



Developers as Users: Exploring the Experiences of Using a New Theoretical Method for Usability Assessment

Downloaded from: <https://research.chalmers.se>, 2024-04-24 22:36 UTC

Citation for the original published paper (version of record):

Bligård, L., Strömberg, H., Karlsson, M. (2017). Developers as Users: Exploring the Experiences of Using a New Theoretical Method for Usability Assessment. *Advances in Human-Computer Interaction*, 2017.
<http://dx.doi.org/10.1155/2017/6131575>

N.B. When citing this work, cite the original published paper.

Research Article

Developers as Users: Exploring the Experiences of Using a New Theoretical Method for Usability Assessment

Lars-Ola Bligård, Helena Strömberg, and MariAnne Karlsson

Chalmers University of Technology, 412 96 Gothenburg, Sweden

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

Received 22 September 2016; Revised 4 January 2017; Accepted 31 January 2017; Published 5 March 2017

Academic Editor: Jean Vanderdonckt

Copyright © 2017 Lars-Ola Bligård et al. 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.

There is a need for appropriate evaluation methods to efficiently identify and counteract usability issues early in the development process. The aim of this study was to investigate how product developers assessed a new theoretical method for identifying usability problems and use errors. Two cases where the method had been applied were selected and the users of the method in them were asked to fill in a questionnaire and were then interviewed about their experiences of using the method. Overall, the participants (students and professionals) found the methods useful and their outcome trustworthy. At the same time, the methods were assessed as difficult to learn and as cumbersome and tedious to use. Nevertheless, both students and professionals thought that the methods would be useful in future development work. Suggestions for further improvement included provision of further instructions, for example, on how to adapt the methods and development of an IT-support tool.

1. Introduction

For most products, from simple artefacts to complex technical systems, safe and easy handling is essential. Therefore products need to be designed with a high level of usability [1]. A step towards creating products that are safe and easy to use is to try to identify and counteract mismatches in the interaction between users and products as early as possible in the development process, long before the product is to be used in a real use situation. The earlier in the process that problems can be detected, the better the possibilities to adjust the design [2, 3].

To be able to identify and counteract possible usability issues, there is thus a need for usability evaluation methods that can be applied early in the development process. A number of methods have been developed for this specific purpose, including theoretical or expert-based methods such as heuristic evaluation, link analysis, cognitive task analysis, and cognitive walkthrough [4–8]. In addition to being applicable early in the development process, theoretical or expert-based methods hold additional benefits compared to user-based methods including that they may be performed

without first-hand access to users and that they require less time and effort (e.g., [9]).

However, the benefits can only be experienced if developers in actual development work use the methods. Knowledge on the dissemination of theoretical and expert-based methods in industry appears scarce. Nielsen (as one exception) completed one investigation in 1992, in which the participants in a course on usability inspection methods were surveyed 7–8 months after the course to find whether they used the methods they had been taught, why or why not, as well as which methods they in fact used [10]. According to the survey results, methods such as cognitive walkthrough were considered less useful and were used less than usability testing. In a more recent study Jerome and Kazman [11] found that even though approximately one-fourth of the tools used were cognitive walkthroughs, “... the application and adoption of methods and processes from SE [Software Engineering] and HCI [Human-Computer Interaction] research has not yet trickled down into industry” and further that “HCI methods are being used far too late in the life cycle to be truly cost and time efficient.” There thus appears to be a need to further investigate how to increase the dissemination of such methods.

One way to facilitate the dissemination and adoption of these methods is to consider the methods as a “product” and apply user-centred design principles to the method development. One basic principle of user-centred design is to involve the end-users of a product in the development process. To consider the product developers as the end-users of a particular method and to elicit how these users perceive the method are consequently important in method development but this is most often not the case [12, 13]. One important keystone for a successful method development is therefore to evaluate the new method with the intended end-users. The efficiency of a usability method is obviously important, but the effectiveness and satisfaction are as important. If the developers cannot use the method or if they do not, for some reason or other, like the method, the likelihood that the new method will be used decreases. To investigate how users experience the use of a method is hence a significant activity in method development.

2. Aim and Scope

This paper presents a study comprising two cases where product developers (students and professionals) evaluated a combination of two recently developed theoretical methods: Enhanced Cognitive Walkthrough (ECW, [14]) and Predictive Use Error Analysis (PUEA, [15]). Both methods have primarily been used by their developers in development projects in order to identify usability problems and use errors [16–20] and in some documented cases the methods have been used by other people than the developers, for instance, by Moradi and Pour [21] and by Westerlund et al. [22]. In these cases, the methods have performed well and have provided useful knowledge for the improvement of, for example, medical equipment. However, there is further need to study whether the methods are able to perform equally well for other users than its developers.

The specific aim of the study presented here was to investigate how product developers assess the methods from a usability point of view, what strengths and weaknesses they see, how inclined they are to use the methods in future development work, and what their suggestions for improvements are.

3. Description of the Evaluated Method

The method evaluated in the two cases was a combination of Enhanced Cognitive Walkthrough (ECW) and Predictive Use Error Analysis (PUEA). ECW and PUEA are two submethods designed to be applied together in the methodological framework of CCPE (Combined Cognitive and Physical Evaluation). The CCPE framework consists of four phases: (1) definition of evaluation, (2) description of the human-machine system, (3) interaction analysis (with usability problem analysis and use error analysis), and (4) presentation. ECW and PUEA are used in the interaction analysis phase: ECW for the usability problem analysis and PUEA for the use error analysis. A more detailed description of the procedures is given in the following subsections.

The rationale for the development of the respective methods was to improve existing human factors engineering methods to create a methodology that integrates evaluation of use error and usability problems. Both ECW and PUEA have an analytical approach as they are performed by one or more analysts with support from theoretical models, such as explorative learning [23], the Skill, Rule and Knowledge-model [24], and Generic Error Modelling System [25]. The method is applied early in the development process, for instance, on a low fidelity prototype, with the intention that this should allow for the method to be used proactively and for the developers to detect and counteract usability problems and use errors before they are realized in the design of the product.

The method can be used by a single analyst or by a group of analysts. The group may consist of designers, engineers, human factors experts, and users. The most important factor when putting together the group is that the participating individuals have knowledge of who the users are, as well as the product(s) and its intended use.

The innovative features of the method are considered to be (1) the integrated analysis of usability problems and use error, describing the causes of identified mismatches in the interaction and the effects of the mismatch in the interaction, respectively, (2) the analysis on both a functional and an operational level, (3) the grading and categorisation in the analysis, and (4) the presentation of the result in the form of a matrix. The argued benefits are that the analysis becomes more comprehensive and coherent compared to the former methods.

