Evaluating the Usability of Home Healthcare Applications

Anders Bruun and Jan Stage
Aalborg University, Denmark

ABSTRACT

Home healthcare applications have the potential to reduce healthcare costs and improve the quality of life for elderly people who prefer to stay in their own homes instead of making frequent visits to the hospital. This requires ambient assisted living applications that fulfil relevant needs of the users; yet it also requires applications with a high level of usability in order to achieve user acceptance, especially when the target user group is elderly people. This chapter proposes a method to be used for conducting usability evaluations of smart healthcare applications. It includes a report from a usability evaluation where the method was used to evaluate a simple home healthcare application for collecting personal health data in the home. The usability evaluation demonstrates that the method presented here facilitates identification of key usability problems, while the efforts required to conduct the evaluation are considerably reduced compared to conventional methods.

INTRODUCTION

There is growing interest in devices for home healthcare applications. At world level, the life expectancy will increase from 67.2 years in 2010 to 75.4 years in 2050, and in more developed regions life expectancy will rise to 82.4 years by 2050 (UN, 2006). This has considerable consequences for healthcare budgets. The number of people with chronic illness is also increasing and, due to frequent checkups at hospitals, these patients face reduced quality of life, because they have limited freedom to perform their daily activities.

The aim of smart healthcare applications for the home is to reduce healthcare costs and at the same time increase the quality of life for patients. Home healthcare applications allow patients to conduct measurements from their own home (e.g. glucose measurements for diabetes patients) and send the results to the hospital. Other applications put even more emphasis on self-management by supporting patients in taking care of their own treatment. If home healthcare applications are successful, they will reduce the workload of the medical staff, and relieve the patients from visits to the hospital or even hospitalization (Kaufman et al., 2003).

For home healthcare applications to be successful, they must be safe and provide relevant functionality. Many researchers have studied home healthcare applications and frameworks that aid in reducing the societal and individual costs of chronically ill elderly. The focus has been on functionality and regards ubiquitous biological monitoring using mobile phones, wearable sensory devices, multi modal platforms, framework and architecture descriptions and literature reviews of effects (Eikerling, et al., 2009; Fensli and Boisen, 2008; Jaana and Paré, 2006; Pascual et al., 2008; Sasaki et al., 2009; Sashima et al., 2008; Souidene et al., 2009; Taleb et al., 2009).

There are, unfortunately, numerous examples of systems that fail despite having the right functionality, simply because the prospective users cannot use the system for its intended purpose. A problematic or incomprehensible user interface is a typical source of such problems. Usability is a measure of the extent to which prospective users are able to apply a system in their activities (Rubin, 1994). A low level of usability means that users cannot work out how to use a system, no matter how elaborate its functionality is (Nielsen, 1993).

The potential of home healthcare applications can only be realized if the usability of the applications is adequate. Thus a high level of usability is a prerequisite for achieving savings on healthcare costs and a better quality of life for patients through use of home healthcare applications. A high level of usability is particularly important when the main user group is elderly people, who may be constrained by motor, perceptual, cognitive and general health limitations (Fisk and Rogers, 2002) and, in addition, may have a low level of computer literacy.

In this chapter, we focus on the challenges of conducting usability evaluations of home healthcare applications. We present a resource-economic method for usability evaluation and illustrate its use through a case study where we evaluated quantitative and qualitative usability aspects of a home healthcare application that has elderly people as the target user group. The efforts devoted to the usability evaluation are also presented and compared with the conventional approach to usability evaluation.

BACKGROUND

Usability evaluation is the process of assessing the level of usability of an application. In the field of human-computer interaction (HCI), there is a distinction between two categories of usability evaluations: Formative and summative. The objective of a formative usability evaluation is to uncover issues in a yet unfinished user interface. This type of evaluation is also referred to as an exploratory or assessment test and may range from an evaluation of high level initial design concepts addressing issues such as support of the users mental model, basic screen navigation etc. to more detailed low level interactions. Early formative evaluations may be conducted on paper prototypes whereas later evaluations typically are conducted on implemented user interfaces. The objective of a summative evaluation, on the other hand, is to test compliance of system usability to a set of expressed requirements. This is done late in the development cycle on an almost completed user interface (Rubin, 1994).

The conventional approach to formative usability evaluation is typically conducted in a laboratory setting. The prospective users are being observed as they interact with a software application and/or hardware device while solving realistic task scenarios. A number of usability specialists are involved, where one is acting as a test monitor and others as data logger and technician. The test monitor is sitting next to the test participant (a prospective end user) during the evaluation session introducing task scenarios and making sure, that the users think aloud while interacting with the system. The data logger typically sits behind a one-way mirror in a control room observing and taking notes when usability problems occur. As the evaluation sessions are recorded on video, a technician is also present in the control room to control camera set-up and video recording (Rubin, 1994).

A single evaluation session usually takes about 45 minutes (it should not exceed this limit), and typically five to eight test participants are involved. This sums up to about four or five hours of video material, which is analyzed by two or three evaluators with the purpose of identifying usability problems in the application or device. The evaluators are typically the test monitor, data logger and/or technician that have been involved in the evaluation sessions. According to our experiences one hour of video requires four to six hours of analysis, which means that we typically spend about 15 to 30 hours per evaluator analyzing the results. The resulting list of usability problems is an indispensable resource for improving the application.

The key drawback of the conventional approach to formative usability evaluation is a considerable demand for resources to plan tests, establish a test setting, conduct the tests and identify usability problems in a rigorous analysis of the hours of video of the users' interaction with the system (Bruun et al., 2009). This resource demand prevents many organizations from deploying usability evaluation methods in their development processes. There is simply not enough time and resources to conduct this kind of rigorous analysis (Bak et al., 2008).

A high level of usability is a critical factor in achieving success with home healthcare applications. Unfortunately, many development organizations are not conducting usability evaluations, which lead to products with poor usability. This ultimately affects the end users in a negative manner. This obstacle may, however, be overcome by applying modern methods for usability evaluation that reduce the resource demands.

RELATED WORK

In this section we provide an overview of previous research on usability evaluation of smart healthcare applications and resource-economic usability evaluation methods.

Usability Evaluation of Smart Healthcare Applications

A number of research activities have studied smart healthcare systems and frameworks that aid in reducing the societal and individual costs of chronically ill elderly. The focus here has been on the functionality that is required from such systems. Examples are technology for ubiquitous biological monitoring using mobile phones, wearable sensory devices, multi modal platforms, framework and architectural descriptions and literature reviews of observed medical effects (Eikerling, et al., 2009; Fensli and Boisen, 2008; Jaana and Paré, 2006; Pascual et al., 2008; Sasaki et al., 2009; Sashima et al., 2008; Souidene et al., 2009; Taleb et al., 2009). The target user group of these systems is primarily elderly people. There are, however, few studies of the usability of healthcare applications, in particular where the target user group is chronically ill or elderly people.

