A New Usability Inspection Method: Experience-based Analysis

Anu Piirisild[1][0000−0002−9569−9752], Ana Perandrés Gómez[2][0000−0003−0869−6391], and Kuldar Taveter[1][0000−0003−3074−7618]

[1] University of Tartu, Institute of Computer Science, Tartu, Estonia
(anu.piirisild@ut.ee)
[2] Ageing Lab Foundation, Jaén, Spain
(ana.perandres@ageinglab.org)
[3] (kuldar.taveter@ut.ee)

Abstract. [**Context and motivation**] For software product development, user-based evaluation is often applied to detect usability issues. User-based testing is not cost-effective at every stage of the project, therefore for the early stage of the development, Usability Inspection Methods (UIM) are widely used where the evaluation is performed by experts. For the later phase, Summative Usability Testing (SUT) methods with end user involvement are used. However, UIMs have some drawbacks that make them complicated to use: finding good experts can be costly, results are too limited, and a lack of methods that allow for usability analysis before prototyping. This article introduces the Experience-based Analysis method (EbA) as an UIM that considers those drawbacks. [**Question/problem**] If we can obtain a considerable amount of usability information with EbA, we can use EbA as an effective UIM method. Therefore, the aim of the study is to identify to what extent the EbA can provide the same usability findings that is obtained by testing with users by SUT. [**Results**] An evaluation was done on a case study, where the information obtained by both EbA and SUT was analyzed by Content Analysis method. The results showed that most of the usability findings by SUT report were also identified by EbA. EbA method resulted in quantitatively more findings than SUT. [**Contribution**] The results lead us to recommend using EbA before SUT. As future work, we first plan to find whether the method gives any false positive or negative results. Second, we plan to conduct a survey to find the limitations of the method.

Keywords: Usability Inspection Method · Human behavior · Usability · Requirements engineering.

1 Introduction

For software product development there is a acknowledged need for early phase usability analysis [6, 1]. Making changes in the early stages of development is less costly than in later stages [1]. Usability is defined in ISO standard as the “extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [1]. User-based testing at every stage of the project is not cost-effective [1, 6], therefore for the early stage of development Usability Inspection Methods (UIM) are often used where usability experts or professionals of other types give feedback on the digital or paper-based prototype and likely user interactions [19, 15, 10, 6]. For the later phase of usability analysis, Summative Usability Testing (SUT) methods are used. SUT assumes that the development phase has been completed at a sufficient level to be evaluated [4]. In a SUT, instead of experts, the end users are involved in testing the prototype or finished product. Due to the cost and logistical challenges of recruiting test-users, a SUT can be costly [15]. Using a SUT can be made more cost-effective and bring greater value by applying an UIM beforehand [1, 12, 10, 15]. UIMs can help to plan the SUT by focusing on predicted issues [15]. The aim of an UIM is to detect usability issues before a SUT to make the otherwise costly usability evaluation more accessible for software development projects [15, 10, 19].

There are several UIMs listed in Section 2, each of which identifies usability issues in a different way. However, there is still some dissatisfaction with UIMs that make them complicated to use. The main drawbacks can be generalized into three categories: 1) finding good experts can be challenging and costly [12, 6, 3]; 2) scope and therefore, the results are too limited and too much focused on design artifact [31, 30, 3]; 3) a lack of UIM methods that allow for early usability analysis before prototyping [26, 9].

In this paper we introduce a new UIM named Experience-based Analysis (EbA) that addresses the drawbacks 1-3 outlined above. We also present results that demonstrate that the EbA method can obtain sufficient usability information as expected from an UIM, EbA can help to plan the SUT by discovering what usability information is missing, and take some workload off from a SUT.

EbA is intended to be used by the people who elicit the requirements. As soon as a new solution is described, requirements engineers can validate whether it would work as such or would it be reasonable to change the requirements before they are handed over for development. EbA is addressing the three drawbacks 1-3 of the UIMs listed above as follows:

1. Knowledge that would otherwise be obtained from experts is collected by the people who drafted the requirements. They are able to perform the inspection by the EbA method where the expected behaviours by the end users are verified to be present in situations that users have experienced in the past.
2. EbA helps to find extensive usability issues, since EbA is not limited by testing the prototype, and the analysis of the expectations enables analysis of the product under development more broadly.
3. EbA does not require creating any prototype for usability inspection since the EbA analysis is intended to be performed by the same person or team who described the requirements. Therefore there is no need to demonstrate the solution to someone else by means of the prototype.

The EbA method was created by the first author of this article in 2021 and was initially evaluated in three real life projects [22]. The results of the evaluation included by [22] showed that using EbA provides valuable usability information, which helps to plan further activities more effectively. In addition, evaluating the EbA in the three projects [22] showed that after getting acquainted with the idea of the product, the EbA analysis would take only 1-2 working days depending on the extent of the planned product and the level of detail of the analysis. In this paper, we introduce the EbA method to the scientific community.

If we want to suggest requirements engineers to use the EbA to learn about usability issues in an early stage of software product development and we presume that the EbA is able to identify usability information that is commonly collected with a SUT, we need to find out to what extent the EbA can obtain the same usability information as a SUT. For this purpose, we formulate the first research question as follows:

RQ1: How does the usability information obtained with the EbA compare to the usability information obtained with a SUT?

