Master Thesis

Software Engineering

Thesis no: MSE-2011:78

Usability and user experience: measurement model

Eriks Klotins

School of Computing

Blekinge Institute of Technology

SE-371 79 Karlskrona

Sweden

This thesis is submitted to the School of Engineering at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Software Engineering. The thesis is equivalent to XXX weeks of full time studies.

Contact Information:

Eriks Klotins

Address: 123-46 Lubanas St. LV-1021, Riga, Latvia

E-mail: eriks.klotins@gmail.com

University advisor:

Dr. Tony Gorschek

Department of Systems and Software Engineering

School of Computing

Blekinge Institute of Technology

SE-371 79 Karlskrona

Internet : www.bth.se/com

Phone : +46 455 38 50 00

Fax : +46 455 38 50 57

Abstract

Software quality is one of success factors in software development. Usability and user experience (U&UX) as a part of software quality is becoming more and more important. Although, there has been successful attempts to formalize specific parts of U&UX there is still a need for a systematic framework of U&UX evaluation.

This thesis is aimed to study the state of the art in U&UX evaluation in order to develop a single framework that comprises existing knowledge on the topic. Furthermore, the U&UX evaluation framework is aimed to support product development in industry and provide a versatile guide for U&UX practitioners.

The study is based on reference based systematic review. The literature review covers both scientific publications and industrial grade papers. The papers to be reviewed were selected by their relevance to the study goals and credibility of the source.

The result of this is three layer U&UX evaluation framework. First layer of the Model features breakdown structure of usability and user experience. Total number of usability and context of use attributes is 217. Second layer of the model contains guidelines of how to perform usability evaluation. Third layer features validation strategies and guidelines on how to expand the Model. In order to enable practical use of the Model both static and dynamic validation should take place.

There are many models in place attempting to formalize U&UX evaluation. However, most of them focuses on particular branch of usability or are too broad to be applied practically without adaption. Furthermore, there are many resources offering practical usability and user experience checklists or guidelines. However, most of them lack connection with industry standards such as ISO/IEC 9126. The Model presented in this thesis attempts to fill the gap between high level industry standards and cook book style U&UX guidelines.

Keywords: Usability, evaluation, model, guidelines

Table of contents

Table of figures

1 Introduction

Product quality is one of the main concerns in any development process. Products are naturally compared and evaluated according to their quality; one of most important quality of a product is how well it fits its purpose and how easy is to use it. Such characteristic of a product is covered by usability – a product feature that allows users easily understand and use the product and has been an increasingly seen as a key feature in design.

In early days of computing, computers were operated by trained professionals and the focus was set on developing better functionality. As the computer technology developed, it became a key to develop more intuitive systems that requires less user training and produces more output on the same time. [7] (Douglas, 2006) Furthermore, as computer systems entered a consumer electronics market, user satisfaction became a point of competition.

A consumer purchasing a product is driven by multiple forces. For example consumer has needs to get certain tasks done and has certain expectations on how well the software product will help him. Furthermore, user has background on similar software and he/she wants to reuse already present skills. Besides obvious practical needs of a user, the product may influence other less practical although important needs, for example a need for social status or fun of using the product.[51] User experience embodies a set of user‘s emotions around the product. Thus, to address such need for positive emotions companies are struggling to attract customers by offering products that offer better user experience. This can be achieved by consciously designing user‘s emotions. (Csikszentmihalyi, Harper and Row, Roto)

There have been many attempts to formalize usability and user experience however a comprehensive model is yet missing. Existing models provide in depth focus on particular sub domains of usability however not all domains of usability are covered.

This thesis attempts to study the structure of usability and user experience in order to create an evaluation model to support decision making in software development industry.

1.1 Structure of the thesis

This thesis has six sections. Introduction and Background introduces user into the topic and the background of problem domain. Research methodology defines research questions and describes the applied research methods and discusses a validity threats. Model development discusses the study process itself and shows how exactly the results are produced. Results section outlines the results of this thesis. Last section contains discussion on the results and ideas for further work on usability and user experience field.

2 Background

This section provides the background knowledge on the topics discussed in this thesis. First part provides an overview of user experience and usability evolution goals. Second part provides an introduction in existing usability evaluation methods. The third part covers existing studies on user experience.

2.1 Goals of usability and user experience evaluation

Usability and user experience deals with product quality in use. Understanding the nature of usability and user experience enables to drive product development to achieve desired levels of quality. The goal of usability and user experience evaluation is to acquire a credible input for product development. [1, 13]

2.2 Usability evolution methods

Usability is a long known phenomenon in development industry. First attempts to analyze usability of a product occurred on the time of WWI when aircraft engineers considered the designs of pilot dashboards with intention to identify critical components and make them easier to access and therefore prevent disaster. [63]

Nowadays software is used among many fields and by nearly anyone thus usability becomes an issue to be addressed by software engineering.

In software engineering domain there are two general approaches on usability evaluation. These approaches are checklist based evaluation and attribute based evaluation.

Checklist based approach uses checklists (also called heuristics) and expert opinion to perform usability evaluation. As the name of method suggests it is based on predesigned checklist and one or more experts that provides their opinion about the product based on checklist items. The group of experts may be mixed with developers, actual product users or other stakeholders. [14]

Predefined checklists for different types of products are available from case studies. For example, J. Nielsen has compiled a list of top 10 usability heuristics [16]. Also, a formal standard or user interface design guidelines can be used as a checklist.

Depending on importance of usability evaluation this method can be applied in more or less formal way. The checklist can be prepared with more or less formal way as well as group of experts can act in more or less organized way.

The usability practitioners suggest that the evaluation steps should be performed in systematic manner. Systematic steps can be as follows: definition of product‘s features, identification of critical features, preparation of the checklist, individual walkthrough of the checklist, group discussions and preparation of final report. [31, 32]

Checklist based approach is mentioned as simplest and less resource demanding. [14, 18] Although, the quality of the evaluation results depends on the experts experience. Furthermore, results from multiple evaluations might be hard compare due to different kinds of checklists used and hard to quantify expert opinions. [14]

Attribute based approach compared to the previous method is more mature and provides more objective results. Also this method is more complicated and resource demanding.

This type of method is introduced by ISO/IEC 9126 standard and further developed by independent scholars. The idea behind attribute based approach is to break down the usability to its attributes and then apply experimental methods to evaluate these attributes. The overall usability of a product is determined by a sum of the values from distinct attributes.

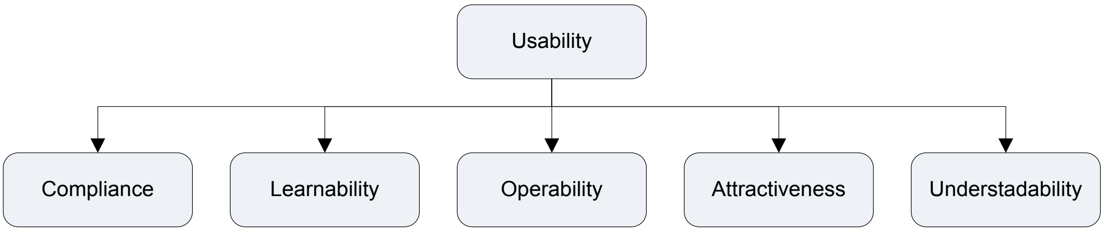

Figure 1, example of usability breakdown structure from ISO/IEC 9126.

Besides ISO/IEC 9126 standard which provides general breakdown structure of usability and basic measures for each of the attributes, there are other more specific models. Furthermore, attribute based approach states that usability cannot be evaluated out of context and analyzes also user and environment characteristics too. Thus, usability attributes are a compilation of a product‘s, users and environment attributes. [18]

MUSiC (Measuring Usability in Context) method is designed to assess the performance related attributes of a product.

The essence of MUSiC method is to observe a subject in simulated environment while the subject (or a group of subjects) is using the product. By analyzing how the user performed (how much mistakes did he made etc.) is possible to quantify specific usability attributes.

This method extends the usability structure by a set of new user performance related attributes and proposes tools and methods to assess these attributes. Tools include video recording and other user observation equipment. The methods describe systematic steps on how to perform usability evaluation. [20]

The systematic steps include definition of the product, its users and environment, identification of critical usability attributes, preparation of the evolution environment, running the tests and analyze the results.

QUIM (Quality in Use Integrated Map) attempts to quantify and organize usability attributes with intention to create application independent usability ontology [31]. Their attempts have resulted in software tool (QUIM Editor) to hold usability attribute structure, their descriptions and other related information. However, it is not clear on how to acquire the editor and whether it can be further developed.

SCANMIC method focuses on analyzing website specific usability attributes. The method describes new and website specific attributes as well as provide the five step website usability evaluation method.

The evaluation method is simpler than one provided by MUSiC, however has the same structure. The evaluation steps are as follows: Decide what to analyze, decide level of analysis, decide on the scale of measures, perform evaluation and analyze the results. [32]

2.3 Studies on user experience

User experience in comparison with usability is a yet a fuzzy concept. [51 - 54] Nevertheless, there have been attempts to define and explore it. This section describes current state of the art in user experience studies.

As defined by ISO 9241 standard, user experience is a set of emotions that arises from user‘s interaction with a product.

V.Roto from NOKIA Group has presented his view and experience on user experience evaluation. The author states that user experience is influenced by multiple factors – building blocks [50].

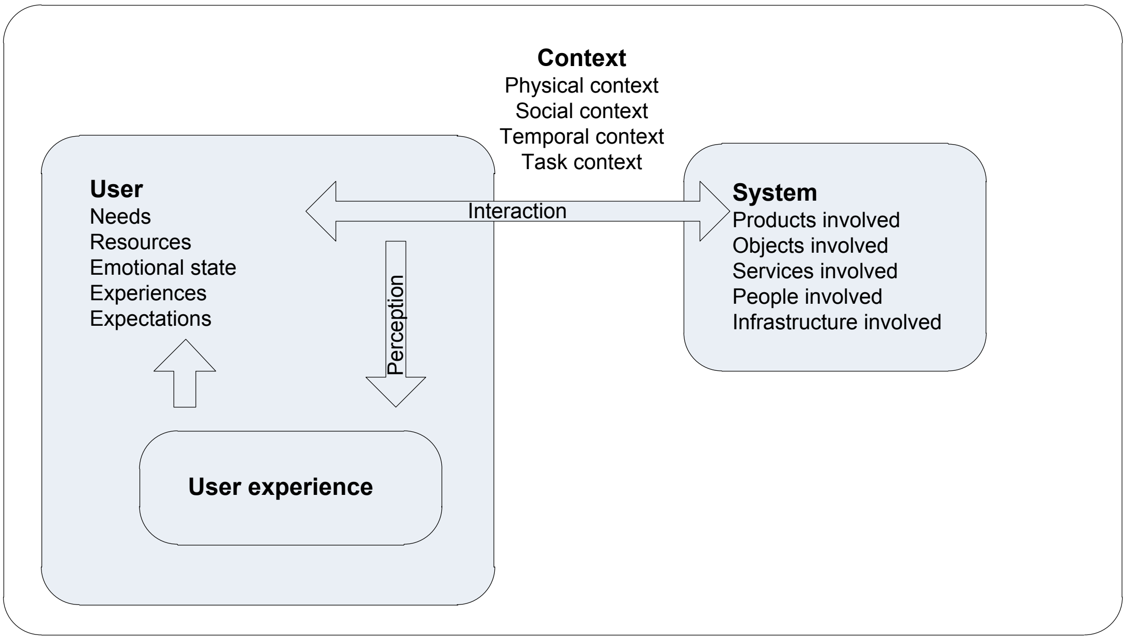

Figure 2, user experience building blocks by V. Roto [50]

The author decomposes each of the building blocks and explains details of each block. The System block refers to a wider scope than the system being used. For example, user experience of a webpage is influenced by webpage itself, browser and network connection who deliver the content.

