjects and none of these 5 problems were critical. Accordingly, all critical non-unique problems were identified in the novice sessions.

The distribution of usability problems experienced by more than one test subject is illustrated in figure 3 below.

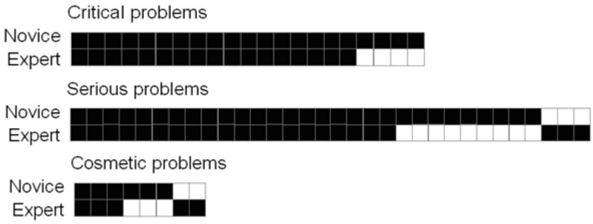


Figure 3. Distribution of usability problems identified by novices and experts in the two studies. Each column represents a usability problem. A black square indicates that the respective user group identified a problem. A white square indicates that a problem was not identified by that user group.

As illustrated in figure 3, 17 critical problems experienced by the novices were still experienced after one year of use. Both novices and experts experienced more than half of the serious problems, while nine serious problems were only experienced by the novices. The expert users, on the other hand, only experienced 3 serious problems not also experienced by the novices. In relation to the cosmetic problems, less than half were experienced by both novices and experts. 3 cosmetic problems were experienced only by novices and 2 only by experts.

In conclusion, there was a huge overlap of both critical and serious usability problems experience by novices and experts. Some problems disappeared over time, but far from all of them. At the same time, new serious and cosmetic problems appeared because more parts of the system were being explored.

Based on our instrumentation for problem identification and categorization, we classified problems according to how the individual test subjects experienced the problems. Thus, the same problem could be critical to one subject while cosmetic to another. 43 of the 103 usability problems were experienced by both the novices and the experts. Attributing the severities values between 1 and 3 where 3=critical, 2=serious, and 1=cosmetic problems, we can count the severity for each of the 43 problems. Considering the number of subjects experiencing the problems, each of the 43 problems was experienced on average by 3.61 (SD=2.19) novice subjects and on average by 3.39 (SD=2.01) expert subjects. But this difference is not significant according to a Wilcoxon signed rank test z=0.722, p=0.470. We further calculated the mean value for each of the 43 problems for the novices and experts. The mean value for novices was 1.91 (SD=0.51) and the mean value for the experts was 1.55 (SD=0.57) and this difference is significant z=3.963, p=0.001. Finally, we analysed the problems experienced in both the first and second evaluation on worst-case for each year. Here we found that the problems on average had a value of 2.19 (SD=0.59) whereas the experts on average experienced the problems to a mean value of 1.84 (0.75). This is significant according to a Wilcoxon signed rank test z=2.690, p=0.007.

In conclusion, a remarkably high number of problems were experienced both by novices and expert users. These problems were experienced significantly more severe for the novices, so the problems that remained became less severe after one year of use.

5. Implications for usability evaluations in health care

The implications for the choice of novice or expert users as test subjects are several. In relation to effectiveness, we found that the expert users completed significantly more tasks and had lower variance in task completion than the novices. This indicates that in situations where it is important for the software development process that every planned aspect of an expert system (such as en electronic patient record) is evaluated, one should consider using experts rather than novices in order for the evaluation sessions not to be held up. As discussed in relation to efficiency, this does not necessarily influence task completion time.

In relation to the identification of usability problems, we found a significant difference between the number of problems experienced by novices and experts. The implications of this finding are debatable. On one hand it can be stated that one should use novices because they enabled more problems to be identified. On the other hand, it could be argued that the use of experts supported the elimination of noise from "false" usability problems (typically rated as cosmetic). Regardless, however, our results show that when evaluating a system designed for highly specialized domain, such as health care, including users who are novices with the system but highly experienced with the use domain as test subjects can support the identification of as many critical and serious usability problems as when using system experts. This finding is important in situations where expert users may be a scarce resource for usability studies.

In relation to problem severity, we found a significant difference between the mean severity ratings for novices and experts, with the latter generally experiencing the usability problems of the system as less severe. The implications of this finding is primarily that when analyzing the data from a usability evaluation with novice users and making suggestions for subsequent response, designers should remember that even though time may not heal a system's usability problems, returning users will get familiar with the system, and that the cost associated with this learning may in some cases outweigh the costs of a redesign that may or may not be significantly better. This is especially important in relation to when responding to cosmetic usability problems.

6. Conclusions

This paper has reported from a longitudinal study in health care where we have compared the usability of an electronic patient record system as experienced by novice and expert users. The usability of the system was measured in different ways. The first measure was effectiveness and efficiency. The expert users were more effective than the novices; they solved significantly more tasks and there was less variation than among the novices. However, we found no significant differences on task completion times for the individual tasks. The second measure was the number and severity of usability problems experienced. The novice subjects experienced significantly more critical and serious problems, whereas the experts experienced significantly more cosmetic problems. Thus there was a huge overlap of both critical and serious usability problems experience by novices and experts.