A significant number of studies deal with health care systems where the target user group is professional medical staff. This includes evaluation of the usability of desktop, mobile and other healthcare systems with the aim of reducing medical errors introduced by technology. Examples are systems designed for supporting handheld prescription writing, decision support, ordering of lab tests, patient records, family history tracking etc. (Ginsburg, 2004; Johnson et al., 2004; Kushniruk et al., 1996; Kushniruk and Patel, 2004; Kushniruk et al., 2005; Linder et al., 2006; Peleg et al., 2009; Peute and Jaspers, 2007).

Less research activity has been devoted to usability problems with systems where the target users are patients. A notable exception is Kaufman et al. (2003) who conducted a case study where a home healthcare system for elderly diabetes patients was evaluated through interviews, cognitive walkthrough and field usability testing. The evaluated system featured video conferencing, transmission of glucose and blood pressure readings, email, online representation of clinical data and access to educational materials. The study focuses on a methodology for conducting usability evaluation. It also provides a basic overview of barriers such as individual competencies, system usability issues and contextual variables. The main objective in Kaufman et al. (2003) is to reveal patient related factors causing usability issues. This was done through comprehensive microanalysis of video material involving 25 test participants, thus the focus in this study was not on resource saving usability evaluations and the authors also acknowledge that the analysis was exceedingly time consuming.

The research presented here represents significant work on the needed functionality of home healthcare applications as well as on methods for evaluating the usability of such systems. The focus is on functionality and usability problems while there is no concern so far for resource demands.

Resource-Economic Usability Evaluation Methods

In the HCI discipline, there has been considerable focus on the resources that are required to conduct usability evaluations.

The inspection methods were originally presented as a radical attempt to overcome the resource-demand barrier, e.g. (Molich, 2000; Nielsen, 1997; Nielsen, 1994). Several varieties of inspection methods exist, see Nielsen (1994) for a short overview. Instead of observing users interacting with a system, the system is "inspected" by usability experts with the goal of identifying potential usability problems, thus this approach is non-user based. Examples of inspection methods are Heuristic Evaluation (HE), Cognitive Walkthrough (CW) and Metaphors of Human Thinking (MOT). HE involves inspection of every program dialogue to check if they follow a set of usability heuristics (Nielsen, 1994). CW tries to simulate the users' problem solving process, thereby identifying problems that the users may encounter when working towards a goal (Nielsen, 1994). MOT is based on five essential metaphors of human thinking, which provide usability experts with guidelines on how to consider the users' thinking process (Hornbæk and Frøkjær, 2004).

Heuristic Evaluation has been shown to facilitate identification of more problems than Cognitive Walkthrough (Bruun et al., 2009; Nielsen, 1994), which is also the case when comparing Metaphors of Human Thinking to Cognitive Walkthrough (Frøkjær and Hornbæk, 2008; Hornbæk and Frøkjær, 2004). Metaphors of Human Thinking is shown to identify the same number of problems as Heuristic Evaluation (Frøkjær and Hornbæk, 2008; Hornbæk and Frøkjær, 2004). While being more time consuming, the conventional user-based laboratory evaluation facilitate identification of considerably more usability problems than Heuristic Evaluation (Karat et al., 1992; Nielsen, 1994).

Iterative Testing and Evaluation (RITE) maintains observation of users while reducing the effort. This method is based on a traditional think-aloud laboratory usability test. The primary strategy is to ensure that identified usability problems are corrected within a short timeframe. A secondary objective is to reduce the resources spent on testing and implementing fixes (Medlock et al., 2002). Using RITE, problems are identified on the fly. If possible, problems are fixed immediately, and a new prototype is used for the following tests. If a problem is not fixed easily, more data about it is collected during the following tests. The RITE approach requires experienced usability experts as well as developer resources during the tests. The advantages of this method are that the identified usability problems are solved and the fixes are tested too.

Cooperative evaluation focuses on user involvement and quick delivery of evaluation results (Monk et al., 1993). The user is actively involved in the evaluation by providing comments and assessments of the system while using it. There is one evaluator involved who helps the user think aloud and takes notes for the evaluation report. A similar approach is the use of sparse prototypes for quick and early evaluation of a system (McCarthy et al., 1993).

Additional focus on resources can be seen with Instant Data Analysis (IDA) The main idea is to reduce the time spent on post-test analysis while still identifying the most critical usability problems, by maintaining a conventional user-based think-aloud test. The identification of usability problems is based only on the observations made by the test monitor and data logger. Thus no video data analysis is done. An empirical study showed that IDA supported identification of almost as many usability problems as the conventional approach with detailed video-based analysis, and this was achieved using only 10 % of the time compared to the conventional approach (Kjeldskov et al., 2004).

The research presented here illustrate that that the HCI discipline has focus on resource demands, and several methods that are more resource-economic than the conventional approach have been proposed.

INSTANT DATA ANALYSIS (IDA)

In this chapter we discuss and illustrate how the Instant Data Analysis (IDA) method can be used for usability evaluation of smart healthcare applications. In this section, we describe the method generally.

The IDA method is presented in Kjeldskov et al. (2004). The method resembles key characteristics of current commercial practice in usability evaluation by avoiding a tedious post-test video analysis. Yet IDA replaces this with a different but systematic form of post-test analysis, as opposed to some approaches in practice where the systematic post-test analysis is discarded completely.

IDA is conducted with four to six prospective end users. Each user participates in a think-aloud usability test session, lasting between 20 and 45 minutes, depending on the system and task scenarios. By using four to six users the test and analysis sessions may be conducted in a single day while ensuring a robust dataset for a formative evaluation. In each test session, the test monitor is present next to the user, to introduce the procedure and make sure the user thinks aloud. Meanwhile the data logger records user behaviour and usability problems, which will be used for problem identification and categorization during the analysis session.

Immediately after the tests sessions are completed, the analysis session is carried out. It is important that the analysis is conducted on the same day as the test sessions. The analysis involves three participants: the test monitor, the data logger and a facilitator, where the latter is usually absent during the user tests. The role of the facilitator is to manage the analysis session. He asks questions for clarification and note identified usability problems on a whiteboard or flip-over while they are presented by the test monitor and data logger. The facilitator should also make sure to categorize problems in themes as the analysis session progresses. He must also avoid redundancy in problem descriptions etc.

The description of IDA in Kjeldskov et al. (2004) provides no further details about the analysis session. We have found it useful to structure it in the following four steps:

1. Brainstorm: The test monitor and data logger brainstorm to identify the usability problems that they remember.
2. Task review: The test monitor and data logger review all tasks, one by one, to remember and discuss usability problems that occurred during each task.

