



CHALMERS

Chalmers Publication Library

Enhanced Cognitive Walkthrough: Development of the Cognitive Walkthrough Method to Better Predict, Identify, and Present Usability Problems

This document has been downloaded from Chalmers Publication Library (CPL). It is the author's version of a work that was accepted for publication in:

Advances in Human-Computer Interaction (ISSN: 1687-5893)

Citation for the published paper:

Bligård, L. ; Osvalder, A. (2013) "Enhanced Cognitive Walkthrough: Development of the Cognitive Walkthrough Method to Better Predict, Identify, and Present Usability Problems". Advances in Human-Computer Interaction, vol. 2013

Downloaded from: <http://publications.lib.chalmers.se/publication/185405>

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source. Please note that access to the published version might require a subscription.

Chalmers Publication Library (CPL) offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all types of publications: articles, dissertations, licentiate theses, masters theses, conference papers, reports etc. Since 2006 it is the official tool for Chalmers official publication statistics. To ensure that Chalmers research results are disseminated as widely as possible, an Open Access Policy has been adopted. The CPL service is administrated and maintained by Chalmers Library.

(article starts on next page)

Research Article

Enhanced Cognitive Walkthrough: Development of the Cognitive Walkthrough Method to Better Predict, Identify, and Present Usability Problems

Lars-Ola Bligård and Anna-Lisa Osvalder

*Division of Design and Human Factors, Department of Product and Production Development, Chalmers University of Technology,
412 96 Gothenburg, Sweden*

Correspondence should be addressed to Lars-Ola Bligård; lars-ola.bligard@chalmers.se

Received 11 October 2012; Accepted 12 August 2013

Academic Editor: Kerstin S. Eklundh

Copyright © 2013 L.-O. Bligård and A.-L. Osvalder. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To avoid use errors when handling medical equipment, it is important to develop products with a high degree of usability. This can be achieved by performing usability evaluations in the product development process to detect and mitigate potential usability problems. A commonly used method is cognitive walkthrough (CW), but this method shows three weaknesses: poor high-level perspective, insufficient categorisation of detected usability problems, and difficulties in overviewing the analytical results. This paper presents a further development of CW with the aim of overcoming its weaknesses. The new method is called enhanced cognitive walkthrough (ECW). ECW is a proactive analytical method for analysis of potential usability problems. The ECW method has been employed to evaluate user interface designs of medical equipment such as home-care ventilators, infusion pumps, dialysis machines, and insulin pumps. The method has proved capable of identifying several potential use problems in designs.

1. Introduction

In the development of user interfaces, it is important to consider the need for these to be simple and safe to handle for the user group in the intended context. This is especially true of safety-critical technical equipment such as medical equipment, where a possibility of harm to patients can arise from erroneous use of the devices [1–4]. Several studies have shown that there is a clear connection between problems of usability and human error; for example, Obradovich and Woods [5], Lin et al. [6], and the FDA [7] have also referred to this problem. Liljegren [2] has shown in a doctoral thesis that medical personnel rank “difficulty of making mistakes” as the most important aspect of good usability for medical equipment.

An important step in the development of usable technology is to try in advance to identify and evaluate the occasions, in the interaction between user and product, when there is a possibility of errors arising [8]. To identify the problems that can give rise to errors in handling a product, evaluations are

normally made of the product's user interface with realistic tasks, that is, in a usability evaluation. Jaspers [9] presents an overview of methods used in medical technology usability evaluations.

The usability evaluation of user interfaces can proceed according to two different approaches: empirical and analytical [3, 10]. Empirical evaluation involves studies of users who interact with the user interface by carrying out different tasks, which is done in what are known as usability tests [11]. Usability tests have been employed to study the usability of medical equipment, such as infusion pumps [12] and clinical information systems [13].

In an analytical evaluation, no users are present as test subjects, and the evaluation of the interface is made by one or more analysts using theoretical models, such as heuristic evaluation [8, 14]. Heuristic evaluation of medical equipment has been done, for instance, with infusion pumps by [15].

An often used analytical method of usability evaluation is cognitive walkthrough (CW) [16–18]. CW is an inspection method for evaluating usability in a user interface.

The method focuses on simplicity in learning, especially through exploratory learning. CW has been employed to evaluate medical equipment, such as clinical information systems, patient information systems, clinical order systems, dialysis machines, and patient surveillance systems [1, 13, 19–22]. An advantage in using CW in healthcare is that the method can be used to identify important usability problems quite easily, quickly, and cheaply when resources for performing real usability tests are limited. Usability tests demand highly skilled usability professionals, end users to test whose time is often hard to book, and plenty of time and effort to perform at high cost [9]. Since CW is also a task-based method and the evaluation follows a linear path [23], it is capable of detecting a greater number of usability problems than in usability tests, where the number of evaluated tasks is lower for time and cost reasons.

However, a limitation of CW is that it focuses mainly on ease of learning; that is, it assesses whether the equipment is simple to use without any previous knowledge. Therefore, the CW results usually need to be complemented by other methods such as heuristic evaluation or usability tests [18]. The users' domain knowledge about the task to be solved is, however, taken into account when evaluating with CW.

A number of other weaknesses have been found in CW. One of the more prominent weaknesses is the emphasis on low-level details [18, 24, 25]. This means that the method detects problems at a detailed level such as the marking of buttons but misses problems of a more general and conceptual nature, for instance, the choice of menu structure and sequences in the user interface. Thus, CW entails a deficient high-level perspective in the evaluation.

The purpose of this paper is to describe and discuss a refined version of the CW method, designated enhanced cognitive walkthrough (ECW). ECW is a method useful in the practical engineering work area and is part of human factors engineering [26] and product development [27] in the industry. ECW provides a more extensive presentation of analysis and results than CW does. ECW was developed especially for the evaluation of medical equipment, but its use is not limited to this application area. As an example, a case study of a fictitious medical equipment interface (a home-care ventilator) will be used in this description.

2. Original Methods

2.1. Hierarchical Task Analysis. In order to make an analytical evaluation of a user interface with CW, knowledge is needed about the task that is to be performed with the aid of the interface. One method employed for analysing work tasks is hierarchical task analysis (HTA) [28]. This method breaks a task down into elements or subgoals [29]. These become ever more detailed as the hierarchy is divided into ever smaller sub-tasks (progressive re-description). The division continues until a stop-criterion is fulfilled, often when the subtask consists of only a single operation.

HTA describes, then, how the overall goal of the work-task can be attained through subgoals and plans. The results are usually presented in a hierarchical tree diagram. HTA is

also used as a basis for other analytical methods in interface design, such as the systematic human error reduction and prediction approach [30] and task analysis for error identification [31].

2.2. History and Description of Cognitive Walkthrough (CW). Cognitive walkthrough is a method for theory-based evaluation of usability in interfaces, developed by Lewis and Wharton [18]. It is employed to identify problems and generate proposals about their causes. The method is primarily intended for application during the process of designing user interfaces. Unlike many other evaluation methods for usability, such as heuristic evaluation, the focus in CW is on the user's cognitive processes and previous knowledge.

CW concentrates on ease of learning by exploration and is based on theories about explorative learning presented by Lewis et al. [16]. Explorative learning means here that a user tries to perform a task by a "trial and error" technique. The method simulates the user's cognitive processes when he/she carries out a sequence of actions in performing a given task. CW determines whether the user's background knowledge, together with hints from the interface, will lead to a correct sequence of goals and actions.

The analysis is best conducted by an evaluator or a group of evaluators, who may consist of designers, software developers, marketing personnel, future users, and persons with expertise in ergonomics and human factors engineering. Since the method simulates a user's thoughts and behaviour, it requires that the person(s) performing the analysis has or have sufficient knowledge about the users. The quality of the results from the analysis depends on how well the evaluators can place themselves in the situation as users.

CW was originally developed in order to bring cognitive theory closer to practical design development and evaluation of user interfaces [24]. The original formulation of CW was intended to evaluate noncomplex interfaces (walk-up-and-use interfaces), meaning that the user is not expected to have any previous knowledge of the technical system in the situation of use. Examples of such interfaces are cash dispensers and coffee machines. CW is also employed to evaluate user interfaces of a more complex character, such as graphic interfaces for Unix operating systems or sales support systems, and research CAD tools, as described by Wharton et al. [24]. Moreover, CW is considered to be usable by software developers without specialised knowledge in cognitive theory and human-computer interaction [32].

The first version of the CW method was presented by Lewis et al. [16]. It has since been refined into a second version by Polson et al. [17]. Today a third version exists, described by Wharton et al. [33] and Lewis and Wharton [18]. Initially the method was developed in order to discover, improve, and counteract defects and problems in user interfaces. The second version of CW was created to make the procedure more formal and detailed, since the initial method presupposed that the evaluators had knowledge of cognitive theory. In addition, the questions in the method were too general. However, the second version became too complex, difficult to apply, and time demanding. This led to the third version of CW by Lewis and Wharton [18], which is simpler and more

effective than its predecessors. The method was consequently also more accessible for usability practitioners.

The third version of CW comprises three steps: preparation, analysis, and followup. The preparatory step deals with identifying and defining the users, choosing tasks to evaluate, determining the correct sequence of actions for these tasks (e.g., with the aid of HTA), and finding out how the user interface presents information during these sequences. In the second stage, the analysis, a walkthrough is conducted of the chosen tasks, where the evaluators pose four questions for each stage in the sequence of action. The questions are aids to simulating the user's cognitive process.