Context block refers to different types of contexts. Physical context refers to physical surroundings of a user, social context refers to the expectations and influence from other people. Temporal context refers to contextual restrictions to a user or task. For example, use mobile browser to find out which bus to take before missing it. Task context refers to a higher goal of a user instead of direct outcome from use of the product. For example, use smart phone capabilities to review a document and send feedback while on the traffic jam. In this case the higher goal is to get the work done, not ability to open an email with an attachment.

User block refers to user‘s internal state. This comprises user‘s expectations, needs, mood and previous experiences as well as available mental resources. User‘s mood and emotions can easily improve or ruin the overall user experience. Low expectations create a space for a positive surprise. Available mental resources refer to user‘s ability to locate resources to perform the task. For example, switch to another CD track while focusing on driving a car in heavy traffic.

V. Roto mentions that user experience has time dimension. When interaction with a product is short (e.g. pressing an elevator button) the user experience is created by sensation of the moment. However, in long term interactions context factors blurs out and context loses its significance. [50]

Youn-kyung Lim et al. describes a study on user experience in which the authors attempted to identify aspects of user experience. The study analyzes user‘s perception on their favorite gadgets. [51]

They have indentified six product attributes that influences user experience. [51]

1. **Interaction** – qualities that enable people‘s interactions with the product, such as interfaces.
2. **Visual** – qualities that people can see such as colors, sizes, shapes, visual material qualities.
3. **Tactile** – qualities that people can feel by touching or grabbing such as weight, texture, etc.
4. **Content** – contents that are carried, accessed, or delivered by the product such as music, news, video, games, etc.
5. **Function** – capabilities and functionality such as playing music, calling to someone, taking pictures, etc.
6. **Performance** – qualities of how well the product performs the expected functions such as resolution, sound quality, speed, etc.

The study results suggest four types of emotions that drive user experience. The visceral level of emotion is emotional responses formed by physical senses such as ―looks nice,” ―feel cold,” etc. The behavioral level of emotion that is formed from cognitive processes such as ―easy to use,” ―simple to use,” ―hard to figure out,” etc. Usefulness - an indirect quality related to their overall needs rather than a certain specific type of emotional response, such as ―fulfills what I need,” ―practical”, ―does what I want it to do,” etc. The reflective level of emotion such as ―it is a trend,” ―creates an artificial world,” ―everyone has these products,” etc. [51]

The authors have attempted to determine the importance of each emotion and concluded that reflective emotion is more important over others.

3 Research methodology

This section describes a research methodology used in this thesis.

3.1 Problem statement

The need of a complete usability and user experience model is initiated and motivated by development of a prototype for Quuber. The concept behind Quuber is to introduce new way users access the Web with emphasis on innovative user interface.

Initial attempts to identify critical attributes of usability and user experience showed the following problems:

Problem I - software usability is defined by ISO standards who gives broad overview of the field, however, the standard needs to be adapted before it can be applied of a specific product in industry setting [14, 18].

Problem II - there is plenty of usability and user experience guidelines available made by usability practitioners. However, many of them lack reliability or connection with ISO standards. [14, 18, 19]

Problem III - lack of guidelines on how to perform usability and user experience evaluation

Problem IV – usability and user experience often comes hand in hand, however there is a large gap between these two concepts. User experience comparing to usability is yet a fuzzy concept that is not described by any industry grade paper. [50, 51]

3.2 Research questions

This thesis attempts to perform a study on usability and user experience and deliver at least partial solution to the identified problems. The study begins with a literature review to identify existing methods on usability and user experience evaluation. The rest of study focuses on compiling existing methods into a comprehensive model. The goal of the model is to provide a framework of usability and user experience evaluation in industrial setting. Research questions

The purpose of this study is to investigate current state of the art in usability and user experience evaluation in order to develop a usability and user experience evaluation framework for software engineering domain. The investigation aims to study existing evaluation methods and find a common background for usability and user experience evaluation.

In order to fulfill the goals of this thesis the following objectives must be achieved:

- Obtain a deep knowledge on usability and user experience evaluation methods

- Identify pros and cons in current evaluation methods
- Propose a evaluation methodology that at least partially solve the identified problems

Further research is guided by goals which are set in a form of a research questions. The purpose of the study is to answer these research questions. After each question a research sub goals and expected outcomes are explained. Sub goals and outcomes are used later in this study to measure the degree in which general goals are achieved.

The research questions are:

RQ1: What are current views upon usability and user experience?

The purpose of this goal is to establish a solid foundation of knowledge in the topic for further studies. The sub goals are the following:

RQ1.1: What is a definition for usability and user experience?
RQ1.2: What are the attributes of usability and user experience?
RQ1.3: What is the connection between usability and user experience?

RQ2: What is the current state of art of usability and user experience evaluation?

The purpose of this goal is to identify current methods for usability and user experience evaluation. The sub goals are as following:

RQ2.1: What are the existing standards, frameworks and models for usability evaluation?

RQ2.2: What are the features of existing usability and user experience evaluation methods?

RQ3: What are the guidelines to apply, validate and adapt usability and user experience validation strategies?

The purpose of this goal is to identify current usability evaluation guidelines.

RQ3.1: What are the guidelines for usability and user experience evaluation?
RQ3.2: What are the methods for usability and user experience model validation?

3.3 Research scope

In this thesis usability and user experience is analyzed from the perspective of software engineering. The developed model is the compilation of existing knowledge of usability and user experience evaluation. The literature review is limited to academic publications and industrial grade papers.

3.4 Validity evaluation

The fundamental concern about any study is validity of the results. It is important to consider validity issues of a study already in planning phase. [Wohlin et al, 38] Early recognition of validity treats allows adjusting the study design accordingly. Validity evaluation in this thesis is based on discussion presented by Wohlin et al. [38] Four types of threats are discussed in further subsections.

3.4.1 Internal validity

The first validity concern in this study is related with literature review. Literature review is the primary source of information in the study, thus biased or incomplete results from the review may influence overall results of the study.

In order to ensure quality of the literature review a snowball sampling review method is elaborated (see section 3.5.1 Literature review for details).

To eliminate biases in the literature review is designed to cover multiple sources of papers – international standards, academia publications and industry related publications. To avoid false dead end in the reference chain, multiple chains in different directions are developed as suggested by Heckathorn. [28] It may be the case that new domains in the topic are identified during the review. In this case, a new chain in this domain is developed.

Nevertheless, referral samples tend to be biased toward the more popular items that are more recognized and referenced by experts of the field. [Heckathorn, Faugier]

3.4.2 Conclusion validity

Conclusion validity concerns relationship between literature review and the results. It is important that the conclusions from literature review are well motivated and backed by more than single source. [38]

This type of threat is addressed by intentionally searching for papers discussing the same issue. Therefore ensuring that each conclusion is motivated multiple sources. If the review shows more than one opinion on the topic then all views are considered for inclusion in the results.

3.4.3 External validity

External validity concerns the generalization of the results. The initial intention of this study is to develop a universal usability and user experience evaluation model. However, the study is based on limited number of papers. This cause inherit treat to the generalizability of the results.

In order to address this threat all results are built on ―solid foundations". The ―solid foundation" is a matter that lies outside the current scope or use case and is recognized as state-of-the-art in global scientific community. For example, the evaluation methodology presented in this thesis is based on empirical investigation strategies – experimentation and case study method [38, 44, 46-48]. Therefore, the results of this thesis can be generalized as far as allowed by universal empirical investigation strategies.

The same principle is applied to the attributes structure. It is based on and compatible with ISO/IEC 9126 standard.

3.4.4 Construct validity

Construct validity concerns the relation between theory and observation. The theoretical model must present results that to some extent represent reality. This threat is directly linked with practical application of the model.

This threat can be addressed by executing a static and dynamic validation. A guidelines and a blueprint of static validation is included in the model.

3.5 Research methods

In order to achieve the goals of this thesis a multi stage study is performed. The study has the following structure, see Figure 3.

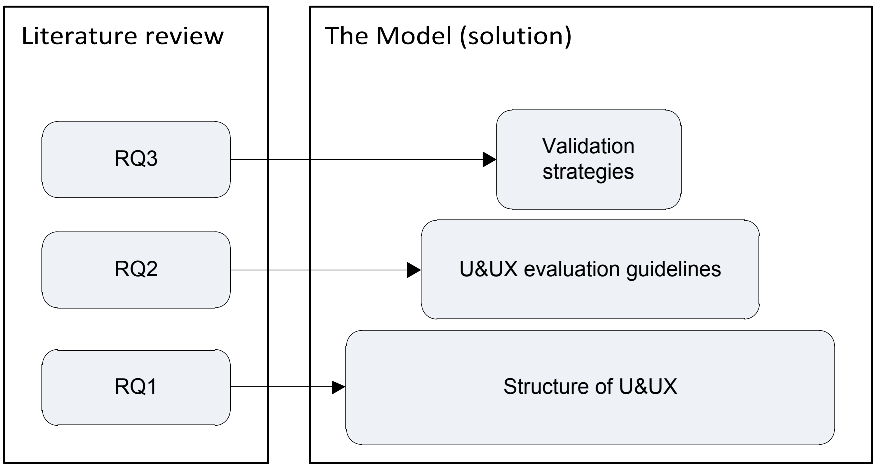

Figure 3, structure of the study

Literature review is used as a primary tool to study existing papers on usability and user experience evaluation. Review of existing studies is motivated by the goal to analyze and compile existing usability and user experience models into one framework. See section 3.5.1

3.5.1 Literature review

Literature review is used as the primary source to answer the research questions. It is performed throughout the entire thesis and consists of three phases, see Figure 4.

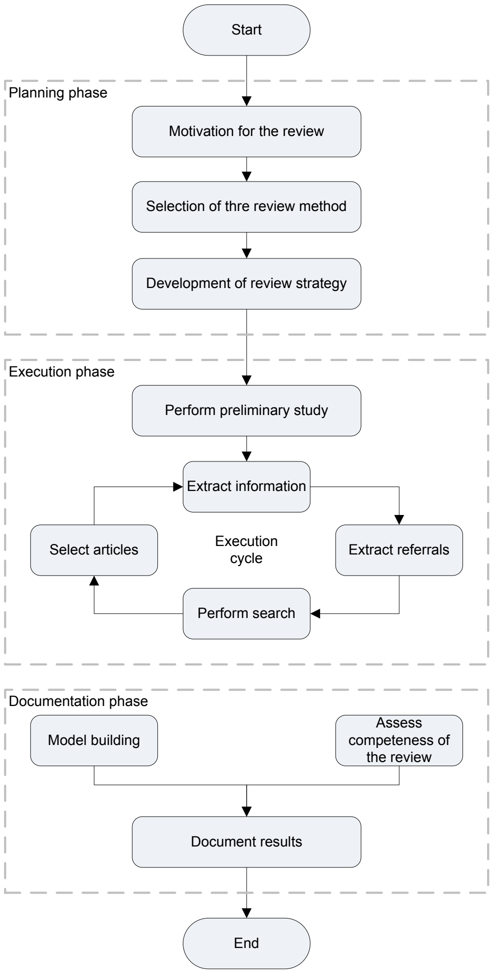

Figure 4, overview of the literature review design

The three stage structure for a literature review is suggested by Kitchenham et al [11]. Although, the review method is not systematic review the same three stage structure by Kitchenham et al. is adapted here. The details of each phase are described in next sections.