3. Note review: The data logger reviews the written notes to remember additional problems and discuss them with the test monitor.
4. Severity rating: The test monitor and data logger discuss the severity of each identified problem and rate it as critical, serious or cosmetic, cf. Molich (2000)

During the analysis session, the facilitator notes and organizes all identified usability problems on the whiteboard, see Figure 1. When the four steps are completed, the test monitor and logger leave the room, and the facilitator writes up the list of usability problems from the notes on the whiteboard. The list of usability problems includes short descriptions and clear references to the application or device evaluated. The list may be ranked according to severity of the problems. The list is subsequently validated and corrected by the test monitor and data logger, typically on the following day.

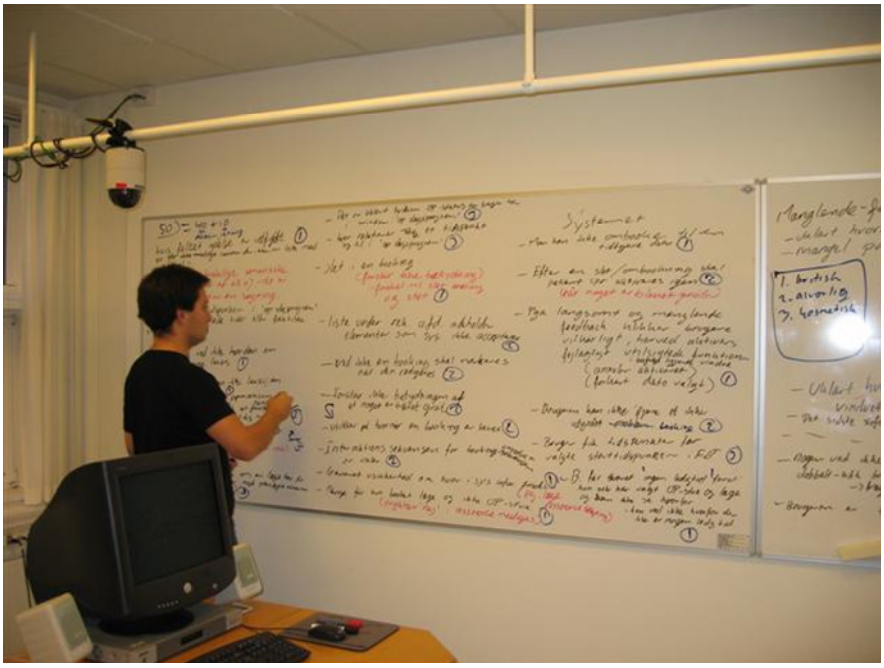

Figure 1. The facilitator notes the identified usability problems on the whiteboard.

CASE STUDY: EVALUATION OF A HOME HEALTHCARE APPLICATION

We have conducted an empirical study where we used the IDA method to evaluate the usability of a home healthcare application. In this section, we use this as a case study to illustrate how IDA can be used to evaluate smart healthcare applications. We also discuss how resource-economic IDA is. This is accomplished by using a conventional video-based analysis (VBA) as benchmark.

Application

The system we evaluated is a system intended for home use by elderly people for monitoring their health. It includes a healthcare system (HCS) device for data collection and transmission with a display, a speaker and four buttons for interaction. The HCS is illustrated in Figure 2. With secondary devices such as blood pressure meter, blood sugar meter and scales, users can make measurements at home and transfer them to the HCS via Bluetooth, an infrared link or a serial cable. At regular intervals, the device also asks the patients various pre-programmed questions regarding their health. The system automatically transfers collected data to a healthcare center via an internet connection, where a nurse, doctor or other person is monitoring the health for a group of patients. The system is sent by mail to the patients in a package with a manual. The usability evaluation focused both on the system and the manual.

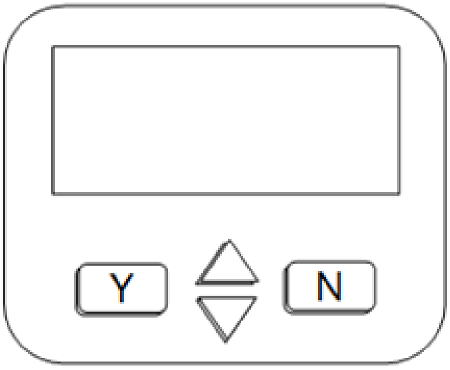

Figure 2. Sketch of the HCS data collection and transmission device.

Setting

The tests sessions were conducted in a usability laboratory, see Figure 3. In Subject room 1, a user was sitting at the table using the system. The test monitor was sitting next to the user. Two data loggers and a technician who controlled cameras and microphones sat in the control room during all tests.

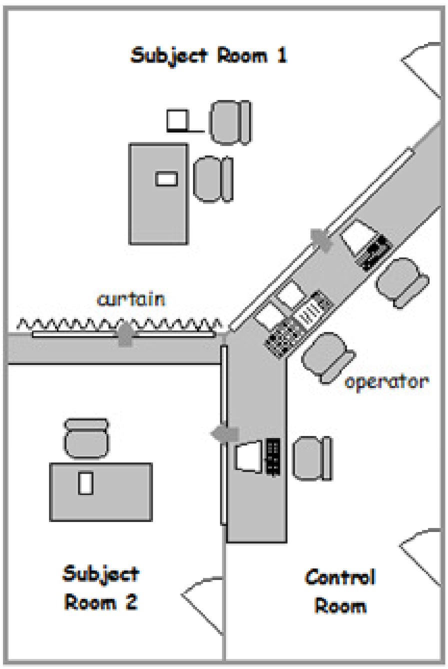

Figure 3. The usability laboratory used for the evaluation.

Participants

The system was evaluated with five users, four male and one female. Since the system primarily is intended for use by elderly people, the test subjects were between 61 and 78 years of age. None of them had previous experience with this or any similar system. Their experience in using electronic equipment in general varied; two were novices, two were slightly experienced and the last was experienced in using electronic equipment in general.

Six usability evaluators were involved in the study in different roles. They were all graduate students specializing in HCI and on their tenth semester, doing a master thesis. They were all experienced in conducting usability evaluations and had the same educational background. None of them had worked in healthcare before, and none of them knew the product in advance. In the evaluation, one of them served as test monitor in all five test sessions and two served as data loggers in all tests. After the tests, the test monitor, one of the data loggers and a facilitator, who did not observe the tests, conducted the IDA analysis. The other data logger and two evaluators, who did not observe the tests, conducted a conventional video-based analysis (VBA). It should be noted that the number of evaluators was unusually high. This was necessary to assess how resource-economic IDA is. A standard IDA evaluation only requires three evaluators.