The validity of the ECW/PUEA has been evaluated in a study by Bligård and Osvalder [26]. The study investigated how well the results from ECW and PUEA matched the results from usability tests on a vacuum cleaner and an office chair, respectively. The conclusions from the study were that ECW/PUEA worked well in finding usability problems (91%) and use errors (59%) compared to problems and errors identified with usability testing. The method also delivered the intended result (presumptive usability problems and use errors) to be a valuable tool for use in a product development process, especially in the early stages, before more extensive empirical evaluations are performed.

3.1. The Definition of Evaluation and Description of the Human-Machine System Phases. The first phase, *definition of evaluation*, establishes the boundaries for the analysis by stating the product, the intended user, use and use context. In the *description phase*, the human-machine system to be evaluated is specified. This includes a specification and a detailed description of the task, the user, the use situation, and the user interface of the product (i.e., “the machine”). The tasks are described using Hierarchical Task Analyses (HTA) [27]. The description phase is considered to have large impact on the quality of the result since the next step, the *analysis phase*, depends upon a correct and exhaustive description of the user, the situation, and the task.

3.2. Interaction Analysis. In the analysis phase, the interaction between the user and the product is evaluated by applying a detailed question process. The evaluation takes

place both on a functional level (level 1) and on an operational level (level 2) where the operation level involves the actual actions and the function level concerns overlying objectives for a set of operations.

3.2.1. Usability Problem Analysis through the ECW Method. The first part of the interaction analysis is to analyse usability problems in the human-machine system. The usability problem analysis is performed by the Enhanced Cognitive Walkthrough (ECW) method ([14]). ECW is a usability inspection method based on the third version of cognitive walkthrough (CW) [8, 23].

ECW is an analytical method which looks into potential usability problems by investigating what prevents the user from performing correct actions and why that happens. A usability problem is, according to Nielsen (1993), any aspect of the design that is expected, or observed, to cause user problems with respect to some relevant usability measure (e.g., learnability, performance, error rate, and subjective satisfaction) and that can be attributed to the design of the product. ECW employs a detailed procedure for simulating the interaction between user and product and the user's problem-solving process in each step of the interaction. Throughout, it is investigated whether the supposed user's goals and previous experience will lead to that the correct action is performed.

To predict usability problems, the analyst works through the question process in ECW for all the selected tasks. The interaction analysis is based on the described correct handling sequences in the HTA. The question process then generates conceivable usage problems. The question process is divided into two levels of questions as follows.

Analysis Questions for ECW

Level 1: Analysis of Tasks/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 the position the user is at 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?
Do the abilities of the user match the demands by the machine?
- (5) Will the user get sufficient feedback to understand that the action has been performed and the goal has been achieved?
Does the user understand, after the performed operation, that he/she has done it correctly?

The first (level 1) is employed for the functions (the nodes in the HTA), and the second (level 2) for the operations (the lowest level in the HTA). In level 1, the machine'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.

The analyst asks the questions for each node and operation, respectively, in the HTA diagram, following one branch all the way down before proceeding to the adjacent node. Each question is answered with a grade (a number between 1, a very small chance of success, and 5, a very good chance of success) and a justification for the grade. These justifications, called *failure/success stories*, are the assumptions underlying the choice of grades, such as that the user cannot interpret a displayed symbol. The grading, called *problem seriousness*, makes it easier to determine what it is most important to rectify in the subsequent reworking of the machine.

The next step is to identify the predicted problems. If the problem seriousness is between 1 and 4, that is, not with "a very good chance of success," it points to the existence of a potential usability problem. Based on the failure story, the usability problem is then described. The problem is the cause which prevents the user from performing the correct action. Each problem is further categorised by a problem type. The categorisation stems from the failure stories and the description of the problem. Depending on the machine and the task that the user is to solve with it, different problem types can be used. For more detailed information, see [14].

3.2.2. Use Error Analysis through the PUEA Method. The second part of the interaction analysis is the use of error analysis. Here, the aim is to predict and identify presumptive use errors in the interaction. PUEA is a human reliability assessment method based on three methods: Action Error Analysis (AEA) [28], Systematic Human Error Reduction and

Prediction Approach (SHERPA) [29], and Predictive Human Error Analysis (PHEA) [30]. Also PUEA utilizes a detailed process to break down the user's tasks when interacting with the product into steps and, for each step, predicts and identifies potential use errors. A user error is an "*act or omission of an act that has a different result than intended by the manufacturer or expected by the operator*" to IEC [31, p 17].

To predict use errors, the analyst works through all the selected tasks. The interaction analysis is based on the correct handling sequences described with an HTA. To predict potential incorrect actions, a question process is employed. The question process is divided into two levels of questions as follows.

Analysis Questions for PUEA

Level 1: Analysis of Tasks/Functions

What happens if the user performs an incomplete operation or omits an operation?

What happens if the user performs an error in the sequence of operations?

What happens if the user performs functions/tasks correctly but at the wrong time?

Level 2: Analysis of Operations

What can the user do wrongly in this operation?

What happens if the user performs the operation at the wrong time?

The first (level 1) is employed for the nodes in the HTA, and the second (level 2) for the operations in the HTA. On level 1, use errors are identified that may arise when actions are performed at the wrong time or in the wrong order. On level 2, use errors are identified that may occur in the individual action.

Guided by the questions, the analysts try to predict as many use errors as possible that can arise in the human-machine interaction. Each predicted use error is noted in a list. During this process, they also eliminate errors that are considered too unlikely to occur. This elimination is done in relation to how the simulated user is expected to make decisions and perform, in view of the machine and the social, organisational, and physical contexts. However, it is important to be careful about dismissing without further investigation improbable errors that would have serious consequences, as these can also constitute a hazard. If there are no use errors corresponding to the answers to the questions, this also should be noted.

The analysis proceeds in the same manner as the ECW, starting on a higher level node and moving down to the operations, before moving along the HTA. For each predicted use error, an investigation is made of eight items: (1) error type; (2) error cause; (3) primary consequence of the error; (4) secondary consequence of the error; (5) error detection; (6) error recovery; (7) protection from consequences of use error; and (8) prevention of use error. The first two concern the error itself, the next two concern its potential consequences, and the last four items concern mitigations of

the errors and consequences. Four of the items also contain a categorisation (1 and 2), a judgment of probability (5), or a judgment of severity (4). This is done to facilitate the compilation and assessment of the investigation. For more detailed information, see [15].

3.3. Presentation. The last phase is the presentation phase in which the results in the form of grades and categories are presented in matrices. The matrices are compiled and, by varying the issues from the results of the analysis in rows and columns, different aspects can be emphasized and make the result easier to overview.

3.4. Application. ECW and PUEA are designed to be applied together and they use a common template (Figures 1 and 2). In this way, the prediction and investigation of usability problems and use errors are conducted in parallel; that is, both the ECW question set and the PUEA items are posed at the same time for each node or operation in the HTA diagram. This simultaneous application is why the paper refers to the methods as one, ECW/PUEA; they are perceived as one method when used.