- (1) Will the user be trying to achieve the right effect?
- (2) Will the user discover that the correct action is available?
- (3) Will the user associate the correct action with the desired effect?
- (4) If the correct action is performed, will the user see that progress is being made?

The questions are answered with YES or NO and reasons why the user will succeed or fail in performing the action ("failure/success stories"). Problems that arise are noted, along with the reason why the problems arose, based on assumptions made by the evaluators. In the last step, followup, proposals are given for how the user interface can be changed so as to eliminate the discovered problems.

2.3. Weaknesses of Cognitive Walkthrough. Even though CW has been refined in a couple of steps, there are still some weaknesses in the method. Three weaknesses have been identified and are thoroughly described in a Master's thesis from Chalmers University of Technology [34].

The first of these is connected with the CW method's lack of high-level perspective. This was pointed out by Bligård and Wass [34] in a heuristic evaluation of a number of user interfaces. Several usability problems were then identified that the analysis with CW itself missed. Wharton et al. [24] also drew attention to this weakness. CW does not answer the question of whether the user knew that an entire function was available or whether the interface provided hints that it was. Thus, CW focuses on the sequence of operations that are needed to carry out a task but does not evaluate the function itself. If the user does not know that a function (a sequence of operations) is available or if the user interface provides no hints that it exists, the user will not be able to carry out the correct sequence of operations, even if the sequence is simple and intuitive. Although the low-level analysis is usually followed by a more high-level perspective when the analyst searches the result for pattern and strategies in the interaction, the CW methods do not help the analyst to achieve the higher levels of perspective.

The next weakness of CW is that the failure-and-success stories provide insufficient information about the difference in the seriousness of problems between different operations. The answers can only be classified as either success or failure, which are regarded as insufficient reply options. It is difficult

to classify the answer as either 100% success or 100% failure. Further, there is a distinction between whether the detected problem concerns an important intended or safety-critical function or whether it involves only a less important operation. Neither is any categorisation made of types of detected usability problems.

The third weakness of CW has to do with the presentation of the results of the evaluation. When the question process in CW is completed, it is hard to obtain a clear and general overview of the results from the analysis. This is true both for a particular user interface and when comparing different interfaces. For example, it is hard to read which are the greatest problems in the design of the interface, which part of the interface the problems are in, or whether interface A has fewer problems than interface B.

To sum up, the weaknesses of CW are as follows.

- (i) CW has a deficient high-level perspective in the evaluation of user interfaces, which was thus manifested.

CW does not answer the question whether the user knew that the function concerned was available.

CW does not answer the question whether the interface provided hints that enabled the user to discover more easily that the function was available.

- (ii) The explanations for success or failure yield insufficient information about the difference in problem severity between distinct operations.
- (iii) It is difficult to obtain an overview of the results, both within a user interface and in comparison between different interfaces.

3. Enhanced Cognitive Walkthrough

The aim in developing ECW was to try to counteract the deficiencies in the third version of CW. The goal was to develop a method that can better detect and identify given presumptive usability problems in an interface and also provide an overview of which types of problems exist and how serious these are.

To attain the goal, three additions to CW were made: (1) a division into two levels of questions, in order to investigate functions and not only operations; (2) grading of tasks, grading of answers for success and failure, respectively, and categorisation of these answers in types of problems; (3) presentation of the results in matrices for a better overview and a possibility of comparison between different interfaces.

3.1. Description of ECW. Enhanced cognitive walkthrough (ECW) is a method of inspection based on the third version of CW, presented by Lewis and Wharton [18]. ECW uses a detailed procedure to simulate the user's problem-solving process in each step of the interaction between user and interface. It is continually checked whether the user's goal and knowledge can lead to the next action being correctly executed.

The method is conducted by an evaluator or a group of evaluators, who may consist of designers, software developers, marketing personnel, presumptive users, and persons with knowledge of ergonomics and human factors engineering. Most important is that real users and/or knowledge about use and the users must be present among those who carry out the analysis.

ECW comprises three parts: preparation, analysis, and presentation of the results in matrices. Before the evaluation begins, however, the artefact (product or technical system to be analysed) and the intended use and the user must be identified and described.

To exemplify the working procedure in the different steps of ECW, a case study is employed, on a fictitious home-care ventilator of the CPAP (continuous positive airway pressure) type. The method is presented in steps, and a general description is first given of how the evaluation is performed for each step, followed by examples from the case study in order to illustrate how the method works more specifically. The case is written as it would be an existing machine when refereeing to, for example, user and manual. The fictitious CPAP is based on the existing home-care ventilator BREAS PV10 [34] which have been moderated to fit as a case here. Finally there are reflections on method development.

3.2. *Intended Use and Users*

3.2.1. Method Description. Before the ECW analysis is begun, it is determined which artefact is to be evaluated and what its intended use is. This applies to both the artefact's main task and other tasks that it can perform. Moreover, the intended users must be defined.

3.2.2. *Case Study of a Home-Care Ventilator*

(1) Analysed Artefact. A home-care ventilator of the CPAP type is used to counteract sleep apnoea. This condition means that a person periodically stops breathing while sleeping, often because the airways collapse. When the body discovers this, the patient's degree of consciousness is raised so that breathing resumes but not so much that the patient wakes up. As a result, the patient sleeps uneasily at night and does not rest enough. The CPAP is basically an apparatus with a fan that creates overpressure. The patient is connected to the ventilator with a mask and a tube. The overpressure helps to open the patient's airways and thus counteract the apnoea.

(2) Intended Use. When a patient is to be treated with a CPAP at home, a tryout is first conducted at a hospital. There, the CPAP is adjusted so that it delivers the right pressure and counteracts the apnoea. Then the patient is sent home with the ventilator. The case study deals with the use of the home-care ventilator which takes place during the tryout.

(3) Intended Users. The staff at the hospital who handle the home-care ventilator during the tryout are specially trained nurses and physiotherapists. Their main task at the hospital is to diagnose and treat sleep-related ailments.

3.2.3. Reflection. In summary, it can be said that the choices of artefact, use, and user that are made before the ECW analysis begins are decisive for the quality of the coming analysis. If the choice of intended use or user is inadequately or wrongly made in relation to the actual conditions, the entire evaluation of usability in the artefact will be erroneous. ECW as a method is highly dependent on the input data that the analysis is made upon.

3.3. Preparation. The phase of preparation for ECW consists of four steps: (1) selection and grading of tasks for evaluation, (2) specification of these tasks, (3) specification of the artefact's user interface, and (4) specification of the user and use situation. These steps are described below in a sequence, but they should be taken in parallel and jointly during the preparatory phase. It is thus not necessary for the specification of tasks to be fully completed before the specification of the artefact's user interface is begun.

3.3.1. *Selection and Grading of Tasks for Evaluation*

(1) Method Description. Depending on the aim and goal of the study, the first step is to define which tasks should be evaluated. It is important to choose realistic tasks for the evaluation, including tasks that are carried out often as well as tasks that are safety critical and carried out more rarely. The selection of tasks is a very important part of the ECW method and merits careful attention, since it is not often feasible to evaluate all tasks that can be performed with a device. It is impossible to give strict general criteria to which tasks to select, since it strongly depends on the specific human-machine system, but it is often good to evaluate tasks that are critical for the intended use, tasks that are frequent in use, and tasks that are hazardous for the user or the environment.

The selection of tasks must be based on the intended use not on the design or function of the equipment. Each selected task should be given a unique number, known as the task number. Each task is graded from 1 to 5, the "task importance." The grading is based on how important the task is in the intended use of the artefact. The most important tasks are graded 1 and the least important 5. It is important that the tasks which are selected for comparison should have the same task importance for all user interfaces if a comparison between different user interfaces of an artefact is to be performed.

(2) Case Study of a Home-Care Ventilator. With reference to the intended use and the manual, ten ordinary tasks were noted that can be performed on the home-care ventilator. The tasks were then graded according to how important they are for being able to use the ventilator (Table 1). The selection and grading were made in collaboration with two users.

(3) Reflection. To sum up, the tasks that are chosen for analysis have a great impact on the quality of the results. If the wrong tasks are selected, usability will not be evaluated against the correct situation of use. The selection and grading of tasks should be done in cooperation with the intended users or with persons who have knowledge about the users and use. A

TABLE 1: Grading of the tasks performed by the home-care ventilator.

	Task	Grade
1	Start treatment	1
2	Set actual time	4
3	Set wake-up alarm	4
4	Set pressure	2
5	Adjustment of what to show in the display	3
6	Show software version	5
7	Show operating hours	4
8	Lock/unlock the panel	2
9	Use the ramp function	2
10	Turn off treatment	1

correct and exhaustive description of the user is also essential for methods such as CW and ECW, as has been stressed by Liu [35].

3.3.2. Specification of the Tasks

(1) *Method Description.* The next step in the preparations is to determine the intended, correct way in which the task should be performed with the aid of the user interface. Since ECW involves a hierarchic approach in analysis, it is appropriate to describe the task in a tree diagram by employing hierarchical task analysis (HTA) [28]. The bottom level in HTA consists of the individual steps (actions) in the interaction between user and interface, which are termed operations. The tasks and subtasks above this level are called nodes. A node together with the nodes and operations below is known as a function (Figure 1).