Figure 4. A test participant and the test monitor. The picture is from the video recording. The small picture in the upper right hand corner shows the interaction.

Procedure

Before the test sessions started, the test participants were asked to fill in a questionnaire with demographic information about their age and previous experience in using electronic equipment. The test monitor then introduced the system and evaluation procedure. This included an introduction to the think-aloud protocol.

In each test session, the task assignments were given to the test subjects one at a time. The test monitor's job was primarily to ensure that the test participants were thinking aloud and give them advice if they got completely stuck in a task. There were five tasks, see Table 1. The task assignments used in the usability evaluation.Table 1. The first had to be solved entirely, because other tasks depended on its result. After completion of each test session we conducted a post test interview with each of the five participants to get their overall opinion about the system.

Task #	Task
1	Connect and install the HCS and secondary devices.
2	Transfer the data from the blood sugar meter to the HCS. The blood sugar meter is connected using a cable.
3	Measure the weight and transfer the data from the scale to the HCS.
4	A new wireless blood sugar meter is used. Transfer the data from this to the HCS.
5	Clean the equipment.

Table 1. The task assignments used in the usability evaluation.

Data Collection

All test sessions were recorded using video cameras and a microphone. The video showed the HCS screen and the user's face, see Figure 4. We recorded a total of four hours of video. Two data loggers made written log files during the test sessions.

Data Analysis

The data analysis was carried out separately with IDA and VBA by two independent teams. Each team employed the procedure described below. There was no communication between the two teams until both had completed their analysis.

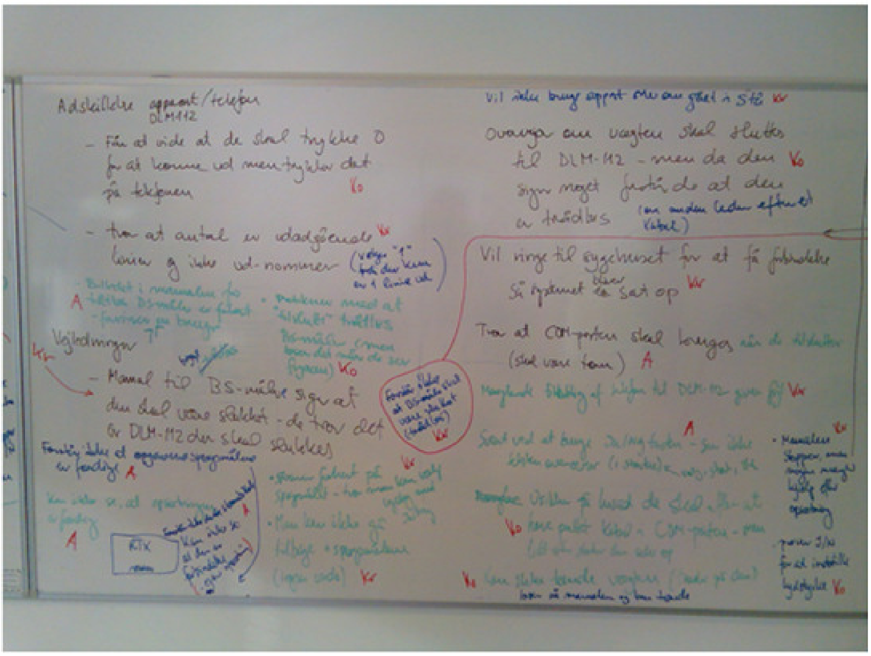

Figure 5. Part of the whiteboard from the IDA analysis with colored problem descriptions.

The data analysis with IDA was carried out as described above. The test monitor, one of the data loggers and a facilitator conducted the IDA session immediately after all test sessions were completed. It involved the four steps described above. The Brainstorm step took 20 minutes, the Task review step took 30 minutes, and the Note review step took 52 minutes. During the IDA analysis session, the facilitator noted and organized all identified usability problems on the whiteboard. The problems identified in each step were marked with different colours (green, blue and black), see Figure 5. This made it possible later on to recognize in which step of the IDA session a usability problem was identified. After completing the fourth step, the test monitor and logger left the room, and the facilitator wrote up the list of usability problems from the notes on the whiteboard. The list was validated and corrected by the test monitor and data logger the following day.

The data analysis with the VBA method was carried out by three evaluators who analyzed the video material individually and made their own list of identified usability problems. The severity of each problem was categorized as critical, serious or cosmetic. The three lists of usability problems were discussed in the team and merged into one list of VBA problems. When there was doubt or disagreement whether problems should be combined or split, or how they should be categorized, the video material was reviewed and discussed until agreement was reached. To measure the evaluator effect, the any-two agreement was calculated. The result was 40.2%, which is well above the minimum of 6% and close to the 42% maximum found in other studies (Hertzum and Jacobsen, 2003). This indicates a robust result as the evaluators identified a large proportion of joint problems based on their individual interpretations of the video material.

To facilitate comparison between IDA and VBA, the two separate lists of usability problems were merged into a total list of usability problems. The test monitor and the data logger from IDA and the three evaluators from VBA did this together. Disagreements were discussed until consensus was reached. In cases where the VBA and IDA lists did not have the same categorization for a particular problem, the proper categorization was discussed until agreement was reached. In this process, some problems were split into more detailed problems or merged with other problems.

PERFORMANCE WITH IDA

In this section, we present the results of applying the IDA method for identification of usability problems for the home healthcare application. We also include data about VBA in order to illustrate how resource-economic IDA is.

	IDA	**VBA**	**Total**
Critical	16	13	18
Serious	13	13	17
Cosmetic	8	18	19
Total	37	44	54

Table 2. Number of identified usability problems.

Number of Identified Usability Problems

With IDA, we identified a total of 37 usability problems, see Table 2. Most of the problems were either critical or serious. Two examples of the identified problems are:

- Connecting: Does not understand the text “Comunicating with the server”.
- Scales: Is missing an “undo” facility when an option has been selected.

To assess the performance of IDA in identifying usability problems, we can use VBA as a benchmark. The two methods together identified a total of 54 usability problems. The difference between the total number of problems identified with the IDA and VBA methods is not significant with a Fisher’s exact test (p=0.1819). For critical problems, there is no significant difference either (p=0.418); this is also the case for serious problems (the p-value is the same). However, for cosmetic problems, the difference is very significant (p=0.0011), as the VBA analysis revealed significantly more of this type. These numbers reflect that there is no significant difference between IDA and VBA except for cosmetic problems, and they are the least important to identify. Thus for identification of important usability problems, IDA's performance is comparable to the conventional approach.