4. Study Procedure

This study is based on two cases where product developers used the ECW/PUEA method. In the first case (A), nine students used the method during a course in order to evaluate a range of user interfaces. In the second case (B), five professional developers used the method to evaluate a prototype in a medical device development project. Both cases were chosen as they represented instances where the method had been applied under circumstances known to the authors. In addition, the two different cases provided the opportunity to get input from individuals under training to become developers and well as from individuals with experience from actual, industrial product development. In order to collect information on the assessment of the method, a combination of questionnaires and interviews was used.

4.1. Participants and Procedure. In case A, nine students were involved in the evaluation. The students attended the second year of their master degree program in industrial design engineering or interaction design, and they were familiar with usability and some other usability methods. Working in pairs, they used the new method as one part of a university course with the aim to perform an extensive cognitive ergonomics evaluation of a human-machine system, including ultrasound machine, disk jockey mixing table, and system camera. The training the students received was limited to a short introduction to the method during a lecture by the first author. They then performed the method by themselves guided by a detailed description of the method. As a part of the examination, students wrote a short reflection on the methods they had used during the course. A month after the end of the course, the students were invited to complete the questionnaire and be interviewed by the second author (first author not present).

Function:	PS	Fail/Success Story	Usability Problem	PT	Notes		
1. Will the user know that the evaluated function is available?							
2. Will the user be able to notice that the function is available?		ECW-part of template					
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?							
<div>• What happens if the user performs an incomplete operation or omits an operation?</div> <div>• What happens if the user performs an error in the sequence of operations?</div> <div>• What happens if the user performs this function at the wrong time?</div>							
Error & Type	Error Cause	Prim. Consequence	Sec. Consequence	Detection	Prevention	Recovery	Protection
	PUEA-part of template						

FIGURE 1: Template for function level analysis for the conjoint use of ECW and PUEA.

Operation:	PS	Fail/Success Story	Usability Problem	PT	Notes		
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 has been performed and the goal has been achieved?							
<ul style="list-style-type: none">What can the user do wrongly in this operation?What happens if the user performs this operation at wrong time?							
Error & Type	Error Cause	Prim. Consequence	Sec. Consequence	Detection	Prevention	Recovery	Protection

FIGURE 2: Template for operation level analysis for the conjoint use of ECW and PUEA.

In case B, five professional developers were involved in the evaluation. The professional developers had slightly different backgrounds, more specifically:

- (i) One system architect with 15 years of work experience and 5 with medical technology
- (ii) One clinical and quality expert, physician with 20-year work experience
- (iii) One quality engineer with 16 years of work experience and 6 years with medical technology
- (iv) Two human factors specialist: one with 7 years of work experience (all in medical technology) and one with 13 years of total work experience and 9 on them with medical technology

They applied the method as part of a risk analysis on an advanced medical device (ventilator), a prototype in real product development project. The professionals received more extensive training than the students. They were taught the method by the first author during half a day. The first author also return on a later occasion to lead the first method application session. The professionals all worked together when performing the analysis. After completing the risk analysis of the medical device they were handed the questionnaire and were then interviewed by the second author (first author not present). The last part of the case B was a focus group with the professional product developers, which was moderated by the first author with the second author present.

4.2. Data Collection and Analysis. The overall topics for the data collection were usability, acceptance, and cost-benefit

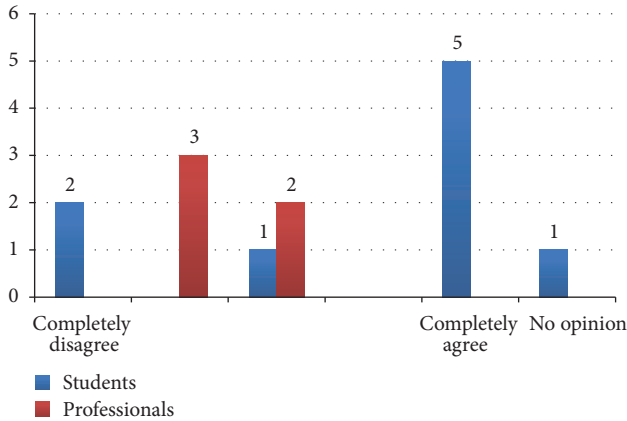


FIGURE 3: ECW/PUEA is very easy to learn.

evaluation. Questions posed concerned learning the method; performing the method; output from the method; general opinions of the method; and possible improvements of the method.

The same questionnaire was used in both cases and it contained altogether 24 items. The items were formulated as statements and the respondents were asked to indicate on a 5-level scale their level of agreement or disagreement (so called Likert items). The option “I have no opinion” was also available. The questionnaire was collected before the interviews were conducted and thus formed the basis for the themes addressed in the interviews. The second author then ran the interviews with one respondent at a time, at the company for the professional and at the university for the students. The interviews were audio-recorded for later analysis.

The results from the questionnaires were compiled and presented in charts. The audio-recordings of the interviews were listened through by the second author, notes were taken, and relevant statements and comments were written down in full. In order to create an overview of the themes that emerged and their interrelations, mind maps of the content were created. All of the authors then together compared the results from the questionnaires and the thematically analysed material from the interviews in order to interpret the meaning and identify strengths, weaknesses, and potential improvements of the methods.

5. Result

The result from the questionnaires and the interviews are presented under the following headings: learning the method, performing the method, output from the method, general opinions of the method, and suggestions for improvement.

5.1. Learning the Method. According to the questionnaires, a majority of the students found the method easy to learn, while the professionals were less positive (Figure 3). There was also a wider distribution among the answers from the students. The same pattern was found regarding both participant groups’ opinion on if the method is easy to carry out once learnt (Figure 4).

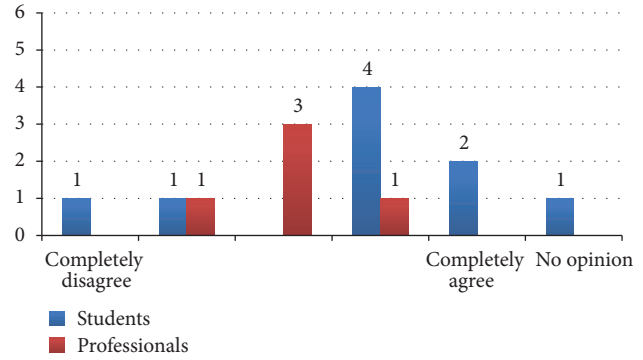


FIGURE 4: Once learnt, ECW/PUEA is very workable.

There was a noticeable deviation between the responses within the student group and also between the students and the professionals. In the interviews, many of the professional product developers commented on how hard it was to get started. They considered the method difficult to use for beginners. The students made similar comments about the learnability of the method.

All of the participants in the study agreed that the part of the method that is easiest to learn is the procedure, that is, learning how to follow the logical sequence of questions. One of the professional participants commented that one “... got a structured help on how to think; one: think like this, two: think like this. Pretty much guiding exactly how you should think, very good guidelines for how to think. It is a problem I generally have in risk analyses that it is hard to think right, it is easy to get lost.”