The review has three major outcomes. First, the definitions and breakdown structure of usability and user experience is compiled into the model. Secondly, the model is complemented with guidelines on how to use and expand it. Thirdly, the model is accompanied with strategies on how to validate the model.

3.5.1.1 Planning phase

The main purpose of the planning phase within the literature review is to select the review method and develop review strategy. The literature review is motivated by the goals of this thesis – to study the state of the art in usability and user experience evaluation.

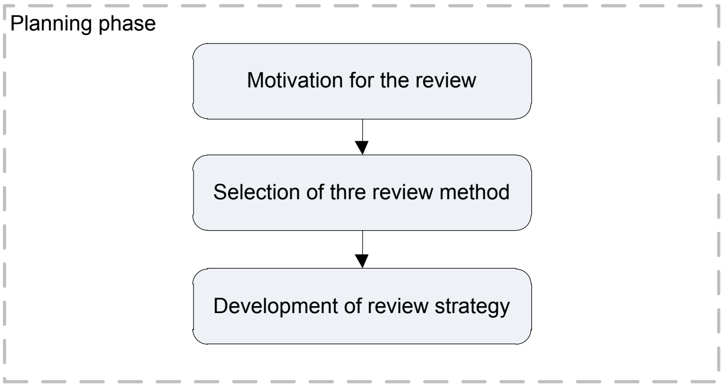

Figure 5, overview of planning phase

Systematic literature review is considered as most mature review method from the empirical investigation toolset. [11, 3] However, this method tends to be very resource demanding due to large number of papers to be reviewed. [11] In order to address this disadvantage, Skoglund and Runneson has studied and evaluated an alternative method for literature reviews. They have found that results from reference based review (referral sampling) have increased precision (less irrelevant papers) and recall is not significantly affected (important papers are not missed) [3].

The literature review within this thesis is conducted by using chain referral sampling method also known as snowball sampling, chain sampling or referral sampling [Skoglund and Runneson]. The idea behind this method is to build sample by studying already known items and extracting referrals to another items. [24, 3] Although, origin of the method can be found in statistics and sociology, the same principle can be applied in different fields.

3.5.1.2 Execution phase

The literature review starts with a preliminary study of the field identifying of most important papers (take of papers) published on usability and user experience. [58] These papers are used as an input for an execution cycle.

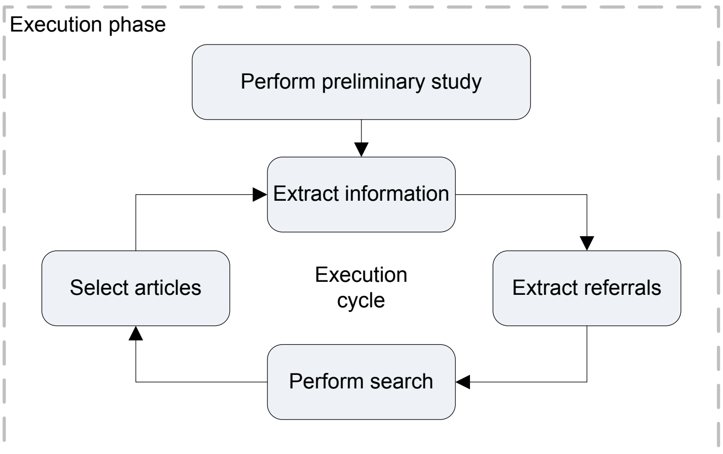

Figure 6, execution phase

Take of papers are selected from multiple domains (international standards, academia papers, studies on usability, studies on user experience). This is done to widen the scope of the review and therefore addressing the risk of missing significant papers.

Execution cycle consists of four steps. First step is to extract information from selected articles. Results from preliminary study are used as sources of information during first iteration. Further iterations uses sources selected during N-1 iteration. The aim of this step is to find answers to research questions specified in section 3.2

During second step references to relevant articles are extracted. These references may include, but not limited to, keywords, relevant standards or methods, concepts, cited or referenced papers and authors. The extracted references are used to develop search queries in third step.

During third step search queries are applied to various databases (primary Arkiv EX, Compendex and Inspec) to extract papers matching these queries. If reference allows to directly identifying a paper, it can be acquired by other means than lookup in a specific database (e.g. ISO standards can be directly acquired from the organizations website)

During fourth step articles returned by search queries are examined whether they target the research questions (specified in section 3.2) Nevertheless, the quality of a paper is also judged by following criteria:

1. Author(s) of a article represents an scientific or industrial organization
2. Article is developed in scientific manner
3. Article is published

This set of criteria allows reviewing articles that are not peer reviewed in academic context. This is been done purposefully to include articles written by industry experts. Besides, any article, book or standard that is acknowledged by industry representatives are reviewed by, at least, other industry experts.

3.5.1.3 Documentation phase

The final phase of the literature review consists of three activities, model building, and assessment of review completeness and documentation of the results, see figure Figure 7, documentation phase.

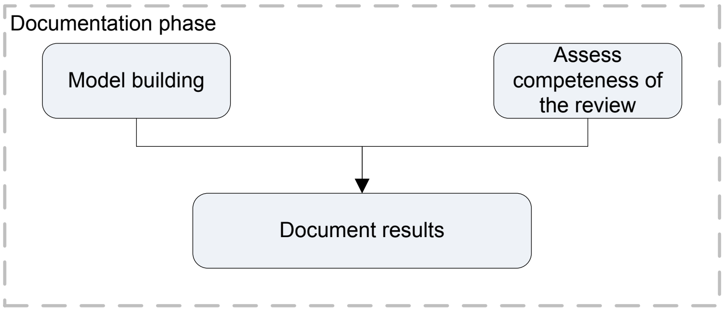

Figure 7, documentation phase

Model building activity is aimed to refine all information extracted during execution phase and compile it into one model. This activity is closely related with execution phase.

Furthermore, the results of review should be assessed for completeness. The literature review is considered complete if the following criteria fulfill:

a) New iterations in execution cycle do not provide any new information of references.
b) All research questions are answered
c) There are no obviously missing parts or holes in the model design.

3.5.2 Model development

This section describes and motivates model development methodology. Subsections of this section focus on each of the model‘s layers.

A model should represent reality to a certain extent. Furthermore, it should be possible to draw conclusions about reality by analyzing the model. The model presented in this thesis features the following:

- Usability and user experience breakdown structure. A tree like structure of usability and user experience attributes.
- Guidelines on usability and user experience evaluation
- Guidelines on how to apply, expand and validate the model

The model is intended to be a useful guide for anyone within software development field who is required to perform usability and user experience evaluation.

To ensure that the model is easy to read and understand, information from the literature review is extracted and the model is developed in layers. The model is developed out of three layers – structure of usability and user experience, usability and user experience validation strategies, and guidelines how to use the model.

3.5.2.1 *Layer I – structure*

The first step towards evaluation of usability and user experience is to understand their structure. The breakdown structure of usability and user experience is developed by reviewing and compiling existing studies on the topic. Identified attributes, factors and measures judged and processed in following way; see Figure 8, model building process and further description.

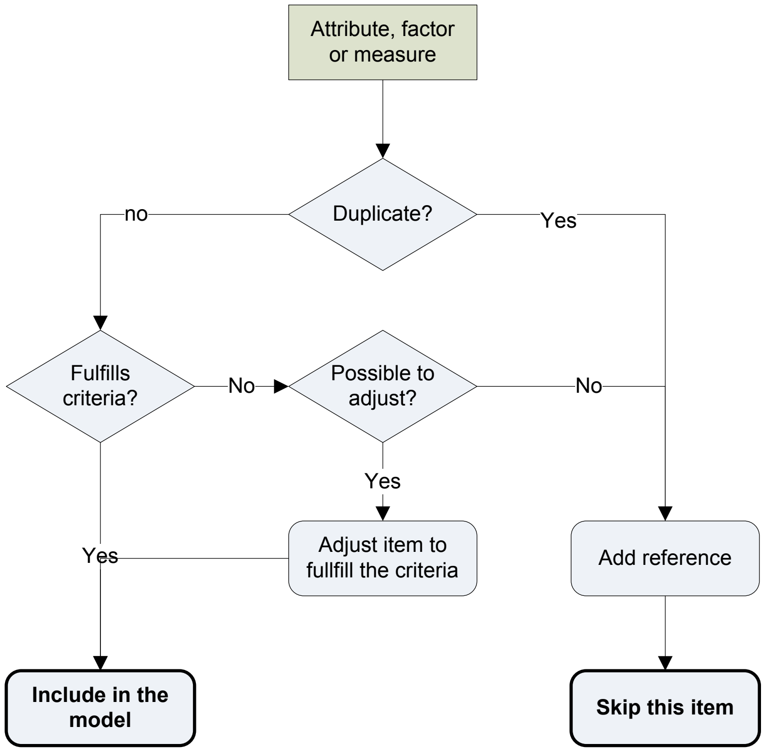

Figure 8, model building process

1. First, activity after an attribute, factor or measure (item) is identified is to check whether it duplicates anything that is already in the model. Items that cannot add anything new to the model are skipped. Nevertheless, a reference is included in the model.
2. Next, activity is to judge the item whether it fulfills criteria to be included in the model
   a. Is item specified clearly and is understandable?
   b. Is item relevant to usability in terms of scope of this model?
   c. Is item specified in uniform detail and level of generalization?
   d. Is item measurable?
3. There can be cases when item should be slightly adjusted in terms of more detailed specification, generalization etc. Such items are adjusted and also included in the model.

When item is added to the model it is described in uniform way by filling up the following fields, see Table 1, attribute description. Since the focus is to measurable sub attributes (terminal nodes in the model) detailed description is given only for them.

Table 1, attribute description

No.	Name	Description
1	No	An unique id of an item
2	Name	A descriptive name of an item
3	Description	Full description of an item
4	Measurement techniques	A list of references to relevant assessment techniques
5	Notes	Notes on adjustments, context or other relevant information
6	Source	A clear reference to another models /papers from which this item is extracted

3.5.2.2 Layer II – evaluation guidelines

The second layer of the model is guidelines on usability and user experience evaluation. The guidelines should provide instructions on how to perform usability and user experience evaluation.

Similar to usability breakdown structure, the guidelines are developed by studying current state of the art in usability and user experience evaluation.

3.5.2.3 Layer III – validation guidelines

There has always been a gap between industry and academia in terms of knowledge transfer. Academia often offers scientifically important results while industry requires practices that are applicable to real projects [60]. Thus, results from academia become useful in industry after pilot testing and adaptation for real life conditions.

The path towards solution is never straight forward. A model of industry - academia collaboration model as presented by Gorschek [61] is depicted in Figure 9, industry - academia collaboration model [61].

A problem is identified within industry and formulated by academia. The solution from academia should pass both static and dynamic validation.