Time Spent on Data Analysis

The purpose of applying the IDA method is to reduce the time spent on usability evaluation without sacrificing the quality of the result. IDA focuses especially on data analysis, because thios is the single most time-consuming activity in a conventional usability evaluation.

	Test monitor	**Data logger**	**Facilitator**	**Total**
Analysis session	2 h	2 h	2 h	6 h
Writing problem list			1.5 h	1.5 h
Validating problem list	1 h	1 h	1 h	3 h
Total	3 h	3 h	4.5 h	11.5 h

Table 3. Time spent on data analysis with the IDA method.

The time spent on data analysis with IDA is shown in Table 3. All three persons involved in the evaluation participated in the analysis session, and this amounts to a total of 6 hours. Then the facilitator wrote up the problem list alone. This was validated with the other two participants. The total time used on data analysis amounted to 11.5 hours.

To understand the reduction of effort better, we have compared this to a conventional video-based analysis. The time spent on that is shown in Table 4. Here, the three persons involved in the evaluation used about 14 hours on average to analyze the video recordings. The merging of their individual problem lists took another 6 hours. The total time spent on data analysis with this approach amounted to 59.75 hours. Thus the time spent on analysis using IDA is roughly five times lower than the time spent on VBA.

	Evaluator 1	**Evaluator 2**	**Evaluator 3**	**Total**
Identifying problems	13.5 h	13.75 h	14.5 h	41.75 h
Merging problem lists	6 h	6 h	6 h	18 h
Total	19.5 h	19.75 h	20.5 h	59.75 h

Table 4. Time spent on data analysis with the VBA method.

Fewer VBA Evaluators as Benchmark

When using VBA as a benchmark for the performance of IDA, it is crucial how the number of usability problems identified and the time spent on data analysis with the two methods are compared. The results presented above are based on three evaluators using VBA as this matches the number of persons involved in the IDA analysis. However, it can be argued that this comparison disadvantages VBA. Moreover, there are only two evaluators involved in identifying problems with IDA, because the facilitator only notes problems identified by the two others.

We have re-analyzed our data and calculated how many problems each pair of the three VBA evaluators identified together, which is shown in Table 5. The best case for VBA is the pair with evaluator 1 and 2 as they identify a total of 43 problems which is only one less than the number identified by all three evaluators, and there is no difference in the number of critical and serious problems. The worst case for VBA is the pair with evaluator 2 and 3 as they only identify a total of 33 problems, but on critical and serious problems they are only one below the number of problems identified by all three evaluators. In both cases the time spent is close to 40 hours.

	Evaluator 1 and 2	Evaluator 1 and 3	Evaluator 2 and 3	All 3 evaluators	IDA
Critical	13	13	12	13	16
Serious	13	12	12	13	13
Cosmetic	17	16	9	18	8
Total	43	41	33	44	37
Time spent	39.25 h	40 h	40.25 h	59.75 h	11.5 h

Table 5. Problems identified and time spent by the three combinations of VBA evaluator pairs, by all three VBA evaluators together and by the IDA evaluators.

A Fishers exact test between IDA and the best VBA pair (evaluator 1 and 2) gives no significant difference, neither for the total number of problems (p=0.2721), nor for critical problems only (p=0.4018). Even with only two evaluators, the time spent on the analysis is four times higher for VBA than with IDA.

It might be argued that we could reduce the number of VBA evaluators to a single person. This would make the time comparable to IDA as there is no need for merging problem lists. Yet a single VBA evaluator would produce low quality results because of the evaluator effect (Hertzum and Jacobsen, 2003).

Unique Problems

The detailed distribution of usability problems is shown in Figure 6. Each cell corresponds to a single usability problem. A black cell means that the method revealed that particular problem, and a white cell means that it was not found. A unique problem is one that is found with only one of the methods. These numbers appear from the figure. It is interesting because it may emphasize shortcomings with the method. We identified five critical problems that were only found with IDA. On the other hand, there were two critical problems not identified with IDA but with VBA. In the case of serious problems, we identified four problems with IDA that were not found with VBA and vice versa.

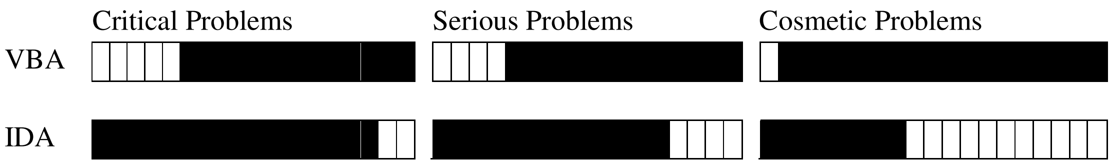

Figure 6. Individual usability problems identified with IDA and VBA.

Four of the five unique critical problems identified with IDA were experienced in the first task, the setup of the HCS. The fifth problem was not directly related to the system as it concerned the participants' reluctance to contact the technical support staff to get help. The two critical problems found with VBA but not with IDA were experienced in the first task. One of them was related to missing information on the display, and the other to a software bug, which caused the system to restart during setup.

The unique serious problems identified with IDA occurred in different tasks and typically relate to missing feedback from the system. One problem was, however, not related to a particular task, but more to the overall nature of the system. The unique serious problems found with VBA were all related to the first task, to physical setup, software bugs, missing feedback and server connection errors.

Quality of Categorizations

Categorization is the activity where usability problems are classified as critical, serious or cosmetic. The categorizations made with IDA are shown in Table 6. In the table, the number after merging the IDA and VBA problem lists are also indicated. It should be noted that when the two lists were merged, some problems were split into several problems or vice versa. This explains the differences in the total number of problems compared to table 2.

	IDA	VBA
Critical	17 (16)	10 (13)
Serious	12 (13)	11 (13)
Cosmetic	6 (8)	25 (18)
Total	35 (37)	46 (44)

Table 6. Severity ratings before merging the IDA and VBA problem lists. The number in parentheses is after merging.

Before merging, 49% of the IDA problems were categorized as critical, 34% as serious and 17% as cosmetic. The similar distribution for the VBA list was 22% critical, 24% serious and 56% cosmetic. Thus the IDA problems were generally categorized as more serious than the same problems were with VBA. After merging the problem lists, 7 of the 35 original IDA-problems were categorized as less serious, and 3 became more serious. 10 of the original VBA problems were, during the merging, categorized as more serious categorization, and none were categorized as less serious. Overall, the problems identified with the IDA method were categorized more seriously than the problems identified with the VBA method.

	Brainstorm 20 min (20%)	**Task Review** 30 min (29%)	**Note Review** 52 min (51%)	**Total** 102 min (100%)
Critical	7	3	7	17
Serious	4	4	4	12
Cosmetic	2	3	1	6
Total	13 (37%)	10 (29%)	12 (34%)	35 (100%)

Table 7. The number of problems identified in the three IDA steps.