The aspect that a major part of the participants considered the most difficult was to remember and differentiate between the different terms and rankings, especially when conducting the PUEA-part. The terminology was difficult to pick up, given the sheer amount of terms and the fact that they were similar sounding, for example, primary and secondary consequence. Since there are many categories and rankings to keep track of, the interviewees said that it was easy to just choose a number or category that you remembered without checking if there was a more appropriate one for the case at hand. Another aspect that was considered difficult was to understand which “user’s mind to enter,” that is, which type of user to imagine. If you as an analyst act as a user with very good knowledge of the product and how it is used, the method may not yield so many problems and errors. On the other hand, if you act a novice user, many of the identified potential problems and errors may never occur in actual use.

5.2. Performing the Method. Most students and professionals agreed that the method was very good at creating consensus within the group (Figure 5). In the interviews, four out of the 14 participants stated “increased consensus in the group” as an important result. They here referred both to consensus on the problems and benefits of the product and consensus on what constitute a usability problem and high usability. The method was also considered to be a very good basis for group discussions by the professionals, whereas the

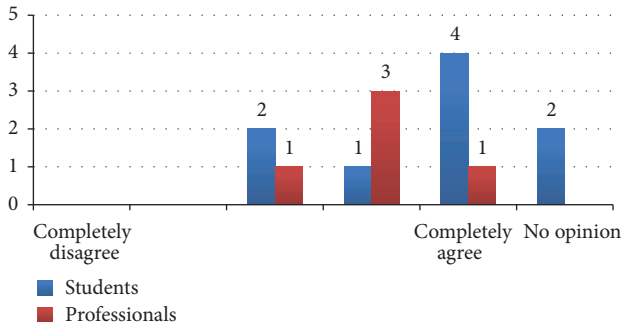


FIGURE 5: ECW/PUEA is a very good at creating consensus within the group performing it.

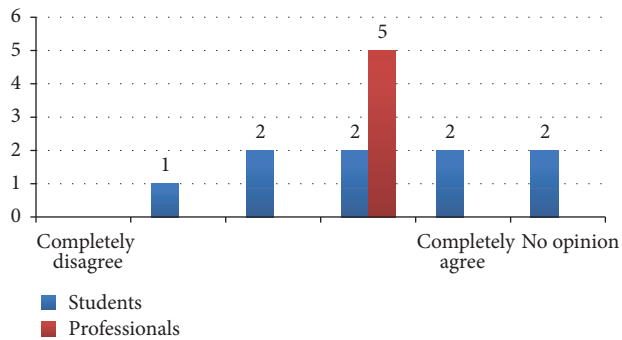


FIGURE 6: ECW/PUEA is a very good basis for group discussions.

students' opinions were more distributed (Figure 6). This might be explained by that the professionals used ECW/PUEA together as one group, while students worked in different groups.

When questioned on the drawbacks of performing the method, the most common answer was the amount of time required: *"(It) takes a very long time because it is so extensive"* (professional). According to the participants, the reason for the time consumed was that the method was comprehensive and repetitive: *"(It is) tedious, the same thing over and over"* (student). This was especially the case when the evaluation was performed on a product with an already satisfactory usability level since not so many problems and errors were detected. In particular, the students found it discouraging to not discover any problems and tedious to find the same type of problems repeatedly.

A few of the interviewees said that keeping up one's concentration level was the most difficult when performing the method. The tediousness of the method and the difficulties to keep alert were pointed out as something that negatively affected the quality of the result. The interviewees explained that after a while you cannot be bothered to find the correct term, judgement, or estimation and just pick one that you have used before. This was especially the case for the PUEA-part.

5.3. Output from Method. In the questionnaires, the participants neither totally disagreed nor agreed completely with the statement that the method results in a great deal of

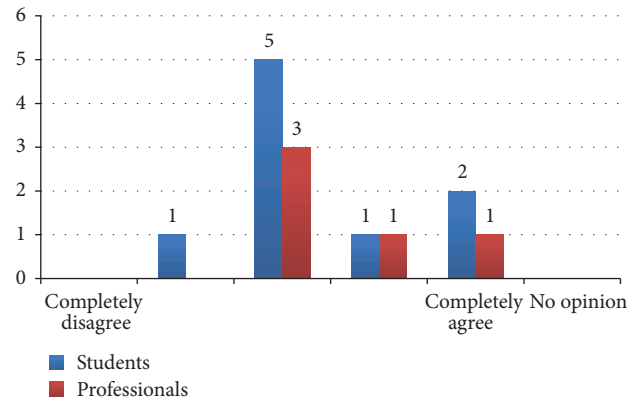


FIGURE 7: Statement: ECW/PUEA gives rise to a lot of new knowledge regarding possible use errors and usability problems.

new knowledge regarding possible use errors and usability problems (Figure 7). Somewhat contradictory, most of them (11 out of 14) commented in the interviews that the primary result from using the method was the discovered problem areas and suggestions for solutions to these problems. To get previously suspected problems confirmed, specified, and written down was found to be an important result.

When asked about their confidence, or trust, in the output from the method, the opinions of the participants differed (Figure 8); some agreed and others disagreed with the statement (with a wider distribution for the students). In the interviews, they explained that it was difficult to make the assessments during the analysis. In addition, it was hard to know if the results were reasonable and viable. Furthermore, they found it difficult to assess whether the method had been performed correctly and to know whether something had been missed. There were also concerns that the method produced results that were subjective, if subjective, one cannot trust the results. However, some participants, who initially considered the method to result in a subjective assessment, changed their mind after having gained experience of applying it: *"I first thought that it would be easily affected and subjective, but I realised that it is objective"* (student).

The inclination to use a method can be assumed to depend upon an assessment of the output compared to the effort required, that is, a kind of cost-benefit analysis [32]. According to the questionnaires, the opinions of the professionals and the students differed slightly when assessing if the method resulted in a large amount of information compared to the effort required. The professionals agreed more than the students with the statement that the method resulted in information of high quality in relation to the effort required (Figure 9).

Both groups disagreed with the statement that the result of the method was independent of the prior knowledge of the participants (Figure 10). They also disagreed with the statement that if different groups complete the method on the same product, they will reach the same result (Figure 11).

When asked about the reasons behind these assessments, a number of aspects were mentioned. One of the reasons was that the method was dependent on the knowledge of the

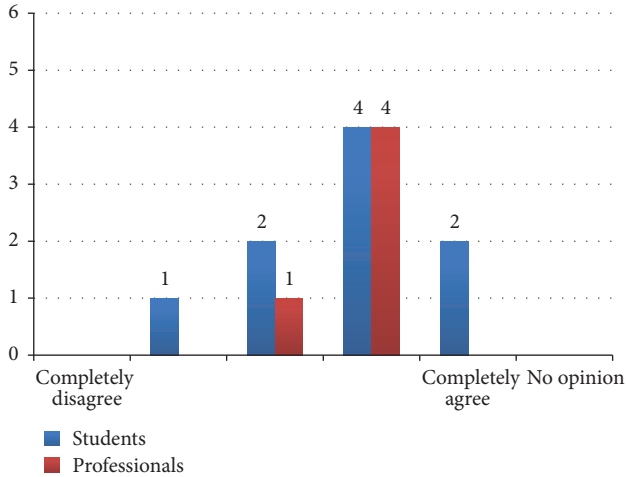


FIGURE 8: Statement: I have great confidence in the result of ECW/PUEA.