Since we know from [22] that one can collect with the EbA a large amount of usability information and we want to understand the quality of such information, we formulate the second research question as follows:

RQ2: To what extent is the usability information obtained by EbA valuable for product owners?

By product owners we mean people who are responsible that the solution under development will work in an efficient way.

To answer RQ1, usability information obtained by both the EbA and a SUT from a complex case study on technologies for older adults was analyzed by the Content Analysis method [17] and the results of the analysis were then compared to each other. As a result, a considerable amount of overlapping usability findings were identified with the two methods. To answer RQ2, two people from the team of the same complex case study evaluated the information obtained by the EbA. They rated 53% of the findings identified by the EbA as valuable. The results suggest to apply the Experience-based Analysis (EbA) method before starting with prototyping and testing with end users to increase the efficiency of development.

In this paper, we did not compare the EbA with the other UIM methods because this falls outside of the scope of this study. This paper focuses on discovering whether the EbA enables detection of a considerable amount of usability information before testing with end users by a SUT method, and if the EbA is able to identify usability information that is commonly collected with a SUT.

The rest of this paper is organized as follows. Section 2 provides an overview of UIMs and results of similar evaluation studies comparing an UIM and a SUT with each other. Section 3 describes the EbA method and its usage. Section 4 provides a short overview of the case study and describes the validation process. Section 5 reports the results of the validation. Section 6 discusses the results. Finally, Section 7 concludes the paper and outlines the future work.

2 Background and Related Work

In the relevant research literature, the most frequently used UIM methods are Heuristic Evaluation (HE), Cognitive Walkthrough (CW), and Task Analysis (TA). In HE, three to six usability experts evaluate user interface with a heuristic technique [6, 15, 10]. In CW, experts walk through the application to detect whether the user interface directs users to act in the expected way [6, 15]. The purpose of the TA method is to understand the work done by users and how the emerging product would affect the users [15] and to learn how people would actually perform specific types of tasks [6]. Two other UIMs are Pluralistic Walkthrough (PW) and Feature Inspection (FI). In PW, scenarios are reviewed and discussed with end users [21]. FI is assessing a proposed feature set for understandability, usefulness, and availability for the user [21]. Newer UIM methods, which can be used e.g. for the development of web applications, can be summarized as the methods that perform usability inspection and problem identification for a given task, aspect or question, such as identifying problematic and confusing links [26].

A number of authors have mentioned the following drawbacks of the existing UIMs: 1) involving experienced experts can be challenging and costly [12, 6, 3]; 2) evaluators need to have a very good knowledge about the intended product and its users, otherwise the inspections can fail [15]; 3) scope and therefore, the results are too limited [31, 30, 3]; 4) too much focused on design artifact and user interface interaction instead of covering wider context of use [31]; 5) the TA method could be too time-consuming and skill dependent [6]; 6) the HE method has a reliability problem [6]; 7) the CW method has too much documentation and limits the problem domain [3]; 8) interpreting the notion of usability too narrowly [31, 9]; 9)most of the methods that are used for inspecting a model, code or prototype focus on the design and implementation phase, which is a later phase of the software development process [26, 9].

In addition to UIMs, user-based testing with early-stage prototypes can also be performed. However, user-based testing at every stage of the project is not always cost-effective or practical [1].

In the literature, studies can be found that have compared with each other the UIM and SUT methods and have identified their overlaps. Khajouei and Farahani [16] combined HE as an UIM method and Think Aloud (TA) as a SUT method to see the potential of identifing usability issues by jointly applying both methods. Resulting from the study, by HE they identified more issues related to user satisfaction, learnability, and error prevention, whereas by the SUT TA method they detected more issues concerned with effectiveness and efficiency. Also, the SUT TA method identified problems with user interaction while HE did not detect these problems. Both methods identified 25% of the same usability issues, and HE and the SUT TA method separately identified 39% and 36% of the usability issues, respectively. Tan et al. [29] likewise compared with each other HE and a SUT method. Their conclusion was that both methods discovered 10% of the same issues, while separately HE covered 60% and SUT 30% of the issues.

3 Experience-based Analysis (EbA) as a Usability Inspection Method: Overview

The EbA method has been created with the purpose of providing requirements engineers with an early indication of whether the solution will function efficiently and identifying potential usability risks across various aspects.

The EbA method was created by applying the Action Design Research method [27] which allows simultaneously to find a solution to a problem and create a tool facilitating the solution of the problem. The preliminary EbA-tool is free to use and can be found online at [23].

The EbA method is based on the premise that people behave according to experiential patterns: if they have behaved in a certain situation and context in a particular way, they are likely to behave in a similar way in other comparable situations [8, 20, 7, 11].