Problems Identified in Each IDA Step

The number of usability problems identified in the three steps of the IDA analysis session is shown in Table 7. The number of problems identified in each step is approximately the same. But looking at the individual problems, it turns out that some of the identified critical problems during the brainstorm were split into more detailed problems during the last step (reviewing the notes). For example, one critical problem from the brainstorm step occurred also as four separate and more detailed critical problems when the notes were reviewed, and therefore, it was deleted from the list of brainstorm problems. This explains why the number of critical problems identified during the brainstorm is not higher.

The last step, even if it was the most time consuming, turned out to be important, because it contributed with important details to problems already identified and added new problems as well.

DISCUSSION

The goal of this chapter is to present a method that is relevant for usability evaluation of smart healthcare applications. In the Introduction, it was argued that in order to realize the potential of smart healthcare application, it is necessary to evaluate their usability, and that is only realistic if the demand for resources can be reduced. We have presented the IDA method that aims to reduce the resources required to conduct usability evaluations by significantly reducing the time spent on data analysis.

The literature on usability of healthcare systems is very sparse on details about the evaluations, including measurement of the time spent. Kaufmann et al. (2003) describes their analysis process in detail, but there is no measure of time except a statement that it was very time consuming.

The case study described above replicates an earlier study by Kjeldskov et al. (2004) that was conducted under similar conditions. The main difference is that the system was an administrative healthcare system, and the number of evaluators differed. In our study, we identified 89% of all critical problems with the IDA method, which is very similar to Kjeldskov et al. (2004) where they identified 85% of all critical problems. For serious problems, we found 76%, which is close to the 68% found by Kjeldskov et al. (2004). This illustrates that the performance we achieved with IDA in terms of problem identification is comparable to the previous study, and in both studies the number of problems identified with IDA is comparable to the conventional approach.

The effort required to conduct data analysis of a usability evaluation with IDA was five times less than the conventional approach. Even if we reduced the number of VBA evaluators to two and picked the best pair, the VBA method still requires more than four times the effort compared to IDA. In Kjeldskov et al. (2004) the result was that VBA required 10 times more person hours than IDA, which is a considerable difference compared to our study. In Kjeldskov et al. (2004) the evaluators conducted only the brainstorm part of the analysis, whereas we did a brainstorm, task review and note review, see Table 7. The task and note reviews of the analysis activity were the most time consuming (30 and 52 minutes respectively). Thus the more systematic and thorough approach does, not surprisingly, reveal more usability problems, but it also requires more time for analysis. In any case, the results show that IDA reduces the time spent on data analysis with at least 80%. In this respect, the goal of this chapter is fulfilled.

Our study was based on one group of evaluators using the IDA method, and this was compared to one group using VBA. This weak basis with only a single case seems to be an obvious limitation. Yet the evaluation was conducted in full scale and in a real evaluation process in collaboration with a software organization. To conduct multi-case experiments on such a realistic scale is nearly impossible. This is also reflected by the fact that key references on usability evaluation research area assess and compare methods based on a single or at most two cases, e.g. Karat et al. (1992) and McCarthy et al. (1993). A review of empirical studies of usability evaluation methods also accepts that a single case study is difficult enough to establish (Gray and Salzman, 1998).

Another characteristic of our study is that it involved only five users. Here, it should be noted that in the area of Human Computer Interaction it is customary to conduct formative usability evaluations using five test participants as this, from a cost/benefit point of view, is the most feasible. This number is based on studies conducted by Nielsen and Landauer (1993) showing that by using five test participants evaluators are able to identify 85 % of the total number of usability problems.

The system that was evaluated in our study is a new type of healthcare system for monitoring elderly people in their homes. This is not a standard application, neither for desktop computer applications nor for healthcare applications. However, the system used in Kjeldskov et al. (2004) was an administrative system for booking resources in a hospital. Although the two studies focus on very different types of applications we still obtain similar results.

A fundamental characteristic of a usability evaluation and the data analysis is that it is based on subjective views of the evaluators. Evaluators watch individual users working with an application and from that they identify usability problem. Validity is achieved through negotiation and comparison. In our study, there is considerable inter-subjective agreement both across the two analysis methods and between the evaluators that used VBA.

There is considerable activity on smart healthcare applications for home use. The focus is mostly on the functionality of such applications and their potential to improve the quality of life for patients and reduce healthcare expenses. Yet these potentials will never be realized unless the applications are usable. In our study, we evaluated the usability of application that solves a simple task of collecting data from a range of devices and sending them to a health monitoring centre. Nevertheless, we identified a total of 54 usability problem. Out of them, 18 problems were critical which means that the user was unable to complete the task without assistance. Even though the application seems simple, it failed to support the user, and that could have fatal consequences in real life use, for example if data about an ill patient were lost because of a user error. This illustrates a key challenge in developing smart healthcare applications for use by ordinary people. The best way to achieve a higher level of usability in such applications is to conduct formative usability evaluations during development and to use the results in improving the application. In this chapter, we have presented IDA as a modern method for conducting formative usability evaluations.

FUTURE RESEARCH DIRECTIONS

The experiences from our study emphasize several interesting directions for future work. It would be highly relevant to overcome the two main limitation of our study, i.e. the single case and the low number of users. Further studies could also involve longer term collaboration with developers of smart healthcare applications.

There are two technical difficulties that occurred in the use of the IDA method. First, it produces less detailed problem descriptions. It would be interesting to inquire into the way this level of detail influence the usefulness for developers and for drawing up redesign proposals. Second, the categorizations of the identified usability problems seem to be less reliable. The usability problems identified with IDA were initially given a more serious severity rating than the problems found with VBA. The problem with this is that prioritization of the problems to work with first becomes more difficult. It would be interesting to include work on a conceptual tool to support usability problem categorization, cf. Skov and Stage (2009).

The usability evaluation itself could also be the issue for further work. The evaluation with the think-aloud protocol is very obtrusive. It could be interesting to develop methods that are less obtrusive without sacrificing the quality of the results. Also, the laboratory evaluation with its carefully constructed procedures is not realistic. It would be interesting to inquire into evaluations in a more realistic setting and evaluations that involve more demanding user groups, e.g. patients suffering from cognitive decline.

The device we evaluated is a simple application. It is very relevant to work with less obtrusive applications. This relates to the broader issue of technology augmented homes and domestic computing, and the impact on the users, as well as the relationship between the home users and the healthcare personnel and facilities. The interesting lesson from our study is that even with a simple task scenario with use of a simple piece of technology a whole range of usability problems cropped up. This shows that usability is an unavoidable condition for increasing the use of smart healthcare applications in people's homes.