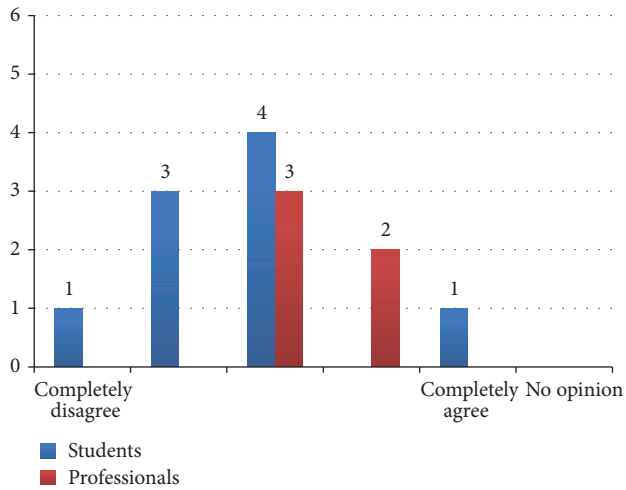


FIGURE 9: ECW/PUEA result in high quality information in relation to the effort it requires.

analysts. If they lacked knowledge of the product, task, and/or user, this was believed to influence the result considerably. One of the students said “[the result] is very much affected by the person performing the method, everybody plays the part of the user differently”. Another explained “you can get false results if the practitioner does not have correct knowledge of the product.” In addition to knowledge, the participants mentioned creativity and imagination, as well as energy to keep concentrated, as important characteristics. Another reason for differences in outcomes was believed to be the analyst’s attitude towards the use of the method and the domain.

Many participants believed that the method required several participants in the group of analysts in order to get valuable output. According mainly to the professionals, the optimum would be that these analysts represented different areas of competence so that different perspectives could be applied. A couple of interviewees from both groups also mentioned that it would be useful to include a representative

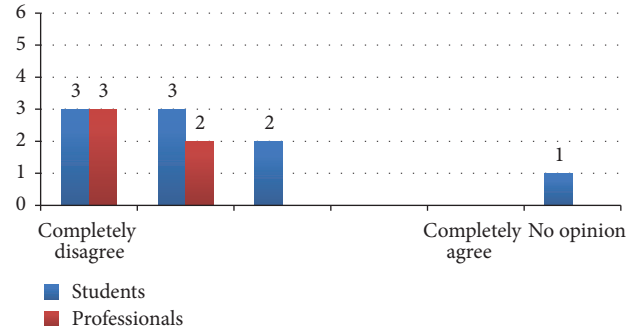


FIGURE 10: Statement: the result of ECW/PUEA is independent of the prior knowledge of the analysts.

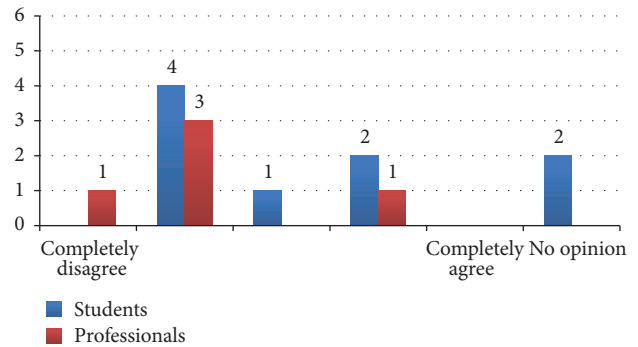


FIGURE 11: Statement: different groups performing ECW/PUEA on the same product will come to the same result.

user in the group, since it could be difficult to judge the reasonableness of which errors could occur and what really is a usability issue on your own.

5.4. Assessment of the Method. A main concern in the evaluation was, evidently, if the participants considered the method useful in product development and if they could consider using it in the future.

A majority of the participants (both students and professionals) agreed with the statement that the method is well worth using (Figure 12). They also agreed with the statement that the method felt like a serious method (Figure 13).

The professionals agreed slightly more so than did the students with the statement that the method felt like a very purposeful method. Most of the participants agreed that the method was a useful method during product development (Figure 14).

In addition, according the questionnaire, most participants could imagine using ECW/PUEA in future projects (Figure 15).

Overall, the method was considered systematic; it provided an easy overview of the issues, and offered clarity and awareness of the problems. There was consensus in the comments that a main strength is that the structured method encourages the developer to consider the usage of the product step by step: “You have to analyse all steps in the task, steps which have become evident to you” (student). The systematic approach makes it easier to think critically of a product: “It

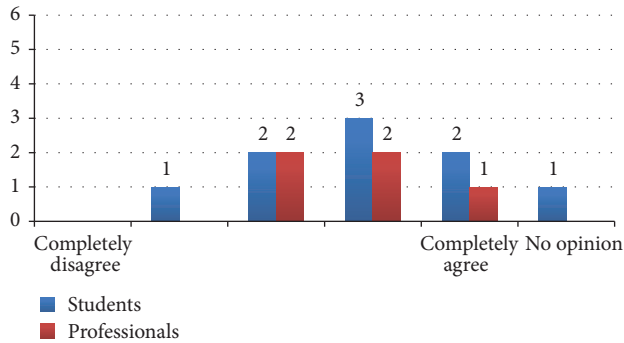


FIGURE 12: Statement: ECW/PUEA is a method well worth using.

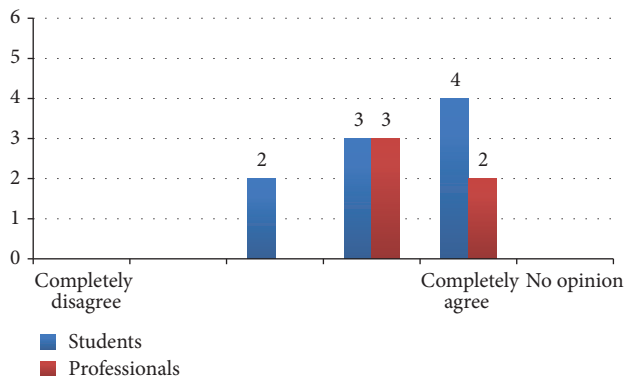


FIGURE 13: Statement: ECW/PUEA feels like a very serious method.

helps you become more critical of something that you have developed yourself” (professional). However, the systematic approach also resulted in the method being perceived as time-consuming, tedious, and unnecessarily complex for application on certain products. The method was by some participants considered to be too “engineery” as it “quantifies everything” (professional) and therefore it was perceived as boring and lengthy. Other participants interpreted this as though the method provides “an objective perspective” and one professional participant argued that the quantification facilitates communicating the concept of usability, which may appear as something “fuzzy” for those unfamiliar with the domain. The result of the method, in terms of a list with individual usability problems and use errors, contributes with a clear picture of the usability of the product. It is “... good way to prove to other people that there are problems and what their problems are” (student). A summary of main strengths and weaknesses of the method according the participants is provided as follows.