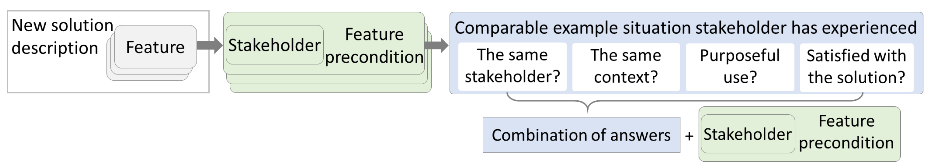

Fig. 1. Experience-based Analysis (EbA) framework [22, 23]

The components of the EbA framework are depicted in Figure 1. All the underlined terms in the following text refer to the terms used in the EbA framework. The process of applying EbA consists of the following steps [22, 23]:

Step 1. The description of the planned solution is divided into features, where each feature represents one or more logically related solution capabilities that provide value to a user and are described by a set of functional requirements [32]. Depending on the peculiarity of the technology to be developed or implemented, or elaboration and the structure of the analysis, these solution descriptions can be statements about what the users can do, user stories, use cases, scenarios, functional goals of goal models, process descriptions or other similar models that describe the planned solution and the intended kinds of users. Hereafter and in Figure 1, these solution descriptions are termed as <u>features</u>.

Example. Let us assume that the platform of an e-service needs to be complemented so that the customers of the e-service could upload files onto the platform instead of sending them separately by e-mail. The product owner of the e-service expresses his opinion as follows: "The IS does not need to limit the format of the files customers upload, because anyway the customers use common software packages. However, the file to be uploaded should be limited to one file and the size of the file can be up to 10 MB." As was stated above, in EbA this kind of requirement is viewed as a feature, which in this case can be formulated as follows: The customers of the e-service can upload one file with no format specified, but a size limit of 10 MB.

Step 2. For each feature it is decided who are the stakeholders involved in the feature or significantly affected by the feature, and what are the essential knowledge, attitudes, skills, awareness, emotions [13], means and activities expected from them within this specific feature, to make this feature efficient. In EbA, such expectations are collected by brainstorming and termed as feature preconditions.

Example. The preconditions for the feature in our running example could be as follows: 1) recipients of the files use the same software package as the customers; 2) the customers are using common software packages; 3) recipients of the files do not need to process the information in the file; 4) the customers can compress files without losing quality of images; 5) the customers know how to include all the information in one file; 6) the customers do not mind including all the information in one file.

Step 3. By using a brainstorming technique, for each feature precondition, the closest comparable example situation should be identified that the stakeholder has experienced in the past and where the same behaviour of the stakeholder appears.

Example. In the running example, each of the six preconditions should be assigned a comparable example situation. For example, precondition 1 is concerned with using software packages, so that the example situation could be: "the recipients of the files have previously opened the attachments sent by the customers via e-mail".

Step 4. It is assessed to what extent the feature precondition and example situation are comparable according to the stakeholders and context, and whether the behavior of the stakeholder was purposeful and what was the satisfaction level of the stakeholder in the comparable example situation. For step 4, the information can be obtained from various sources: published research results and statistics, research articles, opinions left on social media [14], previously collected user behavior and feedback from service providers, information synthesized by the analyst based on public information, and observations in public space. Step 4 is illustrated in Figure 2 by the frame A and consists of the following substeps:

Step 4.1. The same stakeholder? Is the stakeholder of the feature precondition the same as the stakeholder in the comparable example situation?

Step 4.2. The same context? Is the context of the feature precondition the same as the context in the comparable example situation? The context of usage "comprises a combination of users, goals, tasks, resources, and the technical, physical and social, cultural and organizational environments in which a system, product or service is used" [2] and also, for example, time and purpose [22], and if the usage is by novice or experienced users [33]. It is up to the analyst to decide which elements make up the relevant context for the feature preconditions.

Step 4.3. Purposeful usage? How purposeful was the usage of a solution by the stakeholder in the comparable example situation?

Step 4.4. Satisfied with the solution? Was the stakeholder satisfied with the solution in the comparable example situation?

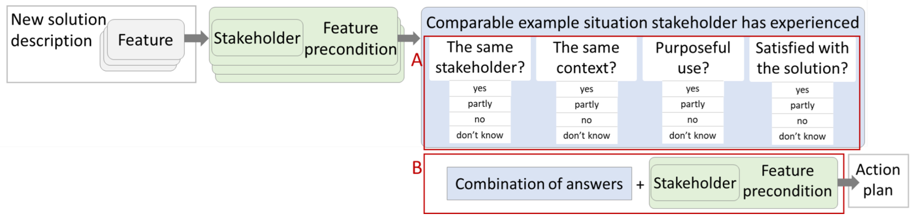

Fig. 2. The EbA framework with the questions and answer options (Frame A) and a combination of the answers (Frame B)

Example. In our example, substeps 4.1-4.4 are applied to the feature precondition number 1 as follows: 4.1.) “Yes”, the stakeholder is the same one who previously opened files sent by e-mail. 4.2.) “Yes”, the context is the same, because the same kinds of files were created and sent by the customers. 4.3.) It was a ”partly” purposeful usage because the recipients of the files claimed they used the Windows operating system, but they occasionally also received files from customers using the IoS operating system that would not open. 4.4.) The answer to the question about satisfaction of the recipients of the files is “no”, because the files that would not open caused them additional work.