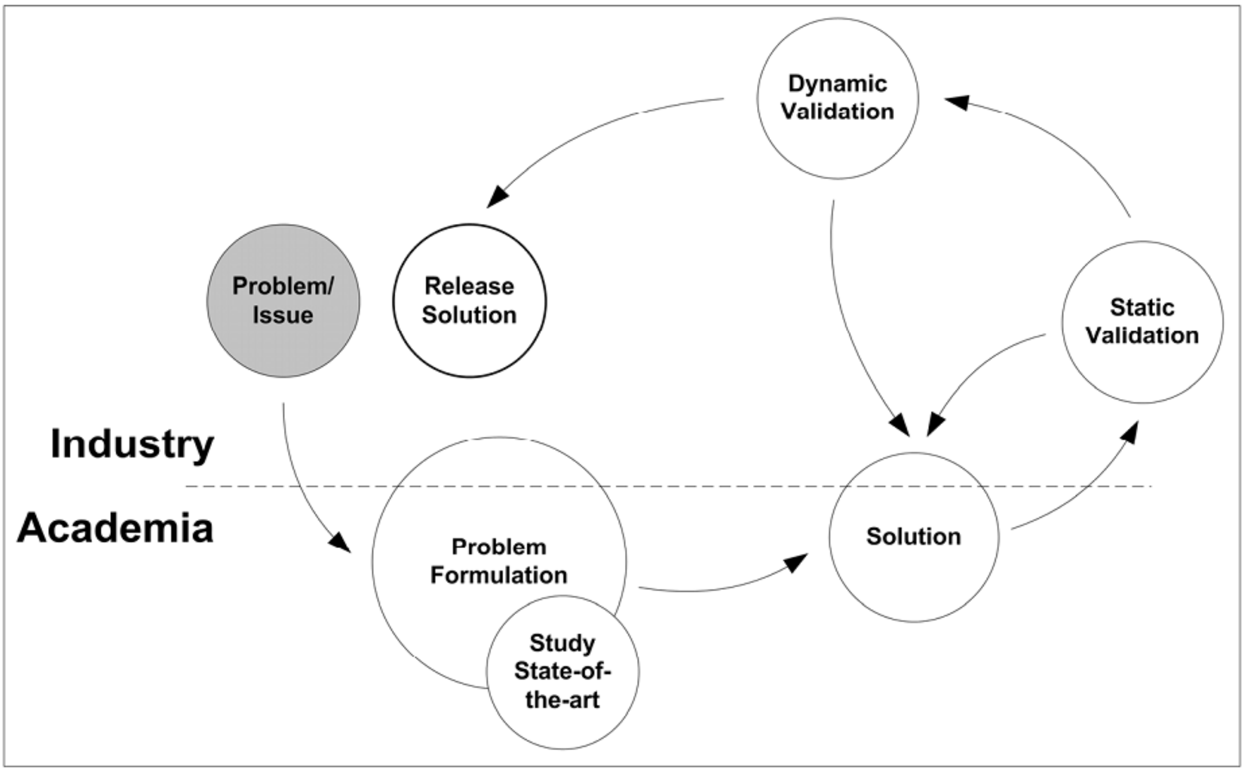

Figure 9, industry - academia collaboration model [61]

The scope of this thesis is limited to academia solution of the problem. However, the solution is accompanied with guidelines on how to conduct further validation of the Model.

The difference between the two types of validation lies in two factors - environment and goals. Static validation tends to be smaller scale, conducted on a synthetic environment and has goals to assess readiness for dynamic validation.

Dynamic validation is always conducted on a real industry setting (or as close to it as possible). The goal is to test does provided solution is applicable in an industry and provides expected results.

4 Model development

This section explains model development process. The model is developed according to the results from literature review. The model which is described in this section is the main result of this thesis.

The literature review is conducted to search for answers to the research questions. The first research question (*RQ1*) is aimed to study the definition of usability and user experience as well as the structure of both. Second research question (*RQ2*) is aimed to study user experience and usability evaluation methods. Third research question (*RQ3*) is aimed to study model validation and adaptation methods. The research questions are stated and motivated in section 3.2.

The model has a structure of three layers. Each of the layers corresponds one of the research questions.

Three level structure of the model will enable modularity of the results. Modularity of the results will allow using the model as whole or combining it with results from alternative studies. For example, the usability and user experience structure layer from this model can be used together with alternative evaluation methodology.

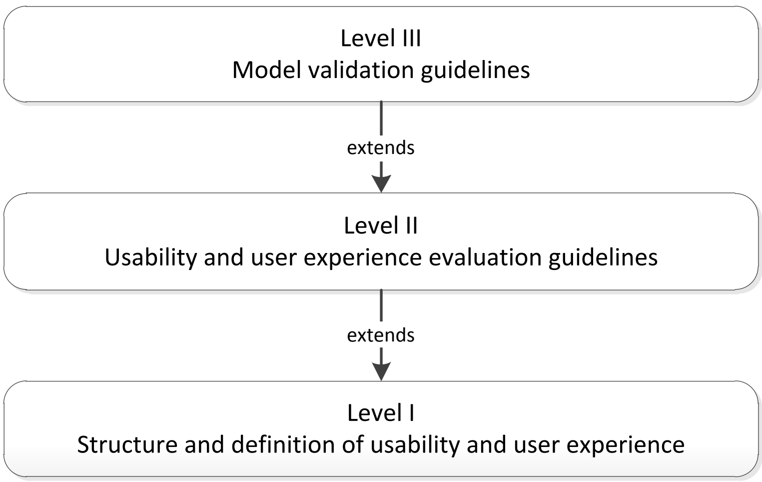

Figure 10, three level structure of the model

4.1 Level I – structure and definition of usability and user experience

First level of the model contains structure and attributes of usability and user experience. Connection between usability and user experience are motivated in section 4.1.4

4 Model development

This section explains model development process. The model is developed according to the results from literature review. The model which is described in this section is the main result of this thesis.

The literature review is conducted to search for answers to the research questions. The first research question (*RQ1*) is aimed to study the definition of usability and user experience as well as the structure of both. Second research question (*RQ2*) is aimed to study user experience and usability evaluation methods. Third research question (*RQ3*) is aimed to study model validation and adaptation methods. The research questions are stated and motivated in section 3.2.

The model has a structure of three layers. Each of the layers corresponds one of the research questions.

Three level structure of the model will enable modularity of the results. Modularity of the results will allow using the model as whole or combining it with results from alternative studies. For example, the usability and user experience structure layer from this model can be used together with alternative evaluation methodology.

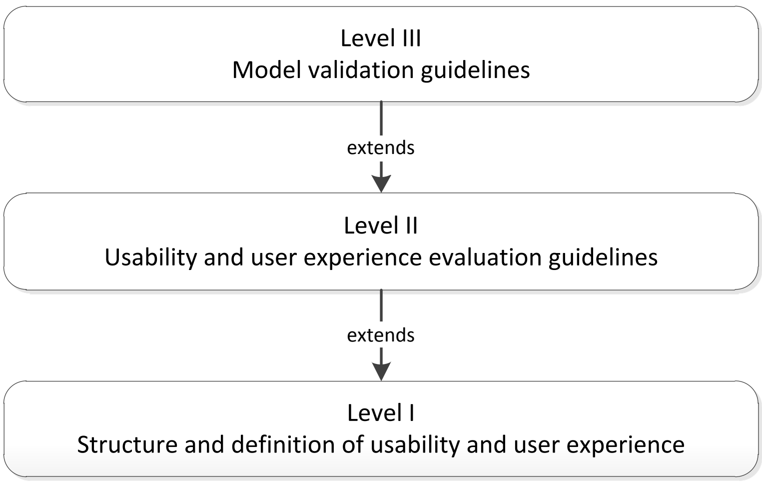

Figure 10, three level structure of the model

4.1 Level I – structure and definition of usability and user experience

First level of the model contains structure and attributes of usability and user experience. Connection between usability and user experience are motivated in section 4.1.4

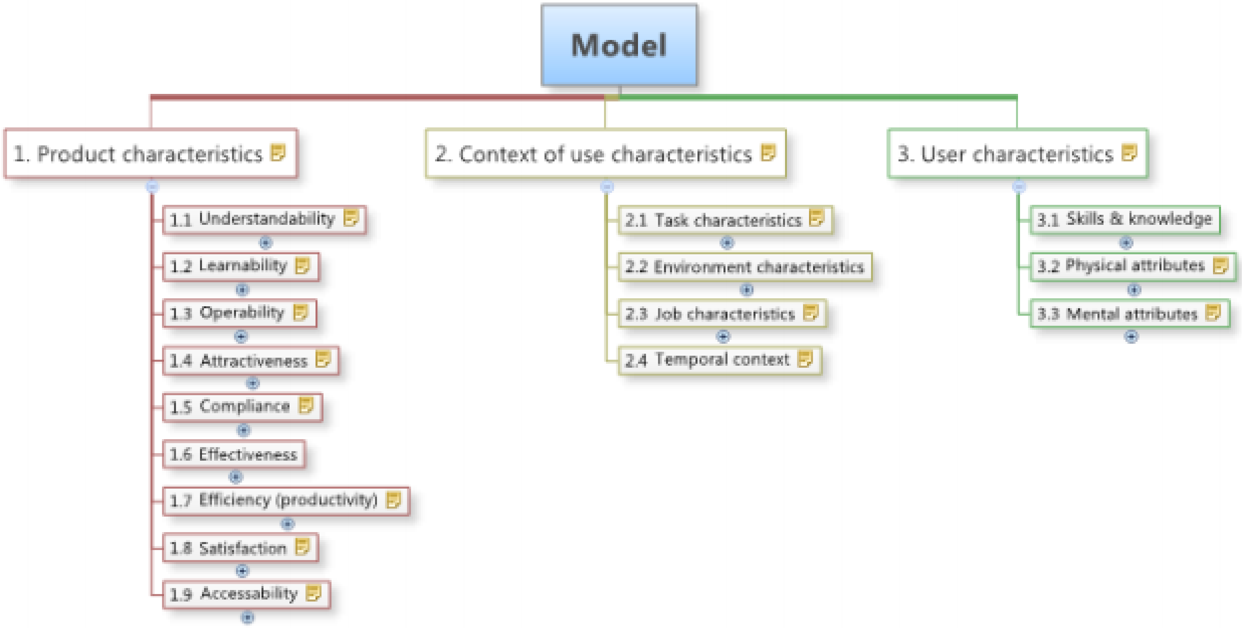

Figure 11, high level structure of usability and user experience attributes

Further subsections cover each group of attributes (product, context of use and user characteristics) providing sources and motivation for inclusion in the model.

4.1.1 Product characteristics

Product characteristics are a group of usability attributes that describes a product itself. The main source for product characteristics is *ISO/IEC 9126:2004* standard. This standard is selected as a backbone for attributes structure due to several reasons. This standard is recognized as industry standard for understanding usability. [1, 2, 20, 21, 26] Furthermore, this standard is also used as a foundation for most of existing usability evaluation methods. Common foundations allow different models to be compatible and the results comparable.

The given standard is a source for idea that usability can be evaluation by analyzing its attributes. Due to that, the standard is a source for many attributes and their structure.