CONCLUSIONS

In this chapter we have presented the Instant Data Analysis (IDA) method that embodies a resource-economic approach to usability evaluation. We have also presented a case study where we evaluated the usability of a smart healthcare application using the IDA method and as a benchmark also the conventional approach to usability evaluation denoted as Video-Based Analysis (VBA). This case study suggests how to overcome the obstacle of high resource demands in usability evaluations, which is crucial because of the need for a high level of usability in smart healthcare applications.

The overall result is that IDA is very time-efficient. The effort required when using IDA is four to five times less than VBA, while the number of critical and serious problems identified is nearly the same. Based on this, we suggest that the IDA method is useful for formative usability evaluations in the development of smart healthcare applications.

The primary limitation of the study reported here is that IDA method has only been used by one group of evaluators, and the same applies to VBA. On the other hand, the study has been carried out in a realistic setting in collaboration with a software organization, and we have obtained results that are comparable to another study of IDA.

The study illustrates some of the challenges developers face in the design of smart healthcare applications. Even the simple task of setting up the application cased severe problems for all of the five users. The evaluation was conducted in a laboratory setting, while an evaluation of technology in a person's home would be of greater value. Yet development and usability evaluation of smart healthcare applications for the home is a relatively new area. Our aim has been to take a step in the right direction and provide a platform on which much more detailed development and evaluation activities can be performed.

REFERENCES

Bak, J. O., Nguyen, K., Risgaard, P. & Stage, J. (2008). Obstacles to Usability Evaluation in Practice: A Survey of Software Organizations. In A. Gulz, C. Magnusson, L. Malmborg, H. Eftring, B. Jönsson & K. Tollmar (Eds.), *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges: Vol. 358.* (pp. 23-32). New York: ACM Press.

Bruun, A., Gull, P., Hofmeister, L. and Stage, J. (2009) Let Your Users Do the Testing: A Comparison of Three Remote Asynchronous Usability Testing Methods. In D. R. Olsen, K. Hinckley, M. R. Morris, S. Hudson & S. Greenberg (Eds.), *Proceedings of the 27th international conference on Human factors in computing systems (CHI '09): Vol 1* (pp. 1619-1628). New York: ACM Press.

Eikerling, H-J., Gräfe, G., Röhr, F. & Schneider, W. (2009). Ambient Healthcare Systems - Using the Hydra Embedded Middleware for Implementing an Ambient Disease Management System. In L. Azevedo & A. R. Londral (Eds.), *Proceedings of the Second International Conference on*

Health Informatics (HEALTHINF 2009): Vol 2 (pp. 82-89). Funchal, Madeira, Portugal: INSTICC Press.

Fensli, R. & Boisen, E. (2008). Human Factors Affecting the Patient's Acceptance of Wireless Biomedical Sensors. In A. Fred, J. Filipe & H. Gamboa (Eds.), *Biomedical Engineering Systems and Technologies, International Joint Conference (BIOSTEC 2008): Vol 25* (pp. 402-412). Berlin: Springer Verlag.

Fisk, A. D. & Rogers, W. A. (2001). Health care of older adults: The promise of human factors research. In W. A. Rogers & A. D. Fisk (Eds.), *Human factors interventions for the health care of older adults* (pp. 1–12). Mahwah, NJ: Erlbaum.

Frøkjær, E., Hornbæk, K. (2008). Metaphors of Human Thinking for Usability Inspection and Design. *ACM Transactions on Computer-Human Interaction*, *14*(4).

Ginsburg, G. (2004). Human Factors Engineering: A Tool for Medical Device Evaluation on Hospital Procurement Decision-Making. In *Journal of Biomedical Informatics, 38*(3), 213-219. Elsevier.

Gray, W. D. & Salzman, M. C. (1998). Damaged merchandise? A review of experiments that compare usability evaluation methods. *Human-Computer Interaction, 13*(3), 203-261.

Hertzum, M. & Jacobsen, N. E. (2003). The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. *International Journal of Human Computer Interaction*, *15*(1), 183-204.

Hornbæk, K. & Frøkjær, E. (2004). Usability Inspection by Metaphors of Human Thinking Compared to Heuristic Evaluation. *International Journal of Human-Computer Interaction*, *17*(3), 357-374.

Jaana M. & Paré, G. (2006). Home Telemonitoring of Patients with Diabetes: A Systematic Assessment of Observed Effects. *Journal of Evaluation in Clinical Practice, 13* (2), 242-253.

Johnson, C. M., Johnson, T.R. & Zhang, J. (2004). A User-Centered Framework for Redesigning Health Care Interfaces. In *Journal of Biomedical Informatics, 38*(1), 75-87. Elsevier.

Karat, C.M., Campbell, R. & Fiegel, T. (1992). Comparison of empirical testing and walkthrough methods in user interface evaluation. In P. Bauersfeld, J. Bennett & G. Lynch (Eds.), *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '92),* (pp. 397-404). Monterey, CA: ACM Press.

Kaufman, D.R., Patel, V.L., Hilliman, C., Morin, P.C., Pevzner, J., Weinstock, R.S., Goland, R., Shea, S. & Starren, J. (2003). Usability in the Real World: Assessing Medical Information Technologies in Patient's Homes. *Journal of Biomedical Informatics, 36*(1-2), 45-60.

Kjeldskov, J., Skov, M. B. & Stage, J. (2004). Instant Data Analysis: Evaluating Usability in a Day. In A. Hyrskykari (Ed.), *Proceedings of the third Nordic conference on Human-computer interaction: Vol 82* (pp. 233-240). New York: ACM Press.

Kushniruk, A. W., Patel, V., Cimino, J. J. & Barrows, R. A. (1996). Cognitive Evaluation of the User Interface and Vocabulary of an Outpatient Information System. In *Proceedings of AMIA Annual Fall Symposium*, 22-26.

Kushniruk, A.W. & Patel, V. L. (2004). Cognitive and Usability Engineering Methods for the Evaluation of Clinical Information Systems. In *International Journal of Medical Informatics, 37*(1), 56-76. Elsevier.

Kushniruk, A.W., Triola, M.M., Borycki, E.M., Stein, B. & Kannry, J.L. (2005). Technology Induced Error and Usability: The Relationship Between Usability Problems and Prescription Errors When Using a Handheld Application. In *International Journal of Medical Informatics*, *77*(7-8), 519-526. Elsevier.

Linder, J. A., Rose, A.F., Palchuk, M.B., Chang, F., Schnipper, J.L., Chan, J.C. & Middleton, B. (2006). Decision support for acute problems: The role of the standardized patient in usability testing. In *Journal of Biomedical Informatics, 39*(6), 648-655. Elsevier.