Step 5. Based on the answers to the questions asked at steps 4.1-4.4, EbA yields a combined answer as a sentence that is composed of designated words and phrases [4]. Further, combining this answer with the Stakeholder and Feature precondition from step 2, the combination presents a prediction about the usability of the feature on the aspect of the precondition under the analysed circumstances. The combination is illustrated in Figure 2 by the frame B. For example, if the evaluation of the similar situation at steps 4.1-4.4 yields the result ”yes-partly-yes-no”, the combined components will be ”There is dissatisfaction with this *(the core of the feature precondition)*, but in a slightly different context + Stakeholder + Feature precondition”. The Stakeholder and the Feature precondition are the same as what we had in step 2. Based on the findings, the analyst can create an action plan. In particular, the analyst can decide whether the stakeholders would have the expected means and would behave in the way expected by the product designers, or if they are not, how could this be improved?

Example. In our example, after performing steps 1-4, our finding by the EbA method can be expressed as follows: “There is dissatisfaction with this *(using the same software)*, but to an extent, + recipients of the files + are using the same software as clients”. After the remaining five preconditions of this feature have been analyzed, in context of the feature and the aim of the product, the analyst can decide how this dissatisfaction aspect of the solution could be improved. The example is illustrated by Figure 3. A larger scale fictional example, where a solution has been analyzed by EbA, can be accessed online [23].

[4] The full list of all 256 versions of answers can be accessed online [23].

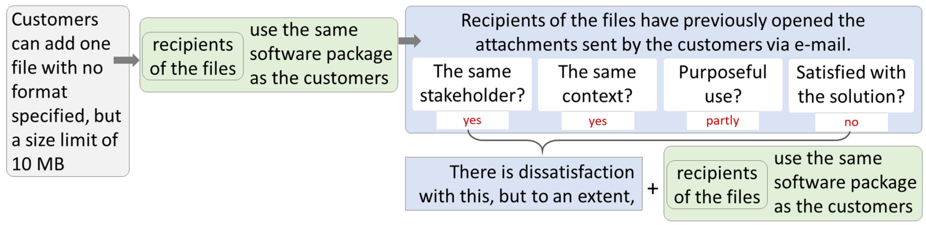

Fig. 3. An example analysed by the EbA framework

The benefits of the EbA method have been described in Section 1. The EbA method has the following limitations [22]:

1. You get an answer to what you ask, both in terms of content and scope;
2. Analysing the usability of features and creating an action plan depends on the skills of the person who is analysing the product under development;
3. The analysis of the usability of features depends on the comparable example situation identified and the ability to adequately evaluate it.

4 Validation

Answering RQ1 leads us to know to what extent the usability information that is obtained with a SUT method could have been discovered already with the EbA method, we compared the usability findings yielded by the SUT and EbA methods on the same complex digital service (”case study” hereafter). For the comparison of both methods, we used the SUT and EbA reports of the case study. The SUT was performed by the team of the University of Jaén [25]. The EbA was performed by the first and second author of this paper. The information obtained by either method was analyzed by the Content Analysis (CA) method [17] and then compared to each other. We chose the CA method because CA enables to determine the presence of certain themes within qualitative data. The first author of this paper defined the relevant categories, applied the CA method to identify usability findings, analyzed the discovered information and compared the findings yielded by the SUT and EbA.

Answering RQ2 leads us to know whether the usability findings obtained by EbA are valuable for product owners. The value of the findings was assessed by two stakeholders of the digital service. One of the evaluators, who is also the second author of this paper, is the product owner of the digital service of the case study.

4.1 The Case Study

The case study was chosen from the industry-oriented research and development project Pharaon[5]. The Pharaon project consists of five large-scale case studies

[5] https://www.pharaon.eu/

or pilots the aim of which is to develop an ecosystem of technological solutions to improve the quality of life of older adults and their caregivers. To answer the research questions, the pilot from Andalusia in Spain was chosen because in addition to applying SUT for validation they were also ready to apply EbA. In the Pharaon project, digital services for older adults and their caregivers integrated in the project were evaluated by SUT. The EbA method is meant to be applied before the SUT. Differently, in this validation these methods were applied in the reverse order because the author of the EbA method joined the Pharaon project after the SUT had already been completed. However, we designed the activities of the validation process in such a way that the credibility of the results would be ensured. These activities are described in Section 6.1.

4.2 Data Collection

Usability analysis by SUT The SUT was performed by the team of the University of Jaén [25]. The aim of the SUT was to highlight bugs, potential bugs, and assess the usability. The team collected feedback about the usage of the digital services during test-sessions in real-life situations according to the requirements represented as use case scenarios (UCS) [28]. Altogether 25 test-scenarios were executed, involving four different technologies provided on one platform. To assess the usability feedback by the end-users, the System Usability Scale (SUS) [5] and After-Scenario Questionnaire (ASQ) [18] interview were applied [25].

Usability analysis by EbA The first author of this paper familiarized herself with the Pharaon project by reading the requirements documents of the project. Based on the information acquired from the documents, she considered UCSs as *features* that were used for representing requirements in the Pharaon project. When applying EbA, the first author was not aware of the results of the validation testing by SUT. After she described preconditions and example situations by following the principles of the EbA method as is described in Section 3, the second author of this article joined who, being the product owner of the analyzed pilot, knew the local context and services under development. To obtain a sufficient number of usability findings for comparison, under the guidance by the second author, seven features were extracted for comparative analysis that had sufficient amount of usability-related information obtained by SUT. The first author identified 48 feature preconditions for the seven features [24]. These feature preconditions were analyzed together by the first and second author by means of EbA. It is important to emphasize here that neither of the authors is a usability expert by profession. The second author as the project owner was aware of the test report results but she did not use this knowledge when analyzing the feature preconditions. Based on the outcome, the first author of this paper devised an action plan describing which features look fine and what design decisions should be made in the future, which parts of features should be further tested, and which support activities should be planned. In total, each of the 48 preconditions analyzed by EbA led to one usability finding.