The nodes and operations in the HTA must be numbered uniquely, in order to facilitate the compilation of results from the analysis. For example, a function can be designated according to the uppermost node. When making comparisons between different interfaces, the design of the HTA will be different for each interface. This is not a problem for the coming analysis, as long as the subdivision and grading of tasks are the same for all the interfaces. But if they are not the same, it is impossible to make a relevant comparison.

(2) *Case Study of the Home-Care Ventilator.* The selected tasks were described by an HTA to obtain the structure before the ECW analysis. The description was made on the basis of how the ventilator must be handled according to the manual. Thus, the evaluation is made in terms of the handling technique that the manufacturer has envisaged. Figures 2 and 3 show the HTA diagram for task 3 (set wake-up alarm) and task 4 (set pressure).

In the subsequent analysis, task 3 will be employed as an example. This task is not central to the intended use but is chosen as a good instance of how the method works. Task 3 is divided into five operations. The node and the operations describe the different goals that the user has in handling. Together, the node and operations constitute a function that will be evaluated in the analysis.

(3) *Reflection.* What is lacking in the classic CW method, then, is an analysis of the user interface on a higher level that

lies above the direct operations performed by the user, that is, an analysis which is oriented more towards tasks or sequences of operations (functions). To make an analysis with several levels, an HTA diagram is therefore needed as a basis. HTA diagrams have also been used in connection with traditional CW analysis [2], but then only for analysing the operations (actions) at the bottom task level of the HTA. ECW also analyses the upper task levels in the HTA diagram.

When describing the task with HTA it is worth noting that a task can be performed in several different ways to reach the same goal. Moreover, the grouping of operations and functions may also vary for the same task. It is therefore important that the design of the HTA diagram corresponds as closely as possible to the sequences of action that occur in reality. This is especially important if different products or technical systems shall be compared after the analysis. In the fictive case here no plans are used in the HTA to make them simpler to understand.

3.3.3. Specification of the User Interface

(1) *Method Description.* The HTA diagrams describe the correct way in which the tasks are to be performed. A specification is then made of how the interface looks for the different operations. In this way it is possible to evaluate the user interface against each task.

(2) *Case Study of the Home-Care Ventilator.* The user interface on the home-care ventilator (Figures 4 and 5) consists of a display, five buttons, and six LEDs (light-emitting diodes), two of which are located in buttons. The display has four seven-segment signs, marking for am/pm, and a symbol showing that the alarm clock is activated.

For each task that is to be evaluated, both the user's actions (Figure 2) and the interface's response for the respective handling in a sequence are specified. Figure 6 presents the sequence for how task 3, "set wake-up alarm", is performed. The sequence begins with presentation of the interface's appearance before the first action.

(3) *Reflection.* The specification of the user interface must be done at the level of detail that is required for the coming analysis. Which level of detail is needed cannot be stated in general; it is up to the evaluator to decide. But each operation has to be described in such a way that its usability can be evaluated. Both the user's actions and the artefact's responses to these must be specified in equal detail.

3.3.4. Specification of the User and Use

(1) *Method Description.* The last step in the preparations is to define what knowledge and experience of the artefact and its use are possessed by the expected user. Further, the context in which the artefact operates is described. The context concerns the physical, organizational, and psychosocial environment during use.

(2) *Case Study of the Home-Care Ventilator.* In the case of home-care ventilators, interviews were held with the users in

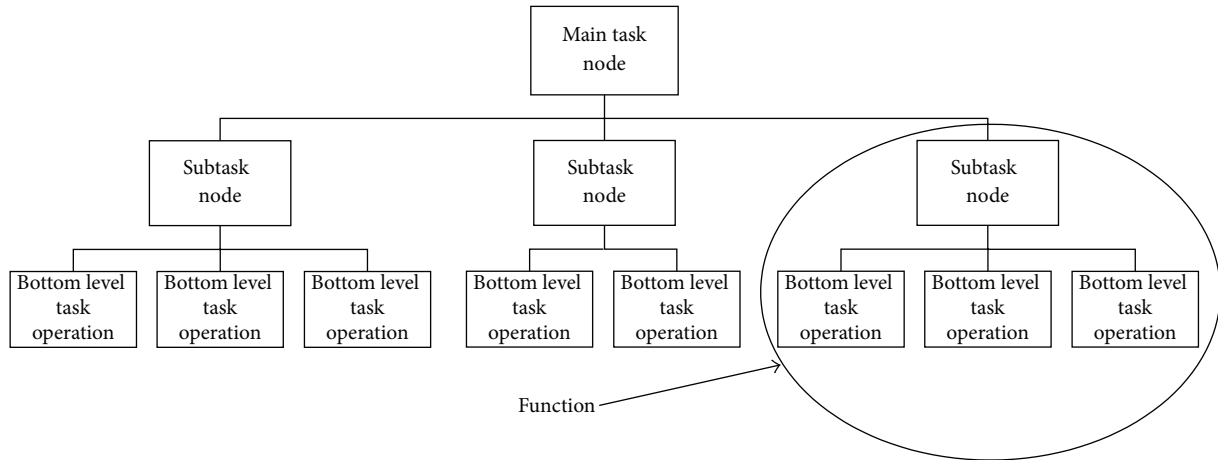


FIGURE 1: HTA diagram with nodes, operations and functions.

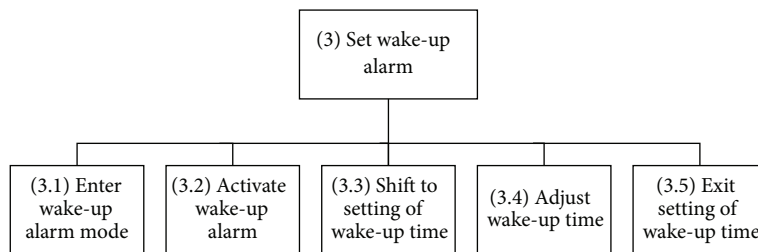


FIGURE 2: The correct handling sequence for the task “set wake-up alarm” presented in an HTA diagram.

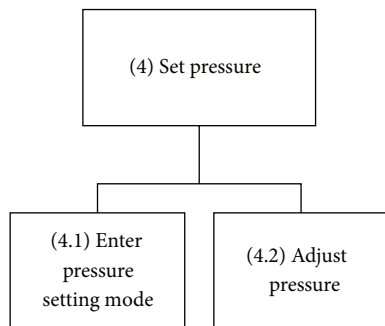


FIGURE 3: The correct handling sequence for the task “set pressure” presented in an HTA diagram.

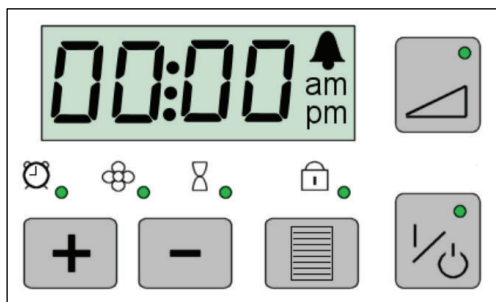


FIGURE 4: The user interface of the home-care ventilator.

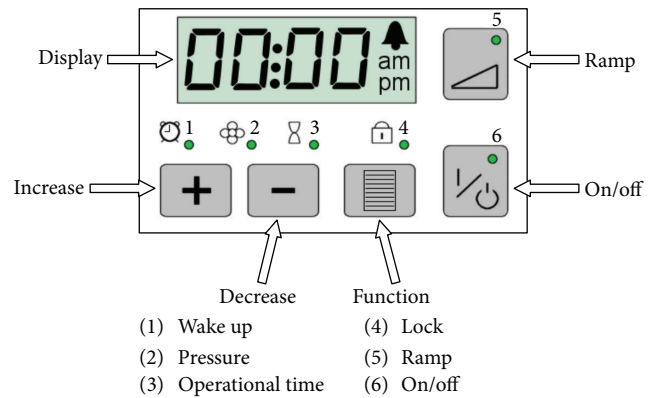


FIGURE 5: The user interface of the home-care ventilator with explanations.

order to learn their background and knowledge about such ventilators. Observations were also made in order to study use. The results can be summarised briefly as follows.