McCarthy, J. C., Miles, V. C., Monk, A. F., Harrison, M. D., Dix, A. J. & Wright, P. C (1993). Text-based on-line conferencing: a conceptual and empirical analysis using a minimal prototype. *Human-Computer Interaction*, 8(2), 147-183.

Medlock, M. C., Wixon, D., Terrano, M., Romero, R. L. & Fulton, B. (2002). Using the RITE Method to Improve Products: A Definition and a Case Study. In *Usability Professionals Association*. Bloomingdale, IL: Usability Professionals Association.

Monk, A., Wright, P., Haber, J. & Davenport, L. (1993). *Improving Your Human-Computer Interface: A Practical Technique*. Upper Saddle River, NJ: Prentice Hall.

Molich, R. (2000) *User-Friendly Web Design* (in Danish). Copenhagen: Ingeniøren Books.

Nielsen, C. M., Overgaard, M., Pedersen, M.B., Stage, J. & Stenild, S. (2006). It's Worth the Hassle! The Added Value of Evaluating the Usability of Mobile Systems in the Field. In A. Mørch, K. Morgen, T. Bratteteig, G. Ghosh & D. Svanaes (Eds.), *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles: Vol 189* (pp. 272-280). New York, NY: ACM Press.

Nielsen, J. (1993). *Usability Engineering*. San Francisco, CA: Morgan Kaufmann.

Nielsen, J. (1994) Heuristic evaluation. In J. Nielsen & R. L. Mack (Eds.), *Usability Inspection Methods* (pp. 25-62). New York, NY: John Wiley & Sons.

Nielsen, J. (1997). Usability Testing. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics, Second Edition* (pp. 1543-1569). New York, NY: John Wiley & Sons.

Nielsen, J. & Landauer, TK (1993). A mathematical model of the finding of usability problems. In S. Ashlund, A. Henderson, E. Hollnagel, K. Mullet & T. White (Eds.), *Proceedings of the INTERCHI '93 conference on Human factors in computing systems,* (pp. 206-213). New York, NY: ACM Press.

Pascual, J., Sanz-Bobi, M.A. & Contreras, D. (2008). Intelligent System for Assisting Elderly People at Home. In L. Azevedo & A.R. Londral (Eds.), *Proceedings of the First International Conference on Health Informatics, (HEALTHINF 2008): Vol 1* (pp. 159-163). Funchal, Madeira, Portugal: INSTICC press.

Peleg, M., Schachak, A., Dongwen, W. & Karnieli, E. (2009). Using Multi-Perspective Methodologies to Study Users' Interactions with the Prototype Front End of a Guideline-Based Decision Support System for Diabetic Foot Care. In *International Journal of Medical Informatics, 78*(7), 482-493. Elsevier.

Peute, L.W. & Jaspers, M.W. (2007). The significance of a usability evaluation of an emerging laboratory order entry system. In *International Journal of Medical Informatics, 76*(2-3), 157-168. Elsevier.

Rubin, J. (1994). *Handbook of Usability Testing*. Toronto, Ontario: John Wiley & Sons.

Sasaki, J., Yamada, K., Tanaka, M. & Funyu, Y. (2009). Experiments of Life Monitoring Systems for Elderly People Living in Rural Areas. In L. Azevedo & A. R. Londral (Eds.), *Proceedings of the Second International Conference on Health Informatics (HEALTHINF 2009): Vol 2* (pp. 430-433). Funchal, Madeira, Portugal: INSTICC Press.

Sashima, A., Inoue, Y., Ikeda, T., Yamashita, T., Ohta, M. & Kurumatana, K. (2008). Toward Mobile Healthcare Services by Using Everyday Mobile Phones. In L. Azevedo & A.R. Londral (Eds.), *Proceedings of the First International Conference on Health Informatics, (HEALTHINF 2008): Vol 1* (pp. 242-245). Funchal, Madeira, Portugal: INSTICC press.

Skov, M. B. & Stage, J. (2009) A Conceptual Tool for Usability Problem Identification in Website Development. *International Journal of Information Technology and Web Engineering*, 4(4):22-35.

Souidene W., Istrate, D., Medjahed, H., Boudy, J., Baldinger, J.L., Belfeki, I & Delavault, F. (2009). Multi-Modal Platform for In-Home Healthcare Monitoring (EMUTEM). In L. Azevedo & A. R. Londral (Eds.), *Proceedings of the Second International Conference on Health Informatics (HEALTHINF 2009): Vol 2* (pp. 430-433). Funchal, Madeira, Portugal: INSTICC Press.

Taleb, T., Bottazzi, D., Guizani, M. & Nait-Charif, H. (2009). ANGELAH: A Framework for Assisting Elders at Home. *IEEE Journal on Selected Areas in Communications, 27*(4), 480-494.

United Nations (2006). World Population Prospects – The 2006 Revision. Department of Economic and Social Affairs, Population Division. Retrieved December 1st, 2009 from http://www.un.org/esa/population/publications/wpp2006/English.pdf

ADDITIONAL READING SECTION

Dumas, J. S. & Redish, J. C. (1999) *A Practical Guide to Usability Testing*. intellect.

Karat, C.M., Campbell, R. & Fiegel, T. (1992). Comparison of empirical testing and walkthrough methods in user interface evaluation. In P. Bauersfeld, J. Bennett & G. Lynch (Eds.), *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '92),* (pp. 397-404). Monterey, CA: ACM Press.

Kjeldskov, J., Skov, M. B. & Stage, J. (2004). Instant Data Analysis: Evaluating Usability in a Day. In A. Hyrskykari (Ed.), *Proceedings of the third Nordic conference on Human-computer interaction: Vol 82* (pp. 233-240). New York: ACM Press.

Nielsen, J. (1993). *Usability Engineering*. San Francisco, CA: Morgan Kaufmann.

Rubin, J. (1994). *Handbook of Usability Testing*. Toronto, Ontario: John Wiley & Sons.

KEY TERMS & DEFINITIONS

Usability: Property of an IT system or electronic device determining to which extent it is easy to learn, easy to use, effective to use and satisfying to use.

Usability evaluation: The process of assessing the level of usability.

Formative evaluation: Exploratory test to uncover usability issues in a yet unfinished user interface.

Summative evaluation: Test to review compliance usability of the final system to a set of requirements.

Instant Data Analysis (IDA): Method for analyzing the results of a usability test session without the use of video recordings.

Video Based Analysis (VBA): Classical method for analyzing results of usability test sessions using video recordings.

Evaluator effect: The name of the fact that evaluators identify different usability problems when evaluating the usability of the same user interface.