4.3 Data Analysis

Before starting the process of data analysis, the relevant data categories for the analysis were defined. The categories and rationale for the selection are explained below for each step of the analysis. The data analysis involved at first extracting the relevant information from the SUT report, and then categorizing the usability findings obtained by SUT and EbA into the appropriate categories according to the CA method [17]. After that, the categorized information obtained by both methods was compared. The process consisted of the following steps:

1. Extracting usability findings from the SUT report representing the seven UCSs considered.

2. Categorizing the usability findings from applying the SUT and EbA into the categories "usability-related issues" or "bugs", depending whether a finding was concerned with an extensive usability issue ("usability" hereafter), or with a technical issue ("bug" hereafter) affecting usability. This categorization is necessary for two reasons: (i) to exclude technical issues from the analysis because identifying technical issues is not the purpose of an UIM; (ii) to know how well suited EbA is also for discovering technical issues.

3. Identifying overlaps between the usability findings by SUT and EbA. To find out to what extent EbA can obtain the same information that is otherwise obtained by SUT ("overlap" hereafter), we should first define what do we mean by "the same information". We evaluate this based on two indicators: object and content. The first indicator expresses whether the overlapping findings by both methods address the same object. The object is something that the finding is focused at, e.g. action or feeling. The second indicator expresses whether the content of the conclusions drawn by EbA matches the content of the conclusions drawn by SUT. For example, if a finding by either method identifies that older adults can get stressed because of the fear of not coping with using the technology, the object of both methods is "getting stressed". At the same time, the content of the conclusions by both methods is also the same: "older adults can get stressed". This means that in the given example there was a "full overlap" between the findings by object and content. Identifying the overlaps by object and content allows us to be more specific as to whether the EbA method leads to focusing on the same issues that are otherwise discovered by SUT and does EbA lead to the same conclusions than SUT?

4. Categorizing the usability findings and action plan suggestions from applying SUT and EbA into the categories "technology-oriented" and "non technology-oriented". "Non technology-oriented" findings allow the discovery of usability issues of a product in a broader sense.

5. Categorizing the usability findings by SUT and EbA into the category "need for action" based on what action plan was decided as a conclusion. The motivation for such categorization is preventing misinterpretation of the results, since UIM and SUT reports traditionally focus on discovering usability problems. However, EbA also yields a significant amount of information as to whether the feature precondition is correct, i.e "it works". If the EbA analysis indicates that the feature precondition is met, it increases confidence to continue with the development of the given feature. The remaining categories enable understanding of the advantages of knowing findings from applying EbA before findings from applying SUT. If a conclusion from applying EbA states that "the technology needs to be complemented" or "the solution does not work this way which means that a different solution should be worked out", early reacting to the situation will be enabled which may avoid re-doing the development process. If "the feature needs a support action" to enables its full value for the product, support actions and resources can be planned at an early stage of the development process. If a finding identifies a critical issue on which we do not have enough information, the issue "needs to be further analyzed or tested" by SUT. This information helps us to plan the test scenarios for the SUT.

6. Categorizing EbA findings into the category "value". The assessment by "value" clarifies whether the usability findings obtained by EbA are valuable for the product owner and other stakeholders. At this step the product owner of the Andalusian pilot, who is also the second author of this paper, along with another member of her team checked all 48 findings identified by EbA and determined their value for the team. They evaluated the usability findings by EbA simultaneously with evaluating the action plan pertaining to these findings proposed by the first author of this paper based on the results of applying EbA. Based on that, all 48 findings by EbA were categorized by "value" into one of the following categories: 1) Important, dealing with it; 2) Important, but cannot deal with it now; 3) Good to know; 4) Irrelevant knowledge; 5) We knew that anyway. The first three categories indicate valuable information, while the remaining two categories indicate the information that was irrelevant or already known.

The original data of the case study can be made public to a limited extent [24], due to the conditions set by the project contract. The analysis of seven features of the case study is available online [24]. The example includes 48 preconditions and comparable example situations. The example also includes categorizations and value evaluations of the findings and suggestions for the action plan.

Based on the information obtained by the data analysis, the research questions RQ1 and RQ2 are answered in Sections 5 and 6.

5 Results

With respect to the research question RQ1, we found that most of the usability information that is obtained with the SUT method can be discovered also with the EbA method. As is reflected by Figure 4, from among the 14 findings by SUT, 11 were also identified by EbA, out of which five represented the same object and six partially the same object. The overlap by content between SUT and EbA was similar: five of them drew the same and six partially the same conclusions. There were no situations where the conclusions reached by SUT claimed something totally different from the conclusions reached by EbA.