MUSiC (Measuring the Usability of Systems in Context) method extends ISO/IEC 9126 standard with performance, efficiency and effectiveness related attributes. [20] Attributes from *MUSiC* are merged into the model

Quality in Use Integrated Measurement (QUiM) framework analyzes usability as four level (data, metric, criteria and factors) structure. This approach is similar to goal/question/metric (GQM) approach [62] and proposes usability evaluation targeted for specific development goals. [31, 32]

SCANMIC method focuses on analyzing usability for website and proposes a set of specific attributes for the task. [32]

Detailed list of attributes, descriptions and references can be found in appendix X

4.1.2 Context of use characteristics

Context of use describes environment in which the product is being used. The context influences usability and user experience by altering perception of the attributes. For example, a text size on a screen my good enough in office environment, however unreadable in the outdoors where lighting is different. The importance of context of use is empathized by MUSiC method and other scholars. [18, 20, 50]

Context of use is determined by analyzing contextual factors such as physical factors, task specifics, job characteristics and temporal context. The initial attribute structure for context of use is given by ISO/IEC 9126 standard [27] and complemented by attributes from other studies [18, 20, 27, 50]. The attributes structure of context of use is included in the Model.

Definition of context of use is one of the steps in the evaluation process (see section 4.2)

Within usability evaluation studies user characteristics are often added to the context of use. [18, 20] However, user experience studies separates user context as an independent entity. [50, 51] The model supports the latter view due to the need to create a unified usability and user experience evaluated model.

4.1.3 User characteristics

User characteristics are a set of attributes that defines a user. The user is defined by his skills and knowledge, mental and physical characteristics. As mentioned in previous section, user is often viewed as a part of context of use. [18, 20, 27] This approach evaluates a product in a context. [18] However, user experience scholars suggest that user is a separate entity. [50, 51] This approach allows evaluating interaction between product, environment and a user.

The attributes from mentioned papers [18, 20, 27, 50-54] are included in the model.

4.1.4 Connection between usability, context of use and user experience

There have been attempts to formalize connection between usability and user experience. [50, 56] In the scope of the model the connection between usability and user experience is expressed as a shift in evolution goals.

Usability according to its definition focuses on analyzing product features with a certain context of use with intention to improve the product. [27]

User experience according to its definition deals with user‘s perceptions from direct or indirect use of a product. Furthermore, studies reveal that user experience is also linked with context of use. [50-51]

Such views conclude with a system in which product, context of use and user influences each other. See the following figure:

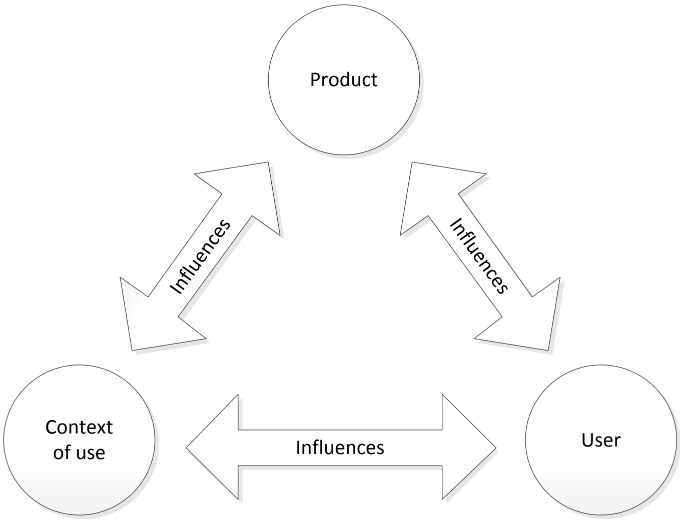

Figure 12, connection between usability, user experience and context of use

The perception of product features is influenced by context of use. The product influences user by satisfying his needs and expectations. User influences context of use by adding his part to temporal context (and maybe other contexts as well).

The other way around, context of use influence user and its perception of a product. The product alters the context.

Such view on the connection between usability and user experience allows putting both concepts in the single evaluation model and apply similar evaluation methods on both.

4.2 Level II – usability and user experience evaluation guidelines

Second level of the model contains guidelines on how to perform usability and user experience evaluation. Such guidelines combined with structure of usability and user experience from Level II allows performing practical usability and user experience evaluation.

Further sections describe and motivate adaptation process. It is assumed that the reader is familiar with concepts of empirical investigation methods. For the purpose of readability some widely known aspects of empirical investigation described here in detail. For more detailed information please refer to the references.

Usability and user experience is a complex concept and can be evaluated by assessing its influencing factors (attributes). [36, 20] However, complete assessment of all usability and user experience attributes is too complex and resource demanding task to be practically applicable.

In order to address this issue a systematic methodology is required to select and assess only critical attributes.

ISO/IEC 9126 proposes attribute structure and simple metrics for each of attributes, however complete usability evaluation methodology is missing in the standard.

In order to address these issues and make the model as universal as possible a solid background is required. By reviewing existing empirical research methodologies a methodology by C. Wohlin et al. [38] is used as a foundation for further development of the guidelines. The authors propose a methodology to carry out an experimental investigation. Furthermore, the experimentation methodology by C.Wohlin et al is compatible with case study investigation methods. [38, 47-48] Both investigation methods are well documented, validated and widely applied. [37-38, 47-48]

In order to apply universal investigation methods for usability and user experience evaluation such methods must be adapted. The adaptation process is based on existing usability evaluation methods.

Authors of MUSiC framework [20] propose seven step methodology to carry out usability evaluation with intention to analyze software efficiency and user performance. Although, performance is only one of usability attributes, the methodology can be applicable in wider scope. This statement is motivated by the following:

a. The methodology itself is not tied to specific attributes or usability domains.
b. The methodology is compatible with empirical investigation methods such as experiments and case studies. [38, 47]

Context of use analysis is critical when performing usability and user experience evaluation. [18, 20, 31] Many papers suggest that it is critical to perform proper context of use analysis in order to assess usability and user experience of a product. [18, 50, 34, 35]

Methodology for context of use analysis is proposed by M. Maguire, C.Thomas and N.Bevan. [34, 35] This methodology complements usability evaluation methodology proposed by MUSiC since one of the steps in MUSiC methodology is context of use analysis. [20]

The Table 2, overview of evaluation steps below provides an overview of usability and user experience evaluation methodology which is developed in this thesis. Further sections provide in detail description of each step.

Table 2, overview of evaluation steps

Step	Name	Description

1	System definition	Initial description of the system[1] including but not limited to: purpose, usage scenarios, goals, stakeholders, user groups etc.
2	Context of use definition	Description of each block is prepared. The descriptions are based on *System definition* (from step 1) and attribute lists (Appendixes D,E and F) Important attributes are identified and their influence described for further investigation.
3	User definition	
4	Product definition	
5	Definition of evaluation goals	Definition of usability and user experience evaluation goals. For example, identification of usability flaws in a product.
6	Development of tests and preparations for test execution	The important attributes from steps 2, 3 and 4 should be put to the test. Proper requirements and test cases should be developed. Furthermore, test environment should be prepared in terms of subject selection, preparation of software and hardware, handouts, observers etc.
7	Test execution	Prepared tests should be executed and measurements collected.
8	Analysis and interpretation	Collected results should be filtered and analyzed. Usually it is done by statistical methods.
9	Reporting the results	Findings from the analysis should be prepared in a form that matches evaluation goals and are understandable by stakeholders.

4.2.1 System definition

System definition is a document describing a purpose for the system, its features, goals, stakeholders, user groups, typical usage scenarios and usage environment. [20] Purpose for the definition is to set limitations and provide a single source of information for an evaluation.

4.2.2 Context of use, user and product definition

Each of three blocks is further described by describing its attributes from appendixes D, E and F. Based on the *system definition* an importance for each attribute is judged according to the following scale:

Important – the attribute has clear influence on the usability and user experience.

Difficult to judge - the attribute is relevant and may have an influence on the usability or user experience.

Irrelevant – the attribute is irrelevant and excluded from further analysis.

If there is more than one distinct user group or usage environment then each of such distinct group or environment should be analyzed separately. For example, if a product is intended for

[1] As described by V. Roto: „The key idea with the System component is to analyze not only the piece of system being investigated but the whole system that this piece is dependent of, or that is involved in the examined use case." [50]

both indoor and outdoor use and then further evaluation should be split to address both environments.

The values are assigned based on expert opinion and prior experience with similar products. [18, 20] For each attribute marked as *important* or d*ifficult to judge* a description of why and how the attribute is relevant should be added.

4.2.3 Definition of evaluation goals

The evaluation goals may vary and influence further evaluation process. At this step the goals should be stated. Examples of the evaluation goals:

- Comparison of similar products in terms of usability and user experience
- Identification of usability and user experience flaws in a product
- Identification of critical usability and user experience factors in a product
- etc

4.2.4 Development of tests and preparations for test execution

The relevant attributes from steps 2, 3 and 4 should be put to the test. Each of the attributes is judged against evaluation goals and added to one of the following groups [34]:

a. **Controlled** – value of the attribute is fixed or limited into an interval by external forces (management, nature, technology, etc) For example, if user already knows how to use any word processing application then that should be sufficient to start using the new one.
b. **Ignored** – the attribute will be ignored in further evaluation
c. **Monitored** – value of the attribute is not limited but monitored to avoid extremes.
d. **Measured** – value of the attribute and exact influence on usability and user experience will be measured within further tests.

In order to document the evaluation process and the test plan it is recommended to list the attributes in the following structure [34]:

a. ID and name of an attribute
b. Influence of the attribute (important, difficult to judge or irrelevant from steps 2,3 and 4)
c. Description of the attribute in the context of system to be evaluated. How and why the attribute is relevant (from steps 2,3 and 4)
d. How the attribute will be treated during further tests (controlled, ignored, monitored or measured)
e. Requirements relevant to the attribute.
f. Test case for attribute and requirement

An example on how to perform evaluation process is provided in section 4.4. Examples on how to document relevant attributes are provided in appendix B.

Further preparation steps and evaluation process is no different from state of the art in empirical investigation in software engineering. This section as well as further evaluation process is based on experimental investigation methodology described by C. Wohlin [34] which is adapted for usability and user experience evaluation [18, 20, 35, 38]

Within previous steps the evaluation goals were set and relevant attributes indentified. Relevant attributes are evaluation experiment variables.

Further preparation steps include:

a. Selection of experiment subjects
b. Development of experiment process
c. Selection of metrics for each relevant attribute
d. Arrangement of experiment environment, materials, hardware and software tools, etc.
e. Validity evaluation

These steps are well described in literature [38-39, 43-48] and will not be covered here in further detail.

Authors of alternative usability evaluation methods encourage usage of audio, video capturing and eye tracking tools. [3, 20, 31] Ability to review and analyze actual experiment process any time later increase chance of finding specific details that may remain undetected under direct observation. Furthermore, adding video or audio of actual user feedback on the final report may spotlight critical findings from the evaluation. [20]

Alternative to an experiment is case study method. A case study is an empirical investigation method that focuses on analyzing one or more specific cases. [44] A case study is a tool of choice when creating an artificial environment for an experiment is not an option.

The differences in planning among experiment and case study lies in execution details. For example, an artificial environment should be created in case of experiment and most interesting cases should be selected in order to conduct a case study. For more information refer to papers [38, 43-48]

4.2.5 Test execution

The experiment should be executed according to the scenario prepared in previous step. Measurement values are recorded during the experiment.

Experiment execution is well described in literature [38-39, 43-48] and will not be covered here in further detail.