Strengths

- (i) It provides a structure for findings potential problems and errors.
- (ii) It helps structure your thoughts.
- (iii) It facilitates reaching consensus within a group of developers.

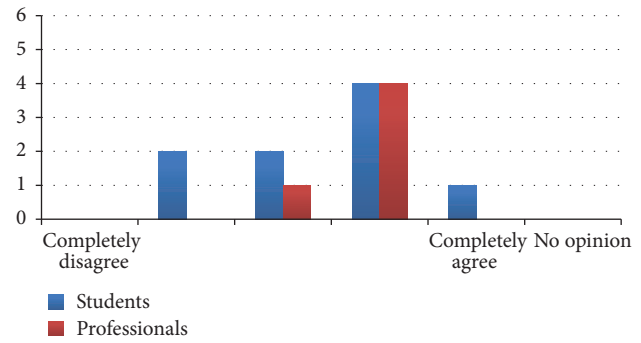


FIGURE 14: Statement: ECW/PUEA is a very useful method during product development.

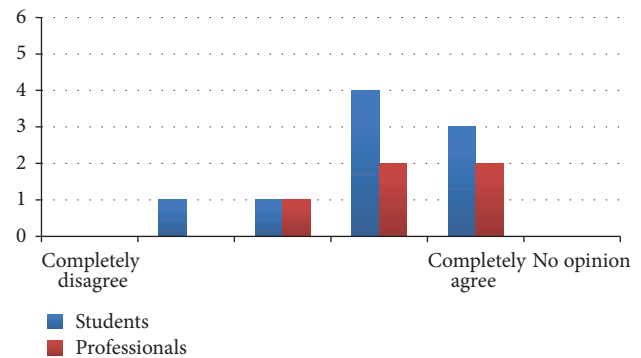


FIGURE 15: Statement: I can imagine using ECW/PUEA in future projects.

- (iv) It forces you to think through the usage of the product.
- (v) It provides convincing arguments.
- (vi) It quantifies results
- (vii) It helps explain the fuzzy concept of “usability.”

Weaknesses

- (i) Time-consuming process
- (ii) Tedious procedure
- (iii) Difficult to grasp the terminology
- (iv) Difficult to assess the quality and reliability of the results
- (v) Results dependent on the competence and experience of the analyst
- (vi) Quantifies results (the need to transform opinions into numbers)

The students and the professionals highlighted the same main strengths and weaknesses.

Three of the problems associated with executing the method were the difficulty to select tasks, prioritize between tasks, and create a good description of the user. The participants believed that the method would work better if the specified product user was a novice rather than an expert, since they found it more difficult to imagine what an expert might do.

5.5. Suggestions for Improvements. Some suggestions for improvements of the method emerged from the interviews. One concerned instructions. Some of the interviewees desired more instructions on how to select and prioritize between the tasks and how to choose user character. One suggestion from a student was “*either adapt [the method] to experienced users or clearly state which use situation and type of user that is suitable.*”

Another suggestion from a student concerned further information on how to adapt the method to the specific product under development: “*[The questions] must be adapted to the specific case.*” Another student indicated that you have to “*modify [the method] according to complexity of product, more like a checklist.*”

Less specific recommendations concerned simplifying the procedure. The method should “*be rendered more efficient, to reduce the time and effort needed*” according to one professional. One idea from a student was to develop an IT-based tool: “*... a programme or advanced Excel-tool,*” so that the analyst can focus on the results and less on the “*administrative task.*”

6. Discussion and Conclusion

The aim of the study was to investigate how two groups of product developers, professional developers and students, assess a new, theoretical method for identifying usability problems and use errors. Two main “conflicts” have been identified: time versus results and structure versus tediousness. These two conflicts are discussed below, followed by further discussion on learning the method, the value of a null result, and some comparison of the students’ and professionals’ experiences. The discussion ends with suggestions for further development of ECW/PUEA and a concluding remark on the paper.

6.1. Time versus Results. The first conflict could be considered an inherent contraposition. Time is generally considered a key issue in product development projects why any methods used in the process must be efficient and provide value for the resources allocated. According to the results from the study, most participants found the new method useful in product development and they trusted the results. At the same time, they found the method tedious and time-consuming. Even so, they indicated that they would consider using the method again in a product development context. This type of “cost-benefit” conflict has earlier been identified regarding structured walkthrough procedures (e.g., by Rowley and Rhoades [33]) and modifications in order to simplify and speed up the process have been suggested. For instance, Rowley and Rhoades [33] proposed a slightly modified “JogThrough” procedure and Spencer [34] proposed the “streamlined cognitive walkthrough.” Little is however mentioned on the relative efficiency of these simplified versions. The input from the potential users of the developed method cannot be neglected, but the fundamental question is if the cost-benefit conflict can be resolved and if simplifications can be made to the ECW/PUEA method without a loss of quality. In fact, the rationale for developing the ECW/PUEA method identified

weaknesses of the existing methods: CW, AEA, SHERPA. and PHEA. The method was not intended to be an optimisation of resources versus detected conceivable use errors, but the aim was that it should detect as *many* problems as possible. In addition, it can be argued that a main part of the participants’ difficulties was related to the preparatory work required. The ECW/PUEA was never designed to be a “standalone” method and much of the information needed for the analysis in ECW/PUEA should be readily available in the project. Knowledge on the users, the tasks, and the context of use is the basis for usability and the foundation for human factors work [35, 36] and should be present in every project. This is a clear indication for a need to make collection and presentation of that knowledge more effective and efficient.

6.2. Structure versus Tediousness. The second conflict could be considered a problem that arise in-between the participants in an ECW/PUEA session. Some participants in the study thought that a particular strength of the method was its structured approach and that the output was presented in a structured way and in numbers, which all contributed to them trusting the outcome. Other participants thought that these same characteristics made the method complex and tedious. Product development is believed to benefit from team members with different backgrounds and different personalities. However, these differences could also cause problems, for instance, when using a method and trusting its result. Members with an engineering background may be more inclined to accept and adopt a method that provides a clear logic and results in “objective” numbers whereas members with, for instance, a traditional design background may be more disposed to trust in intuition and the ability to put oneself in the position of the end-user [37]. If there is such a conflict within the development team, it is not easily solved. Other results from the evaluation points, however, in a more positive direction, such as that the members complemented each other when performing the method.