We also found that in comparison with SUT, the EbA method resulted in quantitatively more and broader usability information, but did not identify any "bugs". In particular, while SUT yielded 14 findings in the "usabil-

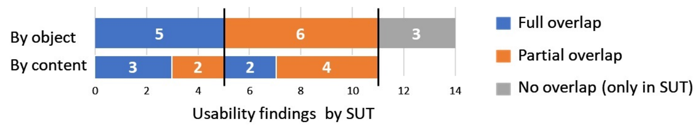

Fig. 4. Overlap of findings by object and content

ity" category and 8 findings in the category "bugs", applying the EbA method resulted in 48 findings in the "usability" category and no findings in the category "bug". The "non technology-oriented" aspects were examined as an indicator of broader usability information. The analysis of feature preconditions with EbA provided more information on "non technology-oriented" aspects than "technology-oriented" aspects (67% vs 33%), testing with SUT was the opposite (44% vs 56%). The findings on overlaps belonged to both aspects. The overlap between "technology-oriented" and "non technology-oriented" feature categories is represented in Figure 5.

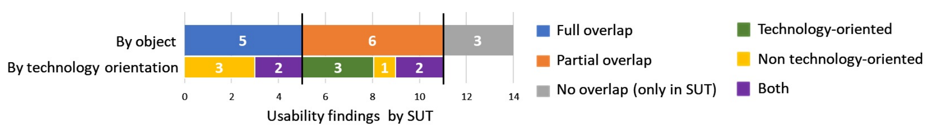

Fig. 5. Overlap of findings in the "technology-oriented" and "non technology-oriented" categories

Both methods provided findings showing what does work in the product, what aspects do not work, and what needs additional support action or change in technology. Figure 6 reflects the balance of the results between the two methods.

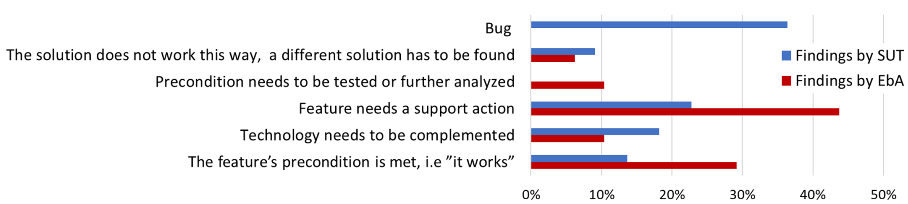

Fig. 6. Percentage distribution of findings in the category "need for action"

With respect to the research question RQ2, we found that more than half of the usability information gathered by EbA – 53% – was evaluated by the product owner and her colleague as valuable, which is also reflected by Figure 7.

	Is valuable 53%			Is not valuable 47%	
	Important, dealing with it	Important, but can't deal with it now	Good to know	Irrelevant knowledge	We knew that anyway
Findings	14	2	9	6	16

Fig. 7. Categorization by value of the findings by EbA

6 Discussion

Overall, we can conclude from our study that the EbA method is capable of providing useful usability information in an early phase of development. The information includes findings that are otherwise obtained with a SUT method.

In summary, we answer the research questions (RQs) that were posed in Section 1 as follows:

- **RQ1:** The usability information obtained by EbA largely overlaps with the usability information obtained by SUT. In our case study, from among the 14 usability findings by SUT, 11 were also identified by EbA. In comparison with SUT, the EbA method resulted in quantitatively more and broader usability findings.
- **RQ2:** Despite its quantitative abundance, the usability information obtained by EbA can be considered to a large extent as valuable for product owners. More than half of the information about the user behavior gathered by EbA – 53% – was evaluated by the product owner and her colleague as valuable.

According to the research literature on UIMs, UIMs offer a greater value if they are used as complementary methods to SUT [1, 12, 15, 29, 16]. This statement also holds in our study, where the analysis by EbA helped discover what usability information was missing, so that the requirements engineering team could select in advance which usability cases had to be tested additionally by SUT or further analyzed.

In Section 2, we referred to the two studies where Heuristic Evaluation (HE) as an UIM method was compared to a SUT method [16, 29]. When comparing our results with those two studies, it must be pointed out that compared to HE, EbA discovered a lot more usability information that is usually discovered only by SUT. Table 1 shows the comparison of the validation results by HE with the validation results by EbA.

	Only in UIM	**Overlap**	**Only in SUT**
UIM HE [16]	39%	25%	36%
UIM HE [29]	60%	10%	30%
UIM **EbA**	72%	22%	6%

Table 1. Validation results by HE compared to validation results by EbA

According to the research literature, UIMs can detect more general usability problems than SUT [10, 29, 16]. In our case study, EbA provided quantitatively more information than SUT. This is presumably due to the nature of either method. In particular, SUT involves specific test scenarios to be performed by the end users testing the service, which may not cover the entire scope of the features included by the scenarios. On the other hand, by presenting all the features of the solution chosen by the analyst, EbA directs the analyst to thinking about each feature separately as a whole with respect to its behavioral preconditions. This explains a larger number of usability findings by EbA compared with SUT. Another important aspect is that while SUT can test the features that exist in the prototype at the given moment, EbA provides an opportunity to analyze also those features that have not yet been included in the prototype. In addition, while SUT by definition considers the technology to be used, EbA does not consider it. Therefore, EbA can be used for obtaining usability information also in cases where technological solutions have not yet been fully decided.

The analysis by EbA takes 1-2 working days per project depending on the volume of the planned product and the level of detail of the analysis [22]. Practitioners can decide whether spending 1-2 working days on EbA can avoid some costs that are otherwise spent on prototyping and testing with end-users.