The users of home-care ventilators at the hospitals are nurses and physiotherapists with special training in the area. At hospitals, use takes place during what are known as tryouts, in which a patient is connected to a ventilator with a mask and tube. The patient sleeps during the tryout, and

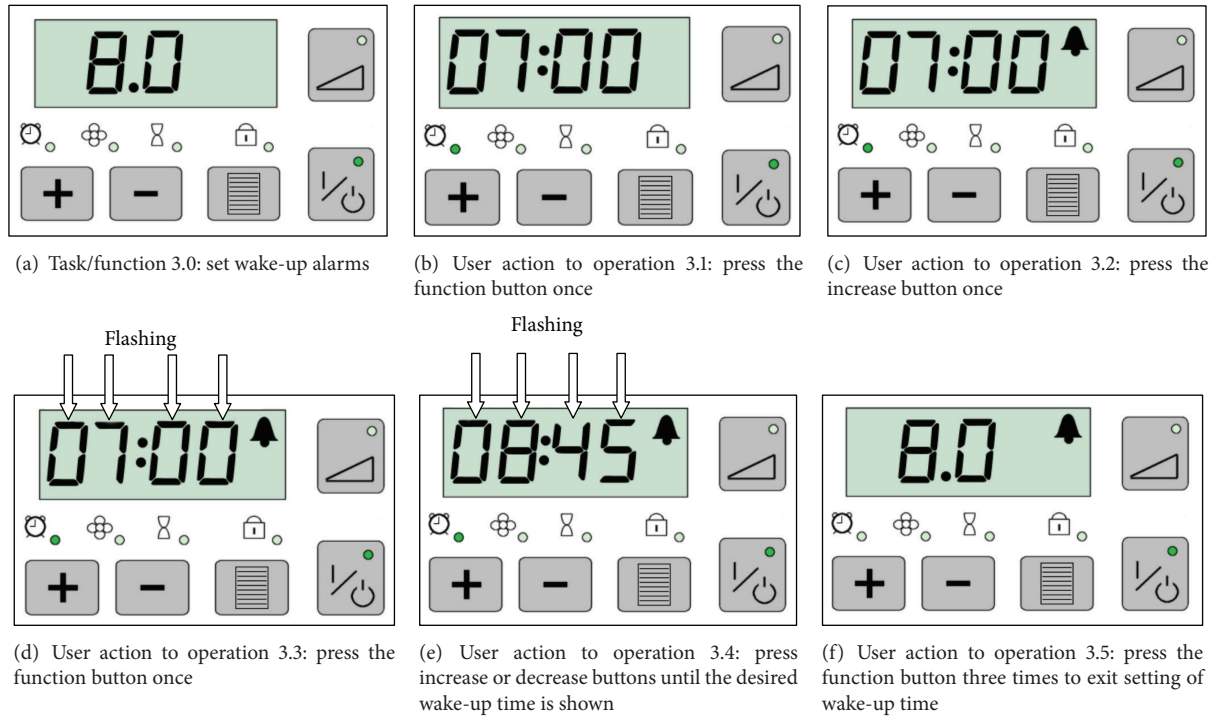


FIGURE 6: Performed user actions and the response of the user interface for the task “set wake-up alarm”.

the user adjusts the pressure that the ventilator delivers to the patient, while the oxygenation of the patient’s blood is monitored. The tryout takes place in a room with lowered lighting. The aim is that the patient should receive full oxygenation during sleep, while the ventilator delivers as little pressure as possible. The intended user makes, on average, one tryout per week with the evaluated home-care ventilator. The alarm clock’s role in the tryout is to waken the patient when the tryout is over.

(3) *Reflection.* The quality of the specification of the users’ background and knowledge is decisive for the validity of the results. This is because the coming analysis builds upon assumptions of what the user thinks and does. Moreover, the context of use should be defined. Knowledge about the context enables the evaluator to take it into account in the analysis. An important part of the preparation for ECW, therefore, is to create a sufficient profile of the user and the context.

3.4. *Analysis.* The analysis is based on the described correct handling sequences in the HTA. The evaluator(s) works or walk through the question process for all selected tasks, and conceivable use problems are then generated, which finally are graded and categorised. To document the analysis and the usability problems, a protocol has been developed (Table 7).

3.4.1. Method Description

(1) *Prediction of Usability Problems with the Aid of a Question Process.* The question process is divided into two levels of

questions. The first (level 1) is employed for the nodes in the HTA and the second (level 2) for the operations in the HTA. In level 1, the interface’s ability to “capture” the user is studied, and in level 2 its ability to lead the user to perform the function correctly is studied.

Analysis Questions for ECW

Level 1: Analysis of functions.

- (1) Will the user know that the evaluated function is available?
Does the user expect, on the basis of previously given indications that the function exists in the machine?
- (2) Will the user be able to notice that the function is available?
Does the machine give clues that show that the function exists?
- (3) Will the user associate the clues with the function?
Can the user’s expectations and the machine’s indications coincide?
- (4) Will the user get sufficient feedback when using the function?
Does the machine give information that the function has been chosen and to what position the user is in the interaction?
- (5) Will the user get sufficient feedback to understand that the function has been fully performed?

Does the user understand, after the performed sequence of actions, that the right function has been performed?

Level 2: Analysis of operations.

- (1) Will the user try to achieve the right goals of the operation?

Does the user expect, on the basis of previously given indications, what is to be performed?

- (2) Will the user be able to notice that the action of the operation is available?

Does the machine give clues that show that the action is available and how to perform it?

- (3) Will the user associate the action of the operation with the right goal of the operation?

Can the user's assumed operation and the machine's indications coincide?

- (4) Will the user be able to perform the correct action?

Does the abilities of the user match the demands by the machine?

- (5) Will the user get sufficient feedback to understand that the action is performed and the goal is achieved?

Does the user understand, after the performed operation, that he/she has done right?

The analysis begins with the evaluator asking the questions on level 1 for the uppermost node in the HTA diagram (Figure 1). Then the analysis continues downward through the HTA diagram, where the evaluator employs questions at level 1 for the nodes and questions at level 2 for the operations farthest down in the tree. The underlying nodes/operations of a given node are analysed completely before the analysis proceeds to the adjacent node.

(2) *Grading of the Answers.* Each question is answered and graded with a number between 1 and 5 together with a justification for the grade. The grading represents different levels of success (Table 2). The justifications are called failure/success stories. The failure/success story describes the assumptions underlying the choice of grades, for example, that the user cannot understand a text message or a symbol.

The grade ranks the different problems found in the interface, that is, the problem seriousness. This type of grading makes it easier to determine what is most important to rectify in the subsequent reworking of the interface. During the analysis, each question is answered—assuming that the preceding questions are answered YES (grade 5)—independently of what the real answer was for the last question. However, in certain cases, the questions may be impossible to answer. These questions must be marked with a dash in the protocol.

(3) *Problem Identification.* If the problem seriousness is between 1 and 4, this suggests the existence of a supposed usability problem. The usability problem is then described based on the failure/success story. The usability problem is

TABLE 2: Grading of the failure/success stories.

Grade	Grade in words	Explanation
5	Yes	A very good chance of success
4	Yes, probably	Probably successful
3	Do not know	Impossible to decide if success or not
2	No, uncertain	Small chance of success
1	No	A very small chance of success

TABLE 3: Examples of problem types.

Problem type	Explication	Origin
User (U)	The problem is due to the user's experience and knowledge, possibly because the user is accustomed to different equipment	Comes primarily from questions 1 and 3
Hidden (H)	The interface gives no indications that the function is available or how it should be used	Comes primarily from question 2
Text and icon (T)	Placement, appearance and content can easily be misinterpreted or not understood	Comes primarily from question 3
Sequence (S)	Functions and operations have to be performed in an unnatural sequence	Comes primarily from question 1
Physical demands (P)	The interface sets too high demands on users' physical speed, motoric skill and force	Comes primarily from question 4 (operation level)
Feedback (F)	The interface gives unclear indications of what the user is doing or has done	Comes primarily from question 4 (function level) and question 5

the factor which restrains the user from performing the correct action.

(4) *Problem Categorisation.* Each problem is then further categorised with a problem type. This is done with the aid of the description of the problem and the failure stories. Depending on the user interface and the task that the user is to solve with the artefact, different problem types can be defined. Suggestions of problem types are described in Table 3.

(5) *Case Study of the Home-Care Ventilator.* The analysis of the tasks from Table 1 was done with ECW's two question levels, and the answers were graded with problem seriousness on the basis of the motivations in failure/success stories. Each detected usability problem (grade 1–4) was categorised in terms of a problem type.

In the analysis of task 3 (set wake-up alarm), several usability problems were detected and identified. The analysis results are reported in a table, where the answers to the questions (failure/success stories), the usability problem, the problem seriousness (PS), and problem type (PT) are summarised (see Table 4). First in the table come the answers

TABLE 4: The ECW analysis template.