Some of the overall results confirm the outcome of other studies. The most striking results are that the expert users are not more efficient on complex tasks and that a remarkable number of serious and critical problems with the electronic patient record system still remained after one year of extensive use. Thus we conclude that time does not heal usability problems. Even though time allows people to learn strategies for overcoming a system's specific peculiarities, poor design remains poor.

On the basis of our findings, we propose the following five take away points for usability evaluations in health care:

1. Time does not heal. Although some problems were not experiences as severe, they still remained after one year of extensive use. Poor design remains poor.

2. Expertise reduces experienced severity of usability problems. When testing with novice users, evaluators must take into consideration that some problems may not be as severe as it seems.

3. Solve usability problems early. If usability problems do not disappear over time, we should get rid of them as soon as possible. There will always be novice users – new employees, temporary staff, etc.

4. Evaluate with both novice and expert users and use their different experience of the system as a lens to get a more complete picture of a system's usability. They both represent a prospective user group.
5. All experienced problems are relevant regardless of user expertise – but problems for expert users should be given priority in re-design. Problems mostly severe to novice users might be given priority through teaching (if applicable).

References

[1] Allen, B. (1994). Cognitive Abilities and Information System Usability. Information Processing and Management, 30, 177-191.

[2] Bailey, R.W., Allan, R.W., Riello, P. (1992). Usability Testing vs. Heuristic Evaluation: A Head-to-Head Comparison. Proceedings of the Human Factors and Ergonomics Society 36th Annual Meeting, HFES, pp. 409-413.

[3] Bednarik, R. and Tukiainen, M. (2005). Effects of display blurring on the behavior of novices and experts during program debugging. Extended Abstracts of CHI 2005, pp 1204 - 1207, New York: ACM

[4] Bourie, P. Q., Dresch, J., and Chapman, R. H. (1997). Usability Evaluation of an On-line Nursing Assessment. Proceedings of AMIA Symposium.

[5] Dillon, A. and Song, M. (1997). An Empirical Comparison of the Usability for Novice and Expert Searchers of a Textual and a Graphic Interface to an Art-Resource Database. Journal of Digital Information, 1(1).

[6] Dix, A., Finlay, J., Abowd, G.D. and Beale R. (2004). Human-Computer Interaction (third edition). Harlow, England: Pearson Education Limited.

[7] Frøkjær, E., Hertzum, M. and Hornbæk, K. (2000) Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated? CHI Letters, 2(1), 345-352.

[8] Ishii, N. and Miwa, K. (2002) Interactive processes between mental and external operations in creative activity: a comparison of experts' and novices' performance. Proceedings of the 4th conference on Creativity & cognition table of contents, pp 178-185, New York: ACM.

[9] Jacobsen, N.E., Hertzum M. and John, B.E. (1998) The Evaluator Effect in Usability Tests. Proceedings of CHI'98. New York: ACM Press.

[10] Karat, C. M., Campbell, R., and Fiegel, T. (1992). Comparison of Empirical Testing and Walkthrough Methods in User Interface Evaluation. Proceedings of CHI'92, pp. 397-404. New York: ACM Press.

[11] Kjeldskov, J. and Skov, M. B. (2007) Exploring Context-Awareness for Ubiquitous Computing in the Healthcare Domain. Personal and Ubiquitous Computing (in press)

[12] Molich, R. (2000). Usable Web Design (In Danish). Ingeniøren|bøger.

[13] Nielsen, J. (2000). Novice vs. Expert Users. Alertbox, February 6, 2000. http://www.useit.com/alertbox/20000206.html

[14] Nielsen, J. (1993). Usability Engineering. San Diego: Morgan Kaufmann.

[15] Preece, J., Rogers, Y. and Sharp H. (2002). Interaction Design: Beyond Human-Computer Interaction. New York: John Wiley & Sons, Inc.

[16] Prümper, J., Frese, M., Zapf, D. and Brodbeck, F. C. (1991) Errors in computerized office work: differences between novice and expert users. ACM SIGCHI Bulletin 23(2), 63-66

[17] Rubin, J. (1994). Handbook of Usability Testing: How to plan, design and conduct effective tests. New York: John Wiley & Sons, Inc.

[18] Urokohara, H., Tanaka, K., Furuta, K., Honda, M. and Kurosu M. (2000). NEM: "Novice Expert ratio Method" A Usability Evaluation Method to Generate a New Performance Measure. Extended Abstracts of CHI 2000, pp 185-186, New York: ACM