6.1 Threats to Validity

A threat to internal validity is concerned with the scope of the study. Seven use case scenarios (UCS) from one case study were analysed, which is sufficient to draw initial conclusions, but certainly the scope needs to be expanded by subsequent studies.

Another threat to internal validity is concerned with the person who was involved in the validation process. In this study, the same person who invented the EbA method defined the relevant categories, applied the Content Analysis (CA) method to identify usability findings, analyzed the discovered information, and compared the findings yielded by SUT and EbA. The quality of the CA and comparative analysis was ensured by two follow-up data checks after two months' intervals. Next studies need to involve more people for quality control. The same threat to internal validity also comprises how much usability findings this one person was able to discover from the SUT report. Results from SUT were included if the SUT report contained comments explaining the strengths and shortcomings of the tested UCSs. The SUT report of the case study also includes the results from conducting the SUS and ASQ questionnaires. However, these results were difficult to incorporate in analysing usability aspects because they yielded numerical values which are hard to employ for concluding which usability aspects should be improved. For excample, if the SUT report includes a test scenario "Make a videocall" and according to the results of conducting the ASQ questionnaire, the evaluation of this functionality is 5.8 without any further comments on what was missing or what was positive, that information was too general to apply for identifying any usability issues.

The third threat to internal validity was that the SUT results were presented before the EbA analysis for the "value" evaluations of the findings was made. To prevent the risk of using the SUT test report results for evaluating the case study features with EbA, we used the interview format where the first author of this paper, who was not aware of the SUT results, asked questions in a directed manner so that finding answers for validation would be outside of the scope of the test-report. She also made sure to re-rail the discussion when the thought of the respondents seemed to drift to the test-report.

Knowing the SUT results can cause one more threat to internal validity as the value of the findings from applying EbA were evaluated by two people from the Andalusian pilot project of the case study, who were aware of the results of the previously accomplished validation by SUT. However, this did not decrease an advantage of applying EbA, because according to the evaluators, if the EbA analysis and evaluation of the findings had been completed before rather than after SUT, several "we knew that anyway" findings would rather have appeared as "important, dealing with it" findings. This would have increased rather than decreased the share of the "valuable" findings by EbA, which are the findings belonging to the first three categories represented in Figure 7.

A threat to external validity is the generalizability of our results, considering the particular case study from the Pharaon project that we chose for the validation. We will mitigate this threat by applying the EbA method and conducting the same kind of comparative analysis for other digital products and services.

7 Conclusions

This paper introduced a new UIM that addresses the following drawbacks of the existing UIMs: (i) finding experts can be complicated; (ii) the results are too limited; (iii) there is the lack of UIM techniques that would allow for early usability analysis before prototyping. This paper also presented the results of the evaluation where the EbA method was compared with a classic SUT method with the aim to find out to what extent EbA yields the same findings that is obtained by SUT and whether this information is valuable for the product owner.

Based on the results of the validation, we can conclude that EbA is providing extensive usability information already in an early phase of requirements engineering when prototyping and SUT have not yet been planned. Despite a large number of findings, slightly more than a half of the usability findings obtained by the EbA method were considered as valuable by the two evaluators from the Andalusian pilot.

With EbA, a considerable number of the same usability findings were identified as with SUT and the usability information obtained by both methods partially or fully overlapped. Therefore, we can conclude that EbA reduces the workload of the usability testing by SUT by detecting usability issues before SUT and this way can make SUT more cost-effective. Consequently, EbA should be considered as a complementary method to SUT, because EbA cannot detect all of the usability issues.

Important areas for our future work are the following ones: 1) To investigate if the EbA method yields false positive or false negative results. 2) To perform a larger scale validation of EbA by applying it to different types of digital products and services and by various stakeholders.

8 Acknowledgements

The research work reported in this paper has received funding from the Pilots for Healthy and Active Ageing (Pharaon) project of the European Union's Horizon 2020 research and innovation programme under the grant agreement no. 857188. The authors are expressing their gratitude to María Parraga Vico from the Ageing Lab Foundation and the University of Jaen from Spain.