(a)				
3.0	Set wake-up alarm Failure/success story	Set wake-up alarm Usability problem	PS	PT
(1)	Do not know. It depends on whether the user has read the manual or has the expectation of being able to execute this setting.	User does not expect functionality	3	U
(2)	Yes, there is an icon with LED in the user interface.	No usability problem	5	—
(3)	Yes, probably. The icon has the appearance of a classic wake-up bell.	Unclear icon	4	T
(4)	Yes, the LED for wake-up alarm is lit.	No usability problem	5	—
(5)	Yes, the symbol for wake-up in the display appears or disappears.	No usability problem	5	—
(b)				
3.1	Enter wake-up alarm mode	Press the function button once	PS	PT
(1)	Yes, probably. The users know that they have to enter a mode to be able to set the wake-up alarm.	User does not expect action	4	U
(2)	Yes. The function button is specifically marked and accessible.	No usability problem	5	—
(3)	Yes, probably. The function button is marked with a symbol for circular change, and the users can use the method of elimination.	The function button can be hard to interpret	4	T
(4)	Yes, the users have motoric precision and force to press the button.	No usability problem	5	—
(5)	Yes, the lit LEDs change and the wake-up time is shown in the display.	No usability problem	5	—
(c)				
3.2	Activate wake-up alarm	Press the increase button once	PS	PT
(1)	Do not know. It depends on the user's mental model of which order the users expect the sequence of action to be in.	User does not expect action	3	S
(2)	Yes. The increase button is specifically marked and accessible.	No usability problem	5	—
(3)	Yes. This is the only button that is plausible to use to activate the wake-up alarm.	No usability problem	5	—
(4)	Yes, the users have motoric precision and force to press the button.	No usability problem	5	—
(5)	Yes, the symbol for activated wake-up alarm becomes visible in the display.	No usability problem	5	—
(d)				
3.3	Shift to setting of wake-up time	Press the function button once	PS	PT
(1)	Yes, probably. The users have the goal of setting the wake-up time.	User does not expect action	4	S
(2)	Yes. The function button is specifically marked and accessible.	No usability problem	5	—
(3)	No, uncertain. The users may not realise that they must press the function button to be able to set the wake-up time.	Unclear marking of correct action	2	T
(4)	Yes, the users have motoric precision and force to press the button.	No usability problem	5	—
(5)	Yes, probably. The wake-up time starts to flash.	Unclear feedback	4	F
(e)				
3.4	Adjust wake-up time	Press increase or decrease buttons until the desired wake-up time is shown	PS	PT
(1)	Yes. The users have the goal of setting the wake-up time.	No usability problem	5	—
(2)	Yes, the increase and decrease buttons are specifically marked and accessible.	No usability problem	5	—
(3)	Yes, plus and minus are well-recognised symbols for increasing and decreasing parameters.	No usability problem	5	—
(4)	Yes, the users have motoric precision and force to press the button.	No usability problem	5	—
(5)	Yes, probably. The wake-up time shown in the display changes concurrently with the button pressings. However, the pressing changes time to minutes and the hours follow.	Unclear feedback	4	F
(f)				
3.5	Exit setting of wake-up time	Press the function button three times to exit setting of wake-up time	PS	PT
(1)	Yes. It is natural to exit the setting mode.	No usability problem	5	—
(2)	Yes. The function button is specifically marked and accessible.	No usability problem	5	—
(3)	Yes. The function button is the logical selection to go further.	No usability problem	5	—
(4)	Yes, the user has motoric precision and force to press the button.	No usability problem	5	—
(5)	Do not know. The time stops flashing, but the interface shows the set pressure directly.	Unclear feedback	3	F

to the questions for the overall function (node 3.0 of the HTA in Figure 2), answered in relation to the knowledge about the user, the use, and the description of the interface (Figures 4 and 5). Next come the answers to the questions for each step in the interaction (operations 3.1 to 3.4 of the HTA in Figure 2). These questions are answered in relation to the steps of the changes in the interface during the interaction (Figure 6).

(6) *Reflection*. In order to utilise the division between operations and nodes in the analysis of the interface, the question process in ECW has been divided into two levels, as stated: level 1 for nodes and level 2 for operations. Level 1 studies the interface's ability to alert the user to a function's availability and use. The five questions for this level are designed on the basis of the four questions that CW employs but now with a focus on the functions and sequences of operations. Level 2 studies the interface's ability to lead the user to perform the operations correctly. Here nearly the same four questions as in CW are employed with an addition of a question regarding the user ability to perform the action. The user may be fully aware of the correct action but not able to perform it, for example, due to physical impairments.

The introduction of grades and categorisations in ECW makes it easier to rank the different problems that the interface exhibits. Thus it becomes easier to determine what is most important to rectify in the subsequence redesign of the interface or in comparison with other interfaces.

The first grading is done on the basis of how important the task is for the intended use of the artefact. Important tasks are often those that are performed frequently or which may have serious consequences if not done in a correct way: what are known as critical functions.

The next grading is done on the basis of the questions in the question process. CW has only two levels of answers, failure or success. To distinguish better between different levels of success or failure, ECW employs five grades of problem seriousness (Table 1).

The problems that are detected, that is, when the conceivable success is not complete, are then categorised in terms of the problem's cause. Such a problem type may be, for example, that the sequence is illogical or that the interface provides insufficient feedback.

3.5. Compilation in Matrices

3.5.1. Method Description. Matrices are used for presenting the results from the analysis part of ECW. The information gathered from the question process is arranged in different ways in the matrices so as to emphasise different aspects of the analysis. The information that is utilised from each failure's motivation consists of the task number, task importance, problem seriousness, and problem type. The matrices can be combined in several ways. Five distinct proposals of matrices that can be employed for presenting the ECW analysis are shown in Table 5.

An example of Matrix A is shown in Figure 7. The figures in the matrix cells show the detected problems distributed according to the two types of data that are being compared.

TABLE 5: Matrixes for presenting the results from the ECW analysis.

Name	Content	Explanation
Matrix A	Problem seriousness versus task importance	This shows the interface's general condition.
Matrix B	Problem seriousness versus problem type	This shows the overall problems with the interface.
Matrix C	Problem type versus task importance	This shows which problems are most important to rectify.
Matrix D	Problem seriousness versus task number	This shows which tasks have the most problems.
Matrix E	Problem type versus task number	This shows which types of problems are most common in the tasks.

Task importance	Problem seriousness			
	1	2	3	4
1	0	0	0	1
2	0	1	1	8
3	2	2	8	1
4	1	2	3	5
5	1	0	0	0

FIGURE 7: Example of Matrix A—problem seriousness versus task importance.

For instance, this matrix shows how many problems occur with each specific combination of problem seriousness and task importance. As an example, there are 3 problems with task importance 4 and problem seriousness 3 (the cell with diagonal lines).

The five different matrices in Table 5 describe the picture of problems with the interface in different ways. The numbers in the matrices show how many problems exist in the specific combination of analytical results about the entire problem complex. Since the matrices only exhibit the same problem complex in different ways, the sum of the numbers is the same in all matrices belonging to a given interface.

Matrix A (problem seriousness versus task importance) shows whether there are serious usability problems with the interface that can prevent the intended use. If there are many problems in the upper left-hand corner of the matrix, it means that serious problems exist in important tasks. If the problems are in the lower part of the matrix, they come from less important tasks, and if they are in the right-hand part they are not so serious.

Matrix B shows problem seriousness versus problem type. This kind of matrix gives an overview of what sorts of problems exist in the interface and how serious they are. Such a matrix may, for instance, show that most of the problems concern marking of buttons, but that the most serious problems have to do with feedback. By studying the numbers in each matrix, it is possible to find patterns, see how serious the problems are, and understand which types of problems are entailed by the interface's design.

Task importance	Problem seriousness			
	1	2	3	4
1	0	0	0	2
2	0	0	5	7
3	0	2	1	3
4	2	2	8	8
5	2	0	1	1

FIGURE 8: Matrix A from case study—problem seriousness versus task importance.

Problem type	Problem seriousness			
	1	2	3	4
U	0	4	0	2
H	4	0	0	0
S	0	2	2	3
T	0	1	7	9
F	0	1	2	7

FIGURE 9: Matrix B from case study—problem seriousness versus problem type.

Matrix C (problem type versus task importance) shows which problems are most common in the most important tasks. Matrixes D and E reveal more specifically how serious the problems are that occur in each task and what types of problems they are.

3.5.2. Case Study of the Home-Care Ventilator. Figures 8–12 and Table 5 show the five matrices that are created from the analysis in the case study. Matrix A (Figure 8) indicates that the interface presents no serious usability problems, since there are no identified problems in the grey field, that is, no serious problems in important functions (only the figure zero). However, there are numerous usability problems with the interface. In tasks with grades 2–4, many problems exist that are not very serious, as illustrated by the numbers in the two right-hand columns of the matrix. For tasks with grades 4 and 5, that is, less important tasks, there are serious usability problems, shown by the numbers in the lower left-hand corner.

To clarify what lies behind the numbers in the matrix, Table 6 shows the five usability problems that entail the number 5 for task importance 2 and problem seriousness 3 in Figure 8 (the cell with diagonal lines). Table 6 is a compilation taken from the analytical protocol (template) for the ECW analysis (Table 4). The retrospective view of the analysis protocol is valuable in understanding the numbers that are shown in the matrices.

Matrix B (Figure 9) shows that the most common problems with the interface consist of deficiencies in the design of text/symbols (T, 17 problems, the grey-marked cells). The serious problems (the cells with diagonal lines) are, however, due to hidden functions (H, 4 problems) and to the machine not meeting the user's expectations (U, 4 problems), which is found through inspection of the protocol from the analysis.

In Figure 10 (matrix C) it can be seen that problems in important tasks are due mainly to deficient text/symbols and

Task importance	Problem type				
	U	H	S	T	F
1	0	0	0	1	1
2	2	0	2	5	3
3	0	0	2	3	1
4	4	2	2	7	5
5	1	2	0	1	0

FIGURE 10: Matrix C from case study—problem type versus task importance.

Task number	Problem seriousness			
	1	2	3	4
1	0	0	0	0
2	0	0	3	1
3	0	1	3	6
4	0	0	1	2
5	0	2	1	3
6	2	0	1	1
7	2	1	2	1
8	0	0	1	2
9	0	0	3	3
10	0	0	0	2

FIGURE 11: Matrix D from case study—problem seriousness versus task number.

Task number	Problem type				
	U	H	S	T	F
1	0	0	0	0	0
2	1	0	0	2	1
3	2	0	2	3	3
4	1	0	1	1	0
5	0	0	2	3	1
6	1	2	0	1	0
7	1	2	0	2	1
8	0	0	1	1	1
9	1	0	0	3	2
10	0	0	0	1	1

FIGURE 12: Matrix E from case study—problem type versus task number.

inadequate feedback (the cells with diagonal lines). These problems also exist in the less important functions, but additional problems occur there which derive from the user's background knowledge and from hidden functionality in the interface.