4.2.6 Analysis and interpretation

Experiments usually produce a series of data which should be interpreted to provide meaningful and credible results. This is done by applying statistical methods and tools. [38]

4.2.7 Presentation and package

The results from the evaluation should be presented to stakeholders. In order to demonstrate the results in meaningful way to the management a use of figures, graphics and video material is encouraged. The results should satisfy the evaluation goals. [20, 38]

Furthermore, the results should be archived or published for further use. [38]

4.3 Layer III – model validation guidelines

This layer addresses concerns related with practical application of the results of this thesis (the model). This section describes strategy to address validity concerns as well as provide an example of validation design.

The result of thesis is aimed to support usability and user experience evaluation processes in software industry. However, technology transfer between academia and industry is not always a straightforward task. In order to enable practical use of a technology it has to deliver consistent and credible results both in academia and industry setting.

A solution to enable safe technology transfer is presented by T. Gorschek. The author proposes to use industry as laboratory to validate the solution proposed by academia. The validation process itself consists of two stages – static and dynamic. [61] See Figure 9, industry - academia collaboration model [61].

The proposed academia – industry collaboration model states that after a solution is proposed by academia it should pass a static validation. Static validation is a process in which the solution is being tested in artificial environment. The proposed solution may be reviewed and adjusted according to the results from static validation. When static validation is successful, a dynamic validation takes place. Dynamic validation is conducted by applying the solution in real industry case by pilot project or other means. The results from dynamic validation are used for further improvement of the solution until the solution is good enough for release. [61]

This approach is used in the model to verify its applicability in industry setting and specific domains of software development industry.

4.4 Static validation design – an example

As motivated before, a static validation is a first step to transfer knowledge to an industry. In the scope of this thesis an example of static validation design is provided and it is described in this section.

Purpose for a static validation example is to:

a. Demonstrate how to apply the results of this thesis
b. Provide a complete design of static validation which can be immediately executed

Goals of static validation which is described further in this section are:

a. Explore the evaluation process
b. Indentify flaws in the model and evaluation methodology.
c. Collect feedback from the model users

One of key features required for a technology to be accepted by practitioners is how well the potential users understand the technology, how consistent are the results and how easy is to apply the technology. In order to address such concerns a usability of the model is evaluated in the example. Furthermore, the methodology presented in this thesis is applied to itself.

Further sections describe evaluation steps and exemplify the model use. Each of further subsections matches an evaluation step from table Table 2, overview of evaluation steps.

4.4.1 Step 1 - System description

A system being evaluated is the usability and user experience evaluation model presented in this thesis (hereinafter the system).

The system is intended for use in development teams to support decision making in software development process. Typical application for the system is to evaluate a product release with intention to achieve improved the usability and user experience. Results from the evaluation are used to adjust the product development process.

Typical users of the system are software development professionals with background knowledge of usability and user experience evaluation. Project managers, system analysts and testers will make use of the results from the system.

4.4.2 Step 2 – context of use definition

Context of use is described by task goals, task side effects, linked tasks and organizational environment.

Detailed descriptions on individual attributes and their influence are provided in Appendix E

Context of use depends on each use case and such details lies out of this static validation case. The system behavior in realistic environment is analyzed by dynamic validation which is next step in the development of the model.

4.4.3 Step 3 – user definition

User is defined by his knowledge and experience in software quality evaluation; especially usability and user experience evaluation.

Detailed descriptions on individual attributes and their influence are provided in Appendix F

4.4.4 Step 4 – product definition

The system is defined by the following attributes:

a. Completeness of description
b. Understandability of input and output
c. Function understandability
d. Overall satisfaction

Detailed descriptions on individual attributes and their influence are provided in Appendix D

4.4.5 Definition of evaluation goals

The goals for this evaluation are:

a. Assess understandability of the system
b. Assess whether the system can be applied practically
c. Assess user satisfaction

4.4.6 Development of test plan

Tests involve subjects who are asked to use the system and perform various tasks. Subject performance are recorded and analyzed. This section describe test plan in detail. The test plan has several sub steps. Each of sub steps is described further in this section.

Selection of the subjects

Experiment subjects are selected to match the user description from section 4.4.3 (Step 3 – user definition). It may be reasonable to select subjects among software development students. Students from same year are comparable level of experience and knowledge which is important to achieve credible results.

Description of experiment process

The validation experiment is actual application of the system in controlled environment –a classroom. Subjects are invited to form groups of 3 - 5 persons. Each group will act as a team to perform usability and user experience evaluation for a university website. In order to perform the experiment each team must have a computer with internet access.

The experiment begins with a short presentation where the motivation and rules for the experiment are explained. The rules forbid interaction among teams and encourage use of supervisor advice and external sources of information. However, when team uses any help from outside it must be logged in team‘s report explaining the reasons for such action.

After the introduction the evaluation methodology is presented and handouts outlining the methodology are shared to the teams.

Teams are asked to proceed with usability and user experience evaluation for university website. Each team prepares a evaluation report which states and motivates the following:

a. A description of relevant knowledge and experience in usability and user experience evaluation in a team.
b. The university‘s website usability and user experience evaluation report. The evaluation should be performed according to the model guidelines. Team should identify 5 most important attributes in each group (website, user and context of use) and evaluate them according to their experience. Selection of most important attributes and evaluation criteria should be motivated in a report.
c. Free interpretation of the results.
d. Feedback on the model. Team should state their experience working with the model. E.g. quality of descriptions, understandability etc.

The following variables are measured in the experiment.

a. Number of teams that completed the task.
b. Number of teams that understood the methodology and successfully applied it. (Completeness of description)
c. Number of teams that wisely selected the input data and successfully interpreted the results. (Understandability of input and output)
d. Number of teams that successfully applied the model attributes and guidelines. (Function understandability)
e. Overall satisfaction

The following variables are controlled during the experiment:

a. Level of relevant knowledge and experience in the teams
b. Time and effort spent to complete the task

Validity evaluation[2]:

Results of the experiment may be negatively affected by homogeneity of subjects and typicality of the case (the university website). To address such concerns, subjects with similar background and level of knowledge should be selected.

To increase generalizability and credibility of the results experiment should be repeated multiple times analyzing different types of software products and increasing number of participating teams. To achieve statistically significant results, at least 30 teams are required. [38]

Another concern is number of teams that did not complete the experiment and failed to submit the report. Significant number of failed teams may indicate flaws in experiment design or execution as well as flaws in the model. Such incomplete reports should be analyzed and causes of failure investigated.

4.4.7 Test execution

Within this step an actual experiment is conducted.

4.4.8 Analysis and interpretation

Analysis of the results starts when the experiment is completed and all reports are collected. Successful reports should be separated from incomplete or otherwise invalid reports. Invalid reports should be analyzed separately (see validity evaluation in previous subsection).

Successful reports are further analyzed according to criteria given in appendix B. Dominant values for each criteria should be determined by statistical means. If common pattern for problems can be identified it must be logged in the results.

All successful reports should provide similar results on the websites usability and user experience qualities. This is an important requirement for the model to achieve repeatable and credible results.

4.4.9 Reporting the results

The final report of the evaluation consists of exact details of experiment design and execution as well as the interpreted results. The results should be included in the model and used to adjust and improve the model.

---

[2] Validity threats described in this example does not represent an exhaustive list of all possible threats that may influence the results. This list rather exemplifies common issues and ways to address them.

4.5 Skipped usability and user experience evaluation methods

The literature review revealed a variety of methods and approaches to assess usability and user experience. Some of the methods were not included in the model due to reasons explained in this section.

J. Nielsen proposes to perform usability evaluation according to usability heuristics. [16] This methodology is based on 10 usability heuristics which analyzes various aspects of usability. The author in his other paper compares different methods of usability evaluation methods and states that heuristic evaluation is more practical and less systematic method to perform usability evaluation [14]

Heuristic usability evaluation is excluded from the model due to the following reasons:

a. Description on how to systematically apply the method was not identified. Therefore, the results from heuristic evaluation depend on evaluator‘s experience. Such fact raises doubts on credibility and objectivity of the results. [14, 16]
b. Link between ISO/IEC 9126 standard and usability heuristics was not identified.
c. The other methods described by J. Nielsen [14, 16] are part of other more mature and better motivated usability evaluation methods. [18, 20, 31, 34, 36]

The experts of user experience states that user experience is influenced by product‘s physical properties such as weight, color, materials etc.[51] Such physical attributes lies out of scope of this thesis.

5 Results

This section outlines the results of this thesis. Section of main results points out exact results and contribution of this thesis. Second subsection provides answers to the research questions (stated in section Research questions0) Last section describes lessons learned during the development of this thesis.

5.1 Main results

The main result of this thesis is usability and user experience model. The model is developed by studying state of the art in usability and user experience evaluation. Pros and cons of existing models were analyzed with aim to develop a one universal framework for usability and user experience evaluation in software development industry. The model has three layers comprising usability and user experience attributes evaluation methodology and validation guidelines.

The model is compatible with majority of state of the art usability and user experience evaluation methods and standards such as MUSiC and ISO/IEC 9126. Three layers structure enables to use the model as a whole or break it down and mix with other evaluation methods.

5.1.1 Usability and user experience attributes

As revealed by literature review, usability and user experience can be evaluated by analyzing their attributes. Furthermore, three domains of attributes – user, context of use and product were identified. Such three domain structure allows putting usability and user experience attributes in one structure. Joining usability and user experience in a one model is one of the contributions of this thesis. A 217 different attributes of product, user and context were identified and arranged in a hierarchical structure. Appendixes D, E and F describes individual attributes, section 4.1 discusses the development of the attribute tree.

5.1.2 Usability and user experience evaluation methodology

A systematic methodology is developed to perform usability and user experience evaluation. The methodology is based on empirical investigation methods and adjusted for the needs of usability and user experience evaluation.

The methodology defines 9 steps of usability and user experience evaluation. The evaluation steps are independent from specific application domains or investigation methods. Therefore, the methodology can be applied to wider scope of products, for example, the same methodology can be applied to analyze usability and user experience of word processing software and a bicycle.

Furthermore, the methodology is based on state of the art empirical investigation methods. Thus, the methodology is flexible to be extended or adapted to fit requirements of a specific case.

The methodology is explained in section 4.2 and exemplified in section 4.4. Appendix C provides an example on how to document evaluation process.

5.1.3 Validation guidelines

To enable technology transfer from academia to industry a validation strategy is provided as a part of the model. The validation strategy is exemplified with a blueprint of validation design which can be used to ensure validity of the model in certain context.

The exemplified static validation design proposes to perform a classroom experiment and analyze how well the model is accepted and applied by software engineering students. The purpose of such validation is to ensure that there are no flaws in descriptions and that the methodology gives consistent results.

Model validation strategies are explained in section 4.3 and exemplified in section 4.4.

6 Conclusion

6.1 Discussion

The methodology developed in this thesis enables to perform systematic usability and user experience evaluation for software engineering products. Comparing with previous methods this thesis presents more detailed structure of product, user and context of use, actual evaluation methodology and blueprint of static validation.

The literature review revealed gaps in understanding on usability and user experience. Usability well covered in papers and studies. Whereas, user experience is a new topic and solid foundation for common view is missing. Current understanding of user experience is based on ad-hoc case studies. Existing studies reveal that there is a link among usability and user experience, however it is not clearly defined. To achieve goals of this thesis, a usability – user experience – context triangle is presented (see section 4.1 for details), however such aim should be further validated.

Usability and user experience is a complex phenomena and which is difficult to be studied within limits of software engineering domain. Software product is rarely used in isolation. Software normally interacts with other software and is tied to hardware. Hardware is connected to peripherals and interacts with other hardware via wired or wireless communications. All involved components affect user experience and to achieve correct figures all components should be evaluated. Such circumstances create a need to perform a wider study on usability and user experience.