References

1. International organization for standardization: Ergonomics of human-system interaction-part 210: Human-centred design for interactive systems. Standard ISO 9241-210:2019
2. International organization for standardization: Ergonomics of human system interaction: Usability, definitions and concepts. Standard ISO 9241-11:2018
3. Aamir, M.J., Mansoor, A.: Testing web application from usability perspective. In: 2013 3rd IEEE International Conference on Computer, Control and Communication (IC4). pp. 1–7. IEEE (2013). https://doi.org/10.1109/IC4.2013.6653765
4. Bill, A., Tom, T.: Measuring the User Experience : Collecting, Analyzing, and Presenting Usability Metrics. Interactive Technologies, Morgan Kaufmann (2013)
5. Brooke, J.: Sus: a retrospective. Journal of usability studies **8**(2), 29–40 (2013)
6. Cheng, L.C., Mustafa, M.: A reference to usability inspection methods. In: International Colloquium of Art and Design Education Research (i-CADER 2014). pp. 407–419. Springer (2015)
7. Clark, A.: Whatever next? predictive brains, situated agents, and the future of cognitive science. Behavioral and brain sciences **36**(3), 181–204 (2013)
8. Fazio, R.H., Zanna, M.P.: Direct experience and attitude-behavior consistency. In: Advances in experimental social psychology, vol. 14, pp. 161–202. Elsevier (1981), https://doi.org/10.1016/S0065-2601(08)60372-X
9. Fernandez, A., Insfrán, E., Abrahão, S.: Usability evaluation methods for the web: A systematic mapping study. Information and Software Technology **53**(8), 789–817 (2011). https://doi.org/10.1016/j.infsof.2011.02.007
10. Helander, M.G., Landauer, T.K., Prabhu, P.V., (Eds.): Handbook of human-computer interactio. Elsevier Science Technology (1997)
11. Hohwy, J.: The self-evidencing brain. Noûs **50**(2), 259–285 (2016)
12. Hollingsed, T., Novick, D.G.: Usability inspection methods after 15 years of research and practice. In: Proceedings of the 25th annual ACM international conference. pp. 249–255 (2007), https://doi.org/10.1145/1297144.1297200
13. Iqbal, T., Anwar, H., Filzah, S., Gharib, M., Mooses, K., Taveter, K.: Emotions in requirements engineering: A systematic mapping study. In: 2023 IEEE/ACM 16th International Conference on CHASE. pp. 111–120. IEEE (2023). https://doi.org/10.1109/CHASE58964.2023.00020

14. Iqbal, T., Khan, M., Taveter, K., Seyff, N.: Mining reddit as a new source for software requirements. In: 2021 IEEE 29th international requirements engineering conference (RE). pp. 128–138. IEEE (2021)
15. Jacko, J.A.: Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications, Third Edition. CRC Press, Inc. (2012)
16. Khajouei, R., Farahani, F.: A combination of two methods for evaluating the usability of a hospital information system. BMC Medical Informatics and Decision Making **20**(1), 1–10 (2020), https://doi.org/10.1186/s12911-020-1083-6
17. Krippendorff, K.: Content analysis: An introduction to its methodology. Sage publications (2018)
18. Lewis, J.R.: Psychometric evaluation of an after-scenario questionnaire for computer usability studies: The asq. SIGCHI Bull. **23**(1), 78–81 (jan 1991). https://doi.org/10.1145/122672.122692
19. Mack, R., Nielsen, J.: Usability inspection methods: report on a workshop held at chi'92, monterey, ca, may 3-4, 1992. ACM SIGCHI Bull. **25**(1), 28–33 (1993)
20. Murray, K.B., Häubl, G.: Explaining cognitive lock-in: The role of skill-based habits of use in consumer choice. Journal of Consumer Research **34**(1), 77–88 (2007). https://doi.org/doi.org/10.1086/513048
21. Nielsen, J.: Summary of Usability Inspection Methods. https://www.nngroup.com/articles/summary-of-usability-inspection-methods/ (1994), [Online; accessed 1-Oct-2023]
22. Piirisild, A.: Analogy-Based Technology Effectiveness Prediction Model and Tool. Master's thesis, University of Tartu (2021), http://hdl.handle.net/10062/72631
23. Piirisild, A.: Tool for the experience-based analysis (eba) method (2024), http://dx.doi.org/10.23673/re-453
24. Piirisild, A., Gómez, A.P.: Experience-based analysis method - the evaluation data (2024), https://datadoi.ee/handle/33/595
25. Polo-Rodríguez, A., Dionisio, P., Agnoloni, F., Gómez, A.P., Paggetti, C., López, L.G., Lendínez, A.C., Espinilla-Estévez, M., Medina-Quero, J.: Challenges of ubiquitous and wearable solutions to address active ageing in the andalusian community. Journal of Universal Computer Science **28**(11), 1221 (2022)
26. Rivero, L., Barreto, R., Conte, T.: Characterizing usability inspection methods through the analysis of a systematic mapping study extension. CLEI Electronic Journal **16**(1), 12–12 (2013)
27. Sein, M.K., Henfridsson, O., Purao, S., Rossi, M., Lindgren, R.: Action design research. MIS quarterly pp. 37–56 (2011)
28. Sterling, L., Taveter, K.: The Art of Agent-Oriented Modeling. MIT Press (2009)
29. Tan, W.s., Liu, D., Bishu, R.: Web evaluation: Heuristic evaluation vs. user testing. International Journal of Industrial Ergonomics **39**(4), 621–627 (2009). https://doi.org/10.1016/j.ergon.2008.02.012
30. Tarkkanen, K., Harkke, V.: Scope of usability tests in is development. AIS Transactions on Human-Computer Interaction **11**(3), 136–156 (2019). https://doi.org/10.17705/1thci.00117
31. Tarkkanen, K., Harkke, V., Reijonen, P.: Testing the unknown–value of usability testing for complex professional systems development. In: Human-Computer Interaction–INTERACT 2015: 15th IFIP TC 13 International Conference, Part II 15. pp. 300–314. Springer (2015). https://doi.org/10.1007/978-3-319-22668-224
32. Wiegers, K.E., Beatty, J.: Software Requirements. Pearson Education (2013)
33. Wu, J., Du, H.: Toward a better understanding of behavioral intention and system usage constructs. European Journal of Information Systems **21**(6), 680–698 (2012). https://doi.org/10.1057/ejis.2012.15