Figures 11 and 12 (matrices D and E) show, respectively, which types of problems are most common in the tasks and how serious these problems are. The most serious problems occur in tasks 6 and 7 (the cells with diagonal lines), while task 3 is the one with the most problems (the grey-marked cells).

TABLE 6: Usability problems for task importance 2 and problem seriousness 3 in matrix A (Figure 8).

Task no.	Task importance	Usability problem	Problem Seriousness	Problem type
4	2	The icon for CPAP can be hard to interpret	3	T
8	2	Easy for the users to miss that they are in a function mode when the panel should be locked	3	F
4	2	The function button can be hard to interpret	3	T
8	2	The function button can be hard to interpret	3	T
9	2	The function button can be hard to interpret	3	T

It is important to emphasise that the qualitative result of the ECW analysis (Table 4), that is, a description of usability problem in text, is always employed in order to understand and interpret what the matrices illustrate.

When summarising the evaluation, the result shows that the user interface needs to be improved with more clear and informative symbols and also that the feedback to the user needs to be improved. This is due to that the main problem types in important functions were “Text and icon” and “Feedback.” But there were also serious problems with “Hidden” interaction. For the analysed wake-up function there was also a problem found that the user does not expect the function. All together this indicates that the user interface of the device probably needs a conceptual approach.

It is further interesting to make a reflection regarding how the results from CW (version 3) had differed from ECW in the case study and what specific problems would have been missed by CW? The first difference is that CW would not have analysed the function level in the same detail as ECW, so it could have missed that the user did not expect the wake-up function. The other main difference is that CW had not resulted in grading and categorisation of the usability problems found, making it harder to get a good over view of the usability of the user interface.

3.5.3. Reflection. In risk analysis, matrices are a common way of combining results from analyses of probability and consequence, since risk is a combination of these parameters [36, 37]. A matrix offers the opportunity of reporting risk for a specific combination of probability and consequence. As matrices are also a suitable tool for reporting large quantities of data, it is natural to place a number in a matrix that describes the quantity of identified risks for that particular combination. An overview is given of all detected risks.

The matrices in ECW are employed to present a comprehensive picture of problems and tendencies (main emphases of problems) in the interface. By combining the results from the analysis in different setups of matrices, different ways of studying the analysis results are elucidated. This is done by studying the numbers in each matrix to see what patterns

and emphases occur. This is even more advantageous with complex interfaces and complicated interactions between the user and the artefact, that is, when a large amount of information needs to be compiled. Patterns may occur that are difficult to interpret when studying the answers from the individual evaluation of operations and functions.

When interpreting the matrices, an overall picture thus emerges from matrices A, B, and C (Figures 7, 8, and 9) of the status of the interface. These types of matrices can also be employed to compare different interfaces and see which ones have the fewest serious problems or whether any differences in type of problems exist between the interfaces. The comparison is made by looking at the same type of matrix for the different interfaces. Most simply, the total number of problems in the upper left-hand corner (marked grey) in Matrix A—problem seriousness versus task importance—can be compared. The interface displaying the fewest total problems is then judged to be best from a usability standpoint.

The problems that are considered serious for important tasks should then be investigated further in order to decide whether they are also potential usability problems in a real situation of handling. ECW (and CW) can give only an indication of where problems may exist in the interface, since the methods are analytical and not empirical. If the purpose of the analysis is only to trace the individual usability problem, the matrices do not need be created; that is, if there is no need to trace tendencies or to obtain an overview from the analysis, matrices can be excluded.

4. Discussion

When evaluating the usability of medical equipment, the most important aim is thus not to perform an evaluation rapidly but to make it as good as possible. For an analysis of presumptive usability problems, it is more important that the method finds as many problems as possible rather than to avoid finding problems which probably do not occur in a real working situation. A comparison can be made with tests for detecting diseases. Most crucial is that a test identifies all patients with the disease, not that all patients with a positive test result actually have the disease.

Moreover, it is only after a usability problem has been identified that it is possible to decide whether the error is plausible or not. Exposing even improbable problems to further evaluation is also beneficial, as these may have serious consequences that may otherwise be overlooked if only the plausible usability problems are investigated. In the same way, the focus is not so much on having a method that is easy to perform by persons with minimal training but rather on persons with skill and expertise attaining a good result with the method.

The ECW method is developed to be used together with the PUEA method which is described in Bligård and Osvalder [38]. ECW is a part of the CCPE methodology (Combined Cognitive and Physical Evaluation) [39].

4.1. Fulfilment of Purpose. The aim in developing enhanced cognitive walkthrough was to counteract the three identified

TABLE 7: Analysis template ECW.

(a) Analysis template ECW for analysis of functions

Function	Task		Task importance	
	Success/failure story	Usability problem	Problem seriousness	Problem type
Questions				
(1) Will the user know that the evaluated function is available?				
(2) Will the user be able to notice that the functions is available?				
(3) Will the user associate the clues with the function?				
(4) Will the user get sufficient feedback when using the function?				
(5) Will the user get sufficient feedback to understand that the function has been fully performed?				

(b) Analysis template ECW for analysis of operations

Operation	Action			
	Success/failure story	Usability problem	Problem seriousness	Problem type
Questions				
(1) Will the user try to achieve the right goals of the operation?				
(2) Will the user be able to notice that the action of the operation is available?				
(3) Will the user associate the action of the operation with the right goal of the operation?				
(4) Will the user be able to perform the correct action?				
(5) Will the user get sufficient feedback to understand that the action is performed and the goal is achieved?				

weaknesses in CW: (1) deficient high-level perspective in the analysis, (2) insufficient information given by the motivations for success or failure about the difference in problems' seriousness for different operations, and (3) the difficulty of obtaining an overview of the results both for a given interface and when comparing interfaces.

4.1.1. Deficient High-Level Perspective. The division into two question levels in ECW resolves the issue of whether the user knows about or seeks the evaluated functionality and whether the interface provides any indications to help the user detect and use the functionality. The question levels thereby provide a higher-level perspective on the interface than CW does.

A difficulty with CW, according to Lewis and Wharton [18], is in dealing with the user's intentions. It is connected to question 1 in CW (*will the user be trying to achieve the right effect?*). They describe this by giving an example: turning on the light in a room. The user's goal can thus be specified: "*Pat wanted to flip the switch*" or: "*Pat wanted to flood the room with light.*" The difficulty in interpreting the user's intentions is not so conspicuous in ECW, since question 1 at level 1 (Table 1) treats the user's goal in general (to light up the room), while question 1 at level 2 (Table 1) concerns the physical operation (pressing the switch).

In the second version of CW [17], the approach is to investigate the user's initial goal before the detailed analysis is begun. The preparations for the analysis are described thus: "*List the goals the user is likely to establish when starting*

the task". This step in the preparations has, however, been eliminated in the third version of CW [18] on the grounds of making the method simpler and more effective to employ. Since the user's initial goal is described in the HTA diagram in the ECW method, this information is once again part of the method.

Altogether, the division into two question levels and the introduction of the HTA diagram mean that the first purpose of the method development has been fulfilled. The two levels make it easier for the analyst to interpret the result at a higher level of interaction.

4.1.2. Insufficient Information in Failure/Success Stories. To rank detected usability problems better, semiquantitative judgments have been introduced in ECW, which are lacking in the third version of CW [18]. This limitation of CW is hardly discussed in the literature, except by Jeffries et al. [40] who state that CW misses "*general and recurring problems*" and that the problems which the method identifies are not so serious. It is an advantage to be able to semiquantitatively judge whether the problems identified are serious or not, which can be done with the ECW method.

In ECW, failure/success stories are also supported by a grading that makes it possible to compare the seriousness of different problems and operations. The grading also means that the judgment does not need be only YES or NO but that there are levels between these extremes. The ECW method's way of grading failure/success stories is similar to a grading

which existed in earlier versions of CW but which was eliminated in the development of the third version so as to increase the method's effectiveness and simplicity. In the first version of CW, a grading from 0 to 3 was made [16], and in the second version a grading of the user's failure is made [0 25 50 75 100]% [17]. The development of method to ECW was carried out, however, without any awareness of the grading in previous versions of CW. The fact that earlier versions of CW have employed grading strengthens its role in ECW. The expanded analysis in ECW with question levels and matrices means that the grading is a natural feature and not something that renders the evaluation more difficult.

The grading of failure/success stories implies that the second purpose of the method has been fulfilled.

4.1.3. Difficulty of Obtaining an Overview of Results. Due to the grading of tasks and categorisation of problems in ECW, the results can be reported in the form of matrices. ECW, unlike CW, also yields semiquantitative analytical results that enable presentation of the results in matrices. The matrix structures provide a lucid way of evaluating and analysing several aspects of the user interface. Information about which types of problems may arise and their seriousness, for example, can be read from the matrices. Conclusions can thus also be drawn about what problem tendencies exist. When the problems' seriousness is weighed against the importance of the tasks that the problems occur in, this constitutes a way of judging the state of the interface in aspects of explorative learning coupled with usability. Moreover, presentation of the results in matrices can be employed to compare interfaces, both for different types and manufacturers of interface, and when redesigning already existing interfaces. The introduction of matrices implies that the third purpose of the method development is fulfilled.