To ensure quality of this thesis the literature review was focused on high quality papers. However, a large amount of usability and user experience related papers were identified in “gray zone” – blogs, community and company websites. Papers in such resources are valuable and practical “cookbook recipes” acknowledged by industry practitioners. However, information in such “gray” resources often lacks connection with state of the art, for example ISO/IEC 9126 standard. The task of exploring, systematizing and validating information in such resources would be a topic for further studies.

6.2 Further work

Further work of this thesis can be separated in two directions. First direction is to validate and improve the results of this thesis. A static and dynamic validation is required to enable practical use of the results. Static validation can be performed with relatively low efforts by applying guidelines presented in this thesis.

Second direction is to widen the scope beyond limits of software engineering and perform a comprehensive study of user experience and usability. A need for such study is motivated by recent developments in smart phone and tablet market. For such products a distinction between software and hardware is dissolving. Thus, to analyze usability and user experience of such devices hardware properties should be analyzed along with the software.

Appendix A - Guidelines on how to expand the Model

This appendix covers guidelines on how to expand the Model. The guidelines are aimed to support a uniform way to add new attributes to the U&UX breakdown structure or alter the Model guidelines.

In order to expand the breakdown structure, two activities should take place.

a) Assess quality of the attribute to be added
b) Determine location of the attribute within the breakdown structure

The following checklist aims to assess quality of the attribute. (The following checklist represents same methodology as applied for initial model building, see section **Error! Reference ource not found.**)

- **Completeness**. The attribute must have a complete description of the attribute. The description should include information on how the attribute influences usability.
- **Source.** The source of the attribute should be reliable, e.g. a valid study.
- **Measurability**. The attribute should have objective measures. Such measures should be included in the attribute‘s description.
- **Compliance.** The attribute to be added should be compliant with current state of the art in usability evaluation represented by ISO/IEC 9126 standard.

In order to determine a location of the attribute within the breakdown structure the following should be considered:

- **Parent**. Each attribute within the breakdown structure describes an aspect of U&UX. When adding a new attribute its parent node should be determined by reviewing the existing breakdown structure and the attribute‘s description.
- **Overlapping**. It may be the case that the attribute to be added to some extent duplicates with one already in the Model. In such case already present attribute should be updated with new information.

Appendix B – criteria for result evaluation

Variable	Poor	Average	Good	Excellent
Completeness of the description	There are errors that void the results.	There are slips from the Model that influences the final results	The report follows the evaluation guidelines, but there are some slips that do not affect overall results.	The report flawlessly follows the evaluation guidelines.
Understandability of input and output	There are errors that void the results	There are some slips or lack of motivation behind input data or results are misunderstood.	There are slips, however the overall results are meaningful and credible	The report contains motivated and meaningful input to the model and results are interpreted in a meaningful way.
Function understandability	The results are not compatible with the Model functions.	The model functions are misunderstood or missed. External sources of information applied.	The model functions are used accordingly however external sources of information are used.	The model functions are used accordingly and external sources of information are not used.
Overall satisfaction	Team has rated the Model as poor. Comments express serious flaws in the Model.	Team has rated the Model as average. Comments contain well motivated critics	Team has rated the Model as good. Comments contain constructive feedback and some critics	Team has rated the Model as excellent. Comments contain constructive and positive feedback

Appendix C – example of evaluation documentation

This appendix contains documentation relevant to the static validation example described in section 4.4. This documentation is developed according to guidelines presented by M. Maigure [34].

Context of use attributes

Attribute ID	Name, description in the context of given system	Influence, treatment	Requirements and test cases
2.11	**Task goal** Definition of task goals determines to what extent the goals can be achieved by applying the model.	*Influence*: Important *Treatment*: Ignore	
2.1.4	**Task side effects**	*Influence*: Important *Treatment*: Ignore	
2.1.13	**Linked tasks**	*Influence*: Important *Treatment*: Ignore	
2.2.1	**Organizational environment**	*Influence*: Important *Treatment*: Ignore	

User attributes

Attribute ID	Name, description in the context of given system	Influence, treatment	Requirements and test cases
3.1.1.1	**Training in tasks supported by product main functions** Users of the model has previous experience in software quality evaluation	*Influence*: Important *Treatment*: Control	*Requirement:* Model users have previous experience in software quality evaluation. *Test case:* Does previous experience in software quality evaluation is sufficient to successfully apply the model.
3.1.2.2	**Experience with products with similar functions** Users of the model has previous experience in usability and user experience quality evaluation	*Influence*: Important *Treatment*: Control	*Requirement:* Model users have general knowledge on usability and user experience *Test case:* Does general knowledge on usability and user experience evaluation is sufficient to successfully apply the model.

Product attributes

Attribute ID	Name, description in the context of given system	Influence, treatment	Requirements and test cases
1.1.1	**Completeness of description** Understandability of the model depends on its description	*Influence*: Important *Treatment*: Measure	*Requirement*: Description should describe and exemplify the model features. *Test case*: The description is sufficient source of information for a user to understand and apply the model.
1.1.7	**Understandability of input and output** Understandability of input and output influences the quality of results and validity of further conclusions.	*Influence*: Important *Treatment*: Measure	*Requirement*: User should be able to select correct input data and provide meaningful interpretation of the results. *Test case:* Repeated and independent evaluation cases provide similar results.
1.1.6	**Function understandability** User understanding of the model functions influences the ability to successfully apply the model.	*Influence*: Important *Treatment*: Measure	*Requirement:* Users understands and can apply the model‘s functions *Test case:* Users are able to demonstrate their understanding by practical example.
1.8.1	**Overall satisfaction** Positive feedback is a key for the methodology to be accepted and used by practicioneers.	*Influence*: Important *Treatment*: Measure	*Requirement:* The model should satisfy user‘s needs in terms of usability and user experience evaluation. *Test case:* User provides positive feedback after use of the model.

Appendix D – List of product attributes

Product quality attributes – understandability

Understandability is described as: Users should be able to select a software product, which is suitable for their intended use. An external understandability metric should be able to assess whether new users can understand:

- whether the software is suitable
- How it can be used for particular tasks.

No.	Name	Description	Measurement techniques	Notes	Source
1.1.1	Completeness of description	What proportion of functions (or types of functions) is understood after reading the product description?	Conduct user test and interview user with questionnaires or observe user behavior. Count the number of functions which are adequately understood and compare with the total number of functions in the product. X = A / B A = Number of functions (or types of functions) understood B = Total number of functions (or types of functions)	This indicates whether potential users understand the capability of the product after reading the product description.	ISO/IEC 9126
1.1.2	Demonstration	What proportion of the demonstrations/ tutorials can the	Conduct user test and observe user behavior.	This indicates whether users can find the demonstrations and/or	ISO/IEC

	accessibility	user access?	Count the number of functions that are adequately demonstrable and compare with the total number of functions requiring demonstration capability. X = A / B A= Number of demonstrations / tutorials that the user successfully accesses B= Number of demonstrations / tutorials available	tutorials.	9126
1.1.3	Demonstration accessibility in use	What proportion of the demonstrations / tutorials can the user access whenever user actually needs to do during operation?	Observe the behavior of the user who is trying to see demonstration/tutorial. Observation may employ human cognitive action monitoring approach with video camera. X = A / B A= Number of cases in which user successfully sees demonstration when user attempts to see demonstration B= Number of cases in which user attempts to see demonstration during observation period	This indicates whether users can find the demonstrations and/or tutorials while using the product.	ISO/IEC 9126

1.1.4	Demonstration effectiveness	What proportion of functions can the user operate successfully after a demonstration or tutorial?	Observe the behavior of the user who is trying to see demonstration/tutorial. Observation may employ human cognitive action monitoring approach with video camera. X = A / B A= Number of functions operated successfully B= Number of demonstrations/tutorials accessed	This indicates whether users can operate functions successfully after an online demonstration or tutorial.	ISO/IEC 9126
1.1.5	Evident functions	What proportion of functions (or types of function) can be identified by the user based upon start up conditions?	Conduct user test and interview user with questionnaires or observe user behavior. Count the number of functions that are evident to the user and compare with the total number of functions. X = A / B A = Number of functions (or types of functions) identified by the user B = Total number of actual functions (or types of functions)	This indicates whether users are able to locate functions by exploring the interface (e.g. by inspecting the menus)	ISO/IEC 9126
1.1.6	Function	What proportion of the product	Conduct user test and interview user with	This indicates whether users are	ISO/IEC

	understandability	functions will the user be able to understand correctly?	questionnaires. Count the number of user interface functions where purposes are easily understood by the user and compare with the number of functions available for user. X= A / B A= Number of interface functions whose purpose is correctly described by the user B= Number of functions available from the interface	able to understand functions by exploring the interface (e.g. by inspecting the menus).	9126
1.1.7	Understandable input and output	Can users understand what is required as input data and what is provided as output by software system?	Conduct user test and interview user with questionnaires or observe user behavior. Count the number of input and output data items understood by the user and compare with the total number of them available for user. X= A / B A= Number of input and output data items which user successfully understands B= Number of input and output data items available from the interface	This indicates whether users can understand the format in which data should be input and correctly identify the meaning of output data.	ISO/IEC 9126

Product quality attributes – learnability

Learnability is described as: An external learnability metric should be able to assess how long users take to learn how to use particular functions, and the effectiveness of help systems and documentation.

Learnability is strongly related to understandability, and understandability measurements can be indicators of the learnability potential of the software. [27]

No.	Name	Description	Measurement techniques	Notes	Source
1.2.1	Ease of function learning	How long does the user take to learn to use a function?	Conduct user test and observe user behavior. T= Mean time taken to learn to use a function correctly	It is recommended to determine an expected user's operating time as a short time. Such user's operating time may be the threshold, for example, which is 70% of time at the first use as the fair proportion. Effort may alternatively represent time by person-hour unit.	ISO/IEC 9126
1.2.2	Ease of learning to perform a task in use	How long does the user take to learn how to perform the specified task efficiently?	Observe user behavior from when they start to learn until they begin to operate efficiently. T= Sum of user operation time until user achieved to perform the specified task within a short time. T= Sum of user operation time until user achieved to perform the specified task within a short time.		ISO/IEC 9126
1.2.3	Effectiveness of the user documentation and help system	What proportion of tasks can be completed correctly after using the user documentation and/or help system?	Conduct user test and observe user behavior. Count the number of tasks successfully completed after accessing online help and/or documentation and compare with	Three metrics are possible: completeness of the documentation, completeness of the help facility, or completeness of the help and documentation used in combination.	ISO/IEC 9126

			the total number of tasks tested. X= A / B A= Number of tasks successfully completed after accessing online help and/or documentation B = Total of number of tasks tested		
1.2.4	Effectiveness of the user documentation and help system in use	What proportion of functions can be used correctly after reading the documentation or using help systems?	Observe user behavior. Count the number of functions used correctly after reading the documentation or using help systems and compare with the total number of functions. X = A / B A = Number of functions that can be used B = Total of number of functions provided [36]	This metric is generally used as one of experienced and justified metrics rather than the others. [36]	ISO/IEC 9126, QUIM
1.2.5	Help accessibility	What proportion of the help topics can the user locate?	Conduct user test and observe user behavior. Count the number of tasks for which correct online help is located and		ISO/IEC 9126, QUIM

		compare with the total number of tasks tested. X = A / B A = Number of tasks for which correct online help is located B = Total of number of tasks tested		
1.2.6 Help frequency	How frequently does a user have to access help to learn operation to complete his/her work task?	Conduct user test and observe user behavior. Count the number of cases that users accesses help to complete his/her task. X = A A = Number of accesses to help until a user completes his/her task.		ISO/IEC 9126