It is not certain that the result of a user interface evaluation is acknowledged by the designers of the same interface; that is, they may not believe that the result is correct. The result from an ECW/PUEA session can enhance this issue since ECW/PUEA is analytical to its character and lacks the conclusiveness of empirical usability tests. The identified problems of the design may result, as pointed out by, for instance Spencer [34], in more work for a development team already under time pressure. Some team members may try to defend their designs, be argumentative, and may “... reject seemingly obvious observations as being opinions that lack data to support them” (Spencer, 2000). An important comment made by the professional participants in the study was that the ECW/PUEA was believed to be a tool for improving the dialogue between the development team members and contributed to creating consensus within the team. A plausible measure to counteract this issue is therefore to include the designers in the ECW/PUEA session, so they can be part of the dialogue and develop an understanding for the results.

6.3. Learning the Method. Another and related result of the study was that, overall, the method was assessed as

complicated and difficult to learn, in particular the terminology and the rankings. The responses were collected from first time users of the method and the efficiency in performing the method will probably increase over time. It must be acknowledged though that the method is not a “plug-in-and-play” method, that is, the method that can be employed without any initial training. The issue here is evidently how much training is required. It seems as though part of the problem when learning the method is “unfamiliarity” with the structured way of approaching the problem. However, the problem could also be at least partly explained by the participants’ different responsibilities, background, and personalities and hence related to the second identified conflict regarding that the users experience ECW/PUEA in different ways.

6.4. The Value of a Null Result. Another issue worth considering, in particular when teaching the method, is that many of the participating students did not think that the method produced a good result since they did not detect a large number of errors and problems. In a real product development context, this would be a positive outcome. If you do not discover plenty of possible usability problems and use errors, it means that the probability for the product being a safe and useable is high. However, in cases where the method did not detect and problems or errors for the participants, this lack of findings instead resulted in an uncertainty regarding whether or not the method was performed correctly, or a disbelief that you have not discovered the possible problems and errors that are, indeed, there. Thus, when training in the methods, the understanding of the underlying principles of the methods should be emphasised. Even though most participants trusted the outcome of the method, the study revealed that there were those who felt uncertain about how to “approach” the method. Several participants meant that the method was only usable for finding the errors that a first-time user would encounter, not realizing the possibility of assuming a different type of user. This has to be clarified in the future dissemination of the ECW/PUEA method. Other participants found it difficult to know how a more experienced user would think and act because they lacked sufficient knowledge and understanding of end-users, their preunderstanding, the situation in which the product is used, and what effect these factors could possibly have on the behaviour of the user. Some of these are issues that have been identified earlier in relation to structured walkthroughs (e.g. by [38]). The issues mentioned reflect the dependency on the participating analyst(s), a dependency which was mentioned by the participants in the study and which has previously been shown in several studies. For instance, Desurvire et al. [39] concluded that usability experts found more problems than nonexperts. Furthermore, Nielsen (1992) stated that usability experts identified more usability problems than nonexperts when conducting a heuristic evaluation, and further that usability experts who also had expertise with the type of interface (or the domain) being evaluated identified the most. It is therefore reasonable and appropriate to have evaluation methods, like ECW/PUEA, that aim to strengthen the skill of the experts.

6.5. Comparing Students and Professionals. Even though the study did not aim to compare the two cases, there are some discernible differences between the experiences of the students in case A and the professional developers in case B. These differences relate to the last two sections: learning the method and the value of a null result.

A majority of the students experience that the ECW/PUEA was easy to learn, while the professionals found it more difficult to learn. This is probably due to that the students were more accustomed to this type of usability methods, as they that had been taught similar methods earlier (e.g., CW). In addition, they were in a context where they were constantly expected to learn, the university, as well as taking a course focused on learning many different types of methods. The students were thus much more prepared and used to learning new things than the professionals, something that might have affected their appraisal of the method’s learnability.

The professionals experienced more benefit from the method and saw no problem with a null result, in contrast to the students. The professional used method in a real development project where they relied on the results to demonstrate that they had carried out a risk analysis of use, in order to get certain certifications. Furthermore, to them, a null result meant that the evaluated product probably did not contain any flaws, a confirmation of a development work well done. In contrast, the students applied the method in a project aimed at finding design errors to correct, something which may have affected their view of a null result, in addition to the insecurity of having performed the method correctly as discussed above.

6.6. Further Development. The study has contributed to validating the new method with a focus on the users’, that is, the product developers’ experiences. Based on the input from the participants in the study, further development and simplification of the procedure are deemed desirable. A suggestion from the participants in the study was to provide computer support. The creation of an IT-supported version of the tool, where a template could be filled in on the computer screen and then used when creating the different tables and matrices, would most probably reduce the time that has to be allocated the presentation of results. In addition, complementary studies are needed. For instance, in order to assess the effectiveness as well as the efficiency of the new method, comparisons have to be made between the already existing theoretical and expert-based methods and the new, modified one. Such evaluations are important in order to argue the relative benefit of the method, something which is considered a key factor in the dissemination of the method (cf. [40]). Furthermore, the number of participants in the study reported here was limited, why only tentative conclusions can be drawn. The evaluations were carried out after a session where the participants learnt how to use the method and therefore it can be argued that the learnability of the method rather than the usefulness of the method has been addressed in this study. The usefulness of the ECW/PUEA should also be evaluated in actual product development work, without the participation of its developers, and with teams consisting of individuals with different backgrounds, to explore how this effects the procedure and the results.

6.7. Concluding Remarks. The study presented in this paper has shown that a user study of a usability method can be performed in the same manner as a usability study on a product. The study provided good insight into how the developers experienced the method and resulted in useful information about how to improve the method. Thus, showing that user studies with developers is a valuable asset in method development. It is important for developers of usability methods to consider the intended users, that is, the product developers, and not only focus on how well the methods evaluate, for instance, the usability of products. To increase the potential for dissemination of usability methods, the usability of usability methods is an important piece of the puzzle to reach applicability and credibility of methods.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