By presenting the results of the analysis in matrices, the focus is lifted from detailed problems in the interface to a more general holistic level. The hope here is to create a high-level perspective on the analysis of the interface, which is lacking in CW. This, too, contributes to fulfilling the first purpose of the method development.

4.1.4. Relation to Cognitive Walkthrough Version 2. Changes from CW version 3 to ECW to some extent signify a return to CW version 2 in that there is greater focus on the user's goals and the introduction of a grading of problems. Utilising HTA and the two levels (function and operation) means that it becomes simpler to evaluate task structure and task complexity and hence not guessability. This also points to greater similarity with CW version 2.

ECW is, however, more suitable for practical use as the method has a more straightforward approach than CW version 2. This occurs as a result of the use of the question process idea from CV version 3, and no retakes are done in evaluation as in CW version 2; that is, each operation is run through once only. The number of questions is also lower in ECW than in CW version 2. CW version 2 has a more in-depth analysis which may, however, be more suitable for application in research contexts than in practical use.

4.2. Weaknesses and Limitations of ECW. Even though the method development into ECW has counteracted certain weaknesses in CW, there are still some minor weaknesses in the method. These mainly concern the limited extent of the analysis and the fact that the method is tedious, complicated, and time demanding to apply.

The analysis is limited firstly in that ECW, just like CW, primarily studies learnability through investigation, so that only a limited part of the usability is evaluated. ECW evaluates mainly guessability, but not the aspects of memorability, efficiency, error prevention, or satisfaction, which belong to usability as defined by Nielsen [11]. Efficiency in particular has not been evaluated by methods such as CW/ECW [41]. Secondly, the analysis is limited because CW/ECW is primarily an inspection method and not an empirical method. This entails a possibility that CW/ECW will find more problems than those which are relevant in a real user situation.

The ECW analysis needs to be supplemented by more methods (triangulation) in order to achieve a more comprehensive analysis of usability. Such methods are heuristic evaluation (HE) and usability tests (UT), as has also been proposed by Lewis and Wharton [18] to complement CW. Koutsabasis et al. [42] also come to the conclusion that the use of a single method is not enough for a comprehensive usability evaluation. Hollingsed and Novick [43] reported that CW and HE are often combined in development projects. The triangulation with HE and UT also means that the potential usability problems found with ECW can be confirmed or discounted with the aid of supplementary analysis by these methods, which will be discussed in more detail in the next section. Another possibility is to let real users in a focus group discussion decide whether the potential problems detected with ECW are real or not.

A further weakness of the ECW method, as well as of CW, is that they are somewhat complicated to use. For example, Miller and Jeffries [44] state that one of the drawbacks of CW is that the method is tedious. Since the method development into ECW essentially consists for the most part of additions to CW, it has not counteracted this weakness at all. However, ECW gives a very extensive and useful result when the analysis is completed. The analysis template for ECW (Table 7) is a tool that structures and thus speeds up the analytical procedure. ECW is more complicated than CW, but after learning and training the analysis goes more smoothly. The total result of the method's refinement is that better quality and usability are achieved when the evaluation is done with ECW instead of CW.

Even if the previous discussion claims that ECW is better than CW, there are some occasions when CW (version 3) can be considered to be good enough. This is when the evaluated user interface is uncomplicated, for example if there only are one or two functions. Then the ECW approach can be too burdensome. Also if there is no need for such detailed information as the grading and categorisation that ECW provide, it can be perceived as bothersome. But ECW is made to be flexible, so it is possible to only use the parts that are judged to be useful in the specific evaluation. A skilled and creative method user does not get stuck in a complicated

structure but finds ways to use only parts needed in the specific evaluation.

4.3. ECW in Relation to Evaluation of Medical Equipment. The weakness that ECW detects too many potential problems and takes time to conduct the analysis, however, must be seen in a different light when evaluating safety-critical interfaces such as those in medical equipment.

To be sure, Lewis and Wharton [18] wrote: “*The CW is not worthwhile if it is not done quickly.*” Yet in evaluation of safety-critical systems, the key is not to perform the evaluation quickly but to do so as well as possible. All opportunities for wrong operations must be minimised with such systems, and it is then beneficial if even improbable problems are exposed to further evaluation.

A safety-critical area is medical care, where the possibility of use errors in handling technical products can have serious consequences. Well-designed and adapted interfaces reduce probability of use errors, thereby making patients safer and the personnel’s working environment more secure. ECW is very suitable to employ in evaluation of user interfaces for medical equipment [45], as has been done for home-care ventilators [34, 46], infusion pumps [47, 48], insulin pumps [49], and dialysis machines [50, 51]. The method has shown good effectiveness in finding presumptive usability problems during these studies. The ECW method is also suitable for combination with methods for analysis of use errors such as predictive use error analysis [38]. However, to confirm the ECW method’s utility and its advantages over CW with certainty, further empirical validation is necessary.

The drawbacks are that ECW takes a long time to perform, is tedious to apply, and detects usability problems which may not be plausible and are thus less prominent when the method is applied to medical equipment. Therefore, ECW is a suitable method for detecting, identifying, and presenting usability problems in medical equipment. The ECW analysis can therefore be seen as a necessary, but not comprehensive, segment of usability evaluation for medical equipment.

5. Conclusions

Enhanced cognitive walkthrough (ECW) has been developed as an attempt to counteract the weaknesses identified in the third version of CW. These were a deficient high-level perspective in the analysis, insufficient information in the motivations for success or failure about differences in the seriousness of problems for different operations, and the difficulty of obtaining an overview of the results for a given interface or when comparing interfaces.

The goal in developing ECW was to present a method that could better detect and identify individual presumptive usability problems in a user interface and provide a comprehensive picture of which types of problems exist and how serious they are. This goal has been achieved by making three additions to the CW method:

- (1) division into two question levels, allowing investigation not only of operations but also of tasks/functions (*improved high-level perspective*)

- (2) introduction of indices: a grading of tasks and of failure/success stories and categorisation into problem types (*better description of usability problems*)
- (3) presentation of results in the form of matrices (*clearer overview of the results and improvement of the high-level perspective*).

ECW provides an analysis at a higher level of the user interface (i.e., not only a focus on the operations) than the classical CW method does. This is due mainly to the division in two question levels and the analysis of functions. The grading of evaluated tasks and the seriousness of detected problems, as well as the presentation in matrix form, facilitate the interpretation and utilisation of the analytical results.

ECW has been employed in several case studies for evaluating interfaces in medical equipment. The method has worked successfully and found many problems in interface design that can lead to poor usability with probability of incorrect handling. Further validation is, however, needed in order to confirm that ECW provides a more reliable and comprehensive result than the third version of CW. Although the methods have been developed in the field of medical technology, they are appropriate for use in other domains of human-machine interaction and also for consumer products.

References

- [1] D. R. Kaufman, V. L. Patel, C. Hilliman et al., “Usability in the real world: assessing medical information technologies in patients’ homes,” *Journal of Biomedical Informatics*, vol. 36, no. 1-2, pp. 45–60, 2003.
- [2] E. Liljegren, *Increasing the Usability of Medical Technology Methodological Considerations for Evaluation*, Department of Product and Production Development, Chalmers University of Technology, Göteborg, Sweden, 2004.
- [3] Y. Liu, *Usability Evaluation of Interface Design for Medical Devices—Exploring Methodologies for Evaluation Prototypes*, Department of Product and Production Development, Chalmers University of Technology, Göteborg, Sweden, 2004.
- [4] J. Zhang, V. L. Patel, T. R. Johnson, and E. H. Shortliffe, “A cognitive taxonomy of medical errors,” *Journal of Biomedical Informatics*, vol. 37, no. 3, pp. 193–204, 2004.
- [5] J. H. Obradovich and D. D. Woods, “Users as designers: how people cope with poor HCI design in computer-based medical devices,” *Human Factors*, vol. 38, no. 4, pp. 574–592, 1996.
- [6] L. Lin, R. Isla, K. Doniz, H. Harkness, K. J. Vicente, and D. J. Doyle, “Applying human factors to the design of medical equipment: patient-controlled analgesia,” *Journal of Clinical Monitoring and Computing*, vol. 14, no. 4, pp. 253–263, 1998.
- [7] U.S. Food and Drug Administration, Postmarket information—Device surveillance and reporting processes. 09/29/2011, <http://www.fda.gov/medicaldevices/deviceregulationandguidance/humanfactors/ucm124851.htm>.
- [8] J. Zhang, T. R. Johnson, V. L. Patel, D. L. Paige, and T. Kubose, “Using usability heuristics to evaluate patient safety of medical devices,” *Journal of Biomedical Informatics*, vol. 36, no. 1-2, pp. 23–30, 2003.
- [9] M. W. M. Jaspers, “A comparison of usability methods for testing interactive health technologies: methodological aspects