References

- [1] J. Nielsen, *Usability Engineering*, Academic Press, Boston, Mass, USA, 1993.
- [2] H. Johannesson, J.-G. Persson, and D. Pettersson, *Produktutveckling: Effektiva Metoder för Konstruktion och Design*, Liber Läromedel, Stockholm, Sweden, 2004.
- [3] B. L. Miles and K. Swift, "Design for manufacture and assembly," *Manufacturing Engineer*, vol. 77, no. 5, pp. 221–224, 1998.
- [4] N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas, and H. W. Hendrick, Eds., *Handbook of Human Factors and Ergonomics Methods*, CRC Press, New York, NY, USA, 2005.
- [5] N. Stanton, *Human Factors Methods: A Practical Guide for Engineering and Design*, Ashgate, Aldershot, UK, 2005.
- [6] G. Salvendy, *Handbook of Human Factors and Ergonomics*, Wiley, Hoboken, NJ, USA, 2006.
- [7] J. Wilson and N. Corlett, Eds., *Evaluation of Human Work: Practical Ergonomics Methodology*, Taylor & Francis, London, UK, 1995.
- [8] 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.
- [9] L. Kantner and S. Rosenbaum, "Usability studies of WWW sites: heuristic evaluation vs. laboratory testing," in *Proceedings of the 15th Annual International Conference on Computer Documentation (SIGDOC '97)*, pp. 153–160, ACM, Salt Lake City, Utah, USA, October 1997.
- [10] J. Nielsen, "Finding usability problems through heuristic evaluation," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92)*, pp. 373–380, Monterey, Calif, USA, 1992.
- [11] B. Jerome and R. Kazman, "Surveying the solitudes: an investigation into the relationships between human computer interaction and software engineering in practice," in *Human-Centered Software Engineering—Integrating Usability in the Software Development Lifecycle*, vol. 8 of *Human-Computer Interaction*, pp. 59–70, Springer, 2005.
- [12] J. Andersson, L.-O. Bligård, A.-L. Osvalder, M. J. Rissanen, and S. Tripathi, "To develop viable human factors engineering methods for improved industrial use," in *Design, User Experience, and Usability. Theory, Methods, Tools and Practice*, vol. 6769 of *Lecture Notes in Computer Science*, pp. 355–362, Springer, Berlin, Germany, 2011.
- [13] K. J. Vicente, C. M. Burns, and W. S. Pawlak, "Better handbooks, better design," *Ergonomics in Design*, vol. 6, no. 4, pp. 21–27, 1998.
- [14] L.-O. Bligård and A.-L. Osvalder, "Enhanced cognitive walkthrough: development of the cognitive walkthrough method to better predict, identify, and present usability problems," *Advances in Human-Computer Interaction*, vol. 2013, Article ID 931698, 17 pages, 2013.
- [15] L.-O. Bligård and A.-L. Osvalder, "Predictive use error analysis—development of AEA, SHERPA and PHEA to better predict, identify and present use errors," *International Journal of Industrial Ergonomics*, vol. 44, no. 1, pp. 153–170, 2014.
- [16] L. Lundgren, L.-O. Bligård, S. Brorsson, and A.-L. Osvalder, "Implementation of usability analysis to detect problems in the management of kitesurfing equipment," in *Proceedings of the 5th Asia-Pacific Congress on Sports Technology*, Melbourne, Australia, 2011.
- [17] L.-O. Bligård and J. Andersson, "Use errors and usability problems in relation to automation levels of medical devices," *VTT Symposium (Valtion Teknillinen Tutkimuskeskus)*, no. 258, pp. 111–114, 2009.
- [18] 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.
- [19] L.-O. Bligård and A.-L. Osvalder, "Using enhanced cognitive walkthrough as a usability evaluation method for medical equipment," in *Proceedings of the International Ergonomics Association Conference (IEA '06)*, Maastricht, The Netherlands, 2006.
- [20] L.-O. Bligård, *Predicting Mismatches in User-Artifact Interaction—Development of an Analytical Methodology to Support Design Work*, Department of Product and Production Development, Chalmers University of Technology, Gothenburg, Sweden, 2012.
- [21] S. Moradi and N. A. Pour, "Usability problems and use errors appraisal in public transport website," in *Human Centred Automation*, D. De Waard, N. Gerard, and L. Onnasch, Eds., pp. 213–225, Shaker Publishing, Maastricht, the Netherlands, 2011.
- [22] A. Westerlund, W. Tancredi, M. Ransjö, A. Bresin, S. Psonis, and O. Torgersson, "Digital casts in orthodontics: a comparison of 4 software systems," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 147, no. 4, pp. 509–516, 2015.
- [23] P. G. Poison and C. H. Lewis, "Theory-based design for easily learned interfaces," *Human-Computer Interaction*, vol. 5, no. 2–3, pp. 191–220, 1990.
- [24] J. Rasmussen, "Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 13, no. 3, pp. 257–266, 1983.
- [25] J. Reason, *Human Error*, Cambridge University Press, Cambridge, UK, 1990.
- [26] L.-O. Bligård and A.-L. Osvalder, "Evaluating usability problems and use errors—comparing analytical methods and usability test," *International Journal of Ergonomics and Human Factors*, In press.

- [27] N. A. Stanton, "Hierarchical task analysis: developments, applications, and extensions," *Applied Ergonomics*, vol. 37, no. 1, pp. 55–79, 2006.
- [28] L. Harms-Ringdahl, *Safety Analysis—principles and Practice in Occupational Safety*, Taylor & Francis, London, England, 2nd edition, 2001.
- [29] D. 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*, pp. 184–193, American Nuclear Society, Knoxville, Tenn, USA, 1986.
- [30] D. Embrey, "Qualitative and quantitative evaluation of human error in risk assessment," in *Human Factors for Engineers*, C. Sandom and R. Harvey, Eds., p. 361, Institution of Electrical Engineers, London, UK, 2004.
- [31] IEC, *IEC 60601-1-6:2004 Medical Electrical Equipment—Part 1-6: General Requirements for Safety—Collateral standard: Usability*, IEC, Geneva, Switzerland, 2004.
- [32] R. G. Bias and D. J. Mayhew, *Cost-Justifying Usability*, 2nd edition, 2005.
- [33] D. E. Rowley and D. G. Rhoades, "Cognitive jogthrough: a fast-paced user interface evaluation procedure," in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '92)*, pp. 389–395, May 1992.
- [34] R. Spencer, "Streamlined cognitive walkthrough method, working around social constraints encountered in a software development company," in *Proceedings of the Conference on Human Factors in Computing Systems "The Future is Here" (CHI '00)*, pp. 353–359, The Hague, Netherlands, April 2000.
- [35] D. A. Norman and S. W. Draper, *User Centered System Design: New Perspectives on Human-Computer Interaction*, Erlbaum, Hillsdale, NJ, USA, 1986.
- [36] A. Chapanis, "Some reflections on progress," in *Proceedings of the 20th Meeting on Human Factors Society*, pp. 1–8, Santa Monica, Calif, USA, 1985.
- [37] S. Persson, *Toward Enhanced Interaction between Engineering Design and Industrial Design*, Department of Product and Production Development, Chalmers University of Technology, Göteborg, Sweden, 2005.
- [38] R. Jeffries, J. R. Miller, C. Wharton, and K. M. Uyeda, "User interface evaluation in the real world: A comparison of four techniques," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '91)*, pp. 119–124, ACM, New Orleans, La, USA, May 1991.
- [39] H. W. Desurvire, J. M. Kondziela, and M. E. Atwood, "Whats is gained and what is lost when using evaluation methods other than empirical testing," in *People and Computers VII*, A. Monk and D. Harrison, Eds., pp. 89–102, Cambridge University Press, Cambridge, UK, 1992.
- [40] E. M. Rogers, *Diffusion of Innovations*, Free Press, New York, NY, USA, 1983.