- and empirical evidence," *International Journal of Medical Informatics*, vol. 78, no. 5, pp. 340–353, 2009.
- [10] H. R. Hartson, T. S. Andre, and R. C. Williges, "Criteria for evaluating usability evaluation methods," *International Journal of Human-Computer Interaction*, vol. 13, no. 4, pp. 373–410, 2001.
 - [11] J. Nielsen, *Usability Engineering*, Academic Press, Boston, Mass, USA, 1993.
 - [12] K. Garmer, E. Liljegren, A. Osvalder, and S. Dahlman, "Application of usability testing to the development of medical equipment. Usability testing of a frequently used infusion pump and a new user interface for an infusion pump developed with a human factors approach," *International Journal of Industrial Ergonomics*, vol. 29, no. 3, pp. 145–159, 2002.
 - [13] A. W. Kushniruk and V. L. Patel, "Cognitive and usability engineering methods for the evaluation of clinical information systems," *Journal of Biomedical Informatics*, vol. 37, no. 1, pp. 56–76, 2004.
 - [14] J. Nielsen and R. L. Mack, Eds., *Usability Inspection Methods*, Wiley, New York, NY, USA, 1994.
 - [15] M. J. Graham, T. K. Kubose, D. Jordan, J. Zhang, T. R. Johnson, and V. L. Patel, "Heuristic evaluation of infusion pumps: implications for patient safety in Intensive Care Units," *International Journal of Medical Informatics*, vol. 73, no. 11-12, pp. 771–779, 2004.
 - [16] C. Lewis, P. Polson, C. Wharton, and J. Rieman, *Testing a Walkthrough Methodology for Theory-Based Design of Walk-Up-and-Use Interfaces*, ACM, Seattle, Wash, USA, 1990.
 - [17] P. G. Polson, C. Lewis, J. Rieman, and C. Wharton, "Cognitive walkthroughs: a method for theory-based evaluation of user interfaces," *International Journal of Man-Machine Studies*, vol. 36, no. 5, pp. 741–773, 1992.
 - [18] C. Lewis and C. Wharton, "Cognitive walkthrough," in *Handbook of Human-Computer Interaction*, M. Helander, T. K. Landauer, and P. Prabhu, Eds., pp. 717–732, Elsevier Science BV, New York, NY, USA, 1997.
 - [19] A. W. Kushniruk, D. R. Kaufman, V. L. Patel, Y. Lévesque, and P. Lottin, "Assessment of a computerized patient record system: a cognitive approach to evaluating medical technology," *M.D. Computing*, vol. 13, no. 5, pp. 406–415, 1996.
 - [20] J. Horsky, D. R. Kaufman, M. I. Oppenheim, and V. L. Patel, "A framework for analyzing the cognitive complexity of computer-assisted clinical ordering," *Journal of Biomedical Informatics*, vol. 36, no. 1-2, pp. 4–22, 2003.
 - [21] E. Liljegren and A. Osvalder, "Cognitive engineering methods as usability evaluation tools for medical equipment," *International Journal of Industrial Ergonomics*, vol. 34, no. 1, pp. 49–62, 2004.
 - [22] Y. Liu, A. Osvalder, and S. Dahlman, "Exploring user background settings in cognitive walkthrough evaluation of medical prototype interfaces: a case study," *International Journal of Industrial Ergonomics*, vol. 35, no. 4, pp. 379–390, 2005.
 - [23] H. Ryu, "Collective web usability analysis: cognitive and activity walkthroughs," *International Journal of Web Engineering and Technology*, vol. 4, no. 3, pp. 286–312, 2008.
 - [24] C. Wharton, J. Bradford, R. Jeffries, and M. Franzke, "Applying cognitive walkthroughs to more complex user interfaces: experiences, issues, and recommendations," in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '92)*, pp. 381–388, May 1992.
 - [25] A. Holzinger, "Usability engineering methods for software developers," *Communications of the ACM*, vol. 48, no. 1, pp. 71–74, 2005.
 - [26] A. Chapanis, "Some reflections on progress," in *Proceedings of the Human Factors Society 20th Meeting*, Santa Monica, Calif, USA, 1985.
 - [27] K. T. Ulrich and S. D. Eppinger, *Product Design and Development*, McGraw-Hill, Boston, Mass, USA, 2004.
 - [28] N. A. Stanton, "Hierarchical task analysis: developments, applications, and extensions," *Applied Ergonomics*, vol. 37, no. 1, pp. 55–79, 2006.
 - [29] B. Kirwan and L. K. Ainsworth, *Guide to Task Analysis*, Taylor & Francis, London, UK, 1992.
 - [30] D. E. Embrey, "SHERPA: a systematic human error reduction and prediction approach," in *Proceedings of the International Topical Meeting on Advances in Human Factors in Nuclear Power System*, American Nuclear Society, Knoxville, Tenn, USA, 1986.
 - [31] C. Baber and N. A. Stanton, "Task analysis for error identification: a methodology for designing error-tolerant consumer products," *Ergonomics*, vol. 37, no. 11, pp. 1923–1941, 1994.
 - [32] J. May and P. Barnard, "The case for supportive evaluation during design," *Interacting with Computers*, vol. 7, no. 2, pp. 115–143, 1995.
 - [33] C. Wharton, J. Rieman, C. Lewis, and P. G. Polson, "The cognitive walkthrough method: a practitioners's guide," in *Usability Inspection Methods*, J. Nielsen and R. L. Mack, Eds., pp. 106–139, John Wiley & Sons, New York, NY, USA, 1994.
 - [34] L. O. Bligård and S. Wass, *Analysis and Development of User Interface for Home Care Airflow Generators*, Product and Production Development, Chalmers University of Technology, Göteborg, Sweden, 2002.
 - [35] Y. Liu, *Usability Evaluation of Medical Technology: Investigating the Effect of User Background and Users Expertise*, Chalmers University of Technology, Göteborg, Sweden, 2009.
 - [36] National Patient Safety Agency, 2008, A risk matrix for risk managers, <http://www.nrls.npsa.nhs.uk/EasySiteWeb/getresource.axd?AssetID=60149&...>
 - [37] E. W. Israelski and W. H. Muto, *Human Factors Risk Management as a Way to Improve Medical Device Safety: A Case Study of the Therac-25 Radiation Therapy System in Using Human Factors to Improve Patient Safety*, J. W. Gosbee and L. L. Gosbee, Eds., Joint Commission Resources, Oakbrook Terrace, Ill, USA, 2005.
 - [38] L. O. Bligård and A. L. Osvalder, "An analytical approach for predicting and identifying use error and usability problem," vol. 4799 of *Lecture Notes in Computer Science*, pp. 427–440, Springer, Berlin, Germany, 2007.
 - [39] L.-O. Bligård and A.-L. Osvalder, "CCPE-methodology for a combined evaluation of cognitive and physical ergonomics in the interaction between human and machine," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 19, no. 6, 2012.
 - [40] R. Jeffries, J. R. Miller, C. Wharton, and K. M. Uyeda, *User Interface Evaluation in the Real World: A Comparison of Four Techniques*, ACM, New Orleans, La, USA, 1991.
 - [41] Z. Ruttkay and R. Op Den Akker, "Affordances and cognitive walkthrough for analyzing human-virtual human interaction," in *Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction*, Lecture Notes in Computer

- Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 90–106, Springer, Berlin, Germany, 2008.
- [42] P. Koutsabasis, T. Spyrou, and J. Darzentas, “Evaluating usability evaluation methods: criteria, method and a case study,” in *Human-Computer Interaction. Interaction Design and Usability*, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 569–578, Springer, Berlin, Germany, 2007.
- [43] T. Hollingsed and D. G. Novick, “Usability inspection methods after 15 years of research and practice,” in *Proceedings of the 25th ACM International Conference on Design of Communication (SIGDOC '07)*, pp. 249–255, October 2007.
- [44] J. R. Miller and R. Jeffries, “Interface-usability evaluation: science of trade-offs,” *IEEE Software*, vol. 9, no. 5, pp. 97–98, 1992.
- [45] L. O. Bligård, *Prediction of Medical Device Usability Problems and Use Errors—An Improved Analytical Methodical Approach*, Chalmers University of Technology, Göteborg, Sweden, 2007.
- [46] L. O. Bligård, S. Wass, E. Liljegren, and A. L. Osvalder, “Using a human factors engineering process to develop new user interfaces for home car air-flow generators,” in *Proceedings of the 35th Annual Congress of the Nordic Ergonomics Society Conference*, Reykjavik, Iceland, 2003.
- [47] A. Gross and J. Foufas, *Developing a Human Factors Engineering Process for Medical Device Design. Methodological Considerations*, Product and Production Development, Chalmers University of Technology, Göteborg, Sweden, 2003.
- [48] E. Liljegren, L. O. Bligård, and A. L. Osvalder, “Developing user-friendly interfaces for medical devices,” in *Proceedings of the 8th IFAC Symposium on Automated Systems Based on Human Skill*, Göteborg, Sweden, 2003.
- [49] L. O. Bligård and A. L. Osvalder, *Using Enhanced Cognitive Walkthrough as a Usability Evaluation Method for Medical Equipment*, IEA, Maastricht, The Netherlands, 2006.
- [50] L. O. Bligård, M. Eriksson, and A. L. Osvalder, “Cassette versus conventional bloodlines—analytical evaluation of three dialysis machines,” 2006, Internal Report Gambro Lundia AB, Classified.
- [51] L. O. Bligård, M. Eriksson, and A. L. Osvalder, “Work physiology evaluation of new machine prototype II,” 2006, Internal Report Gambro Lundia AB, Classified.