Product quality attributes – operability

Operability is described as: An external operability metric should be able to assess whether users can operate and control the software. Operability metrics can be categorized by the dialogue principles in ISO 9241-10:

- suitability of the software for the task
- self-descriptiveness of the software
- controllability of the software
- conformity of the software with user expectations
- error tolerance of the software
- suitability of the software for individualization

The choice of functions to test will be influenced by the expected frequency of use of functions, the criticality of the functions, and any anticipated usability problems. [27]

Amount of effort necessary to operate and control a software product. [30]

No.	Name	Description	Measurement techniques	Notes	Source
1.3.1	User expectations/ likeability	How product satisfies implied needs of a user? User‘s perceptions, feelings, and opinions of the product			QUIM
1.3.1.1	Operational consistency in use	How consistent are the component of the user interface?	Observe the behavior of the user and ask the opinion. a) X = 1 - A / B A= Number of messages or functions which user found unacceptably inconsistent with the user‘s expectation B= Number of messages or functions b) Y = N / UOT N= Number of operations which user found unacceptably inconsistent with the user‘s expectation UOT= user operating time	User‘s experience of operation is usually helpful to recognize several operation patterns, which derive user‘s expectation. Both of ―input predictability" and ―output predictability" are effective for operational consistency. This metric may be used to measure ―Easy to derive operation" and ―Smooth Communication".	ISO/IEC 9126

			(during observation period)		
1.3.1.2	Trustfulness	Faithfulness a software product offers to its users. [33]			QUIM, SCANMIC
1.3.1.3	Feedback	Responsiveness of the software product to user inputs or events in a meaningful way			QUIM
1.3.1.4	Completeness	Whether a user can complete a specified task			QUIM
1.3.2	Controllability	Can user control the software?			ISO/IEC 9126, QUIM
1.3.2.1	Error correction	Can user easily correct error on tasks?	Conduct user test and observe user behavior. $T = Tc - Ts$ Tc = Time of completing correction of specified type errors of performed task Ts = Time of starting correction of specified type errors of performed task	User of this metric is suggested to specify types of errors for test cases by considering, for example, severity (displaying error or destroying data), type of input/output error (input text error, output data error to database or graphical error on display) or type of error operational situation (interactive use or emergent operation).	ISO/IEC 9126, QUIM
1.3.2.2	Error correction in	Can user easily recover his/her	Observe the behavior of the	When function is tested one by one, the ratio	ISO/IEC

use	error or retry tasks? Can user easily recover his/her input?	user who is operating software. a) X= A / UOT A= number of times that the user succeeds to cancel their error operation UOT= user operating time during observation period b) X = A / B A= Number of screens or forms where the input data were successfully modified or changed before being elaborated B = Number of screens or forms where user tried to modify or to change the input data during observed user operating time	can be also calculated, that is the ratio of number of functions which user succeeds to cancel his/her operation to all functions.	9126, QUIM

1.3.3	Suitable for task operation				ISO/IEC 9126
1.3.3.1	Default value availability in use	Can user easily select parameter values for his/her convenient operation?	Observe the behavior of the user who is operating software. Count how many times user attempts to establish or to select parameter values and fails, (because user cannot use default values provided by the software). X = 1 - A / B A= The number of times that the user fail to establish or to select parameter values in a short period (because user cannot use default values provided by the software) B= Total number of times that the user attempt to establish or to select parameter values	It is recommended to observe and record operator's behavior and decide how long period is allowable to select parameter values as —short period". When parameter setting function is tested by each function, the ratio of allowable function can be also calculated. It is recommended to conduct functional test that covers parameter-setting functions.	ISO/IEC 9126

1.3.3.2	Minimal action	Capability of the software product to help users achieve their tasks in a minimum number of steps.			QUIM
1.3.3.3	Minimal memory load	Whether user is required to keep minimal amount of information in mind in order to achieve a specified task.			QUIM
1.3.4	Self descriptive				ISO/IEC 9126, QUIM
1.3.4.1	Message understandability in use	Can user easily understand messages from software system? Is there any message which caused the user a delay in understanding before starting the next action? Can user easily memorize important message?	Observe user behavior that is operating software. X = A / UOT A = number of times that the user pauses for a long period or successively and repeatedly fails at the same operation, because of the lack of message comprehension. UOT = user operating time	The extent of ease of message comprehension is represented by how long that message caused delay in user understanding before starting the next action. Therefore, it is recommended to observe and record operator's behavior and decide what length of pause is considered a ―long period". When messages are tested one by one, the ratio of comprehended messages to the total can be also calculated. When several users are observed who	ISO/IEC 9126, QUIM

			(observation period)	participants of operational testing are, the ratio of users who comprehended messages to all users can be calculated.	
1.3.4.1.1	Message attentiveness	Attentiveness implies that user successfully recognizes important messages presenting information such as guidance on next user action, name of data items to be looked at, and warning of careful operation. - Does user ever fail to watch when encountering important messages? - Can user avoid mistakes in operation, because of recognizing important messages?			ISO ISO/IEC 9126
1.3.4.1.2	Message memorability	Memorability implies that user remember important messages presenting information such as guidance on the next user action, name of data items to be looked at, and warning of careful operation. - Can user easily remember			ISO ISO/IEC 9126

		important messages? - Is remembering important messages helpful to the user? - Is it required for the user to remember only a few important messages and not so much?		
1.3.4.2	Self explanatory error messages	In what proportion of error conditions does the user propose the correct recovery action?	Conduct user test and observe user behavior. X= A / B A =Number of error conditions for which the user proposes the correct recovery action B =Number of error conditions tested	ISO/IEC 9126
1.3.4.3	Consistency	Degree of uniformity among elements of user interface and whether they offer meaningful metaphors to users.		QUIM
1.3.5	Operational error tolerant			ISO/IEC 9126

1.3.5.1	Operational error recoverability in use	Can user easily recover his/her worse situation?	Observe the behavior of the user who is operating software. X = 1 - A / B A= Number of unsuccessfully recovered situation (after a user error or change) in which user was not informed about a risk by the system B= Number of user errors or changes	The formula is representative of the worst case. User of this metric may take account of the combination of 1) the number of errors where the user is / is not warned by the software system and 2) the number of occasions where the user successfully / unsuccessfully recovers the situation.	ISO/IEC 9126
1.3.5.2	Time between human error operations in use	Can user operate the software long enough without human error?	Observe the behavior of the user who is operating software. X = T / N (at time t during [t-T, t]) T = operation time period during observation (or The sum of operating time between user's human error operations)		ISO/IEC 9126

			N= number of occurrences of user's human error operation		
1.3.5.2.1	Simple human error (Slip)	Error when the user just simply makes errors to input operation;			ISO/IEC 9126
1.3.5.2.2	Intentional error (mistake)	Error when the user repeats fail an error at the same operation with misunderstanding during observation period			ISO/IEC 9126
1.3.5.2.3	Operational hesitation pause	Situations whe the user pauses for a long period with hesitation during observation period		It depends on the function, operation procedure, application domain, and user whether it is considered a long period or not for the user to pause the operation. Therefore, the evaluator is requested to take them into account and determine the reasonable threshold time. For an interactive operation, a "long period" threshold range of 1min. to 3 min.	ISO/IEC 9126
1.3.5.3	User error correction	How frequently does the user successfully correct input errors? How frequently does the user correctly undo errors?	Conduct user test and observe user behavior. a) X= A / B A= Number of input errors which the user successfully corrects	This metric is generally used as one of experienced and justified	ISO/IEC 9126

			B= Number of attempts to correct input errors b) Y= A / B A= Number of error conditions which the user successfully corrects B= Total number of error conditions tested		
1.3.6	Suitable for individualization				ISO/IEC 9126
1.3.6.1	Customizability	Can user easily customize operation procedures for his/her convenience? Can a user, who instructs end users, easily set customized operation procedure templates for preventing their errors? What proportion of functions can be customized?	Conduct user test and observe user behavior. X= A / B A= Number of functions successfully customized B= Number of attempts to customize	1 Ratio of user's failures to customize may be measured. Y = 1 - (C / D) C = Number of cases in which a user fails to customize operation D = Total number of cases in which a user attempted to customize operation for his/her convenience. 0<=Y<= 1, The closer to 1.0 is the better. 2 It is recommended to regard the following as variations of customizing operations:	ISO/IEC 9126, QUIM

				- chose alternative operation, such as using menu selection instead of command input; - combined user's operation procedure, such as recording and editing operation procedures; - set constrained template operation, such as programming procedures or making a template for input guidance. 3 This metric is generally used as one of experienced and justified.	
1.3.6.2	Physical accessibility	What proportion of functions can be accessed by users with physical handicaps?	Conduct user test and observe user behavior. X= A / B A= Number of functions successfully accessed B= Number of functions	Examples of physical inaccessibility are inability to use a mouse and blindness	ISO/IEC 9126
1.3.6.3	Operational procedure reduction	Can user easily reduce operation procedures for his/her convenience?	Count user's strokes for specified operation and compare them between before and after customizing operation. X = 1 - A / B	It is recommended to take samples for each different user task and to distinguish between an operator who is a skilled user or a beginner. Number of operation procedures may be represented by counting operation strokes such as click, drug, key touch, screen touch, etc.	

		A = Number of reduced operation procedures after customizing operation B = Number of operation procedures before customizing operation	This includes keyboard shortcuts.

Product quality attributes - attractiveness

Attractiveness is described as: An external attractiveness metric should be able to assess the appearance of the software, and will be influenced by factors such as screen design and color. This is particularly important for consumer products.

No.	Name	Description	Measurement techniques	Notes	Source
1.4.1	Attractive interaction	How attractive is the interface to the user?	Questionnaire to users to assess the attractiveness of the interface to users, after experience of usage		ISO 9126, QUIM
1.4.1.1	Screen design	How attractive is the layout, colors and the text formatting			SCANMIC
1.4.1.1.1	Readability	Ability for the user to easily acquire information provided by system in terms of font and text layout.			SCANMIC, QUIM
1.4.1.1.2	Scanability	Ability for the user to pick out keywords, sentences and paragraphs without actually reading them.			SCANMIC
1.4.1.1.3	Choice of color	Does use of colors improves accessibility, learnability and readability?			SCANMIC
1.4.1.1.4	Space allocation	This refers to proper allocation of space for function and content display provided in a web		Can be generalized to any other application not	SCANMIC

		page to help users focusing their attention.		only web pages.	
1.4.1.2	Media use	The use of media such as graphics, images, animation and audio in software.			QUIM, SCANMIC
1.4.1.2.1	Audio	Use of audio to suit context, for example, instruction, speeches and songs			SCANMIC
1.4.1.2.2	Graphics and images/ simplicity	Minimal use of cosmetic graphic and images Use of graphics and/or images for emphasis. Use of graphics and/or images to attract attention. Labeling of all graphics and images Use thumbnails to display large graphic/images.			SCANMIC, QUIM
1.4.1.2.3	Animation and video	Relevant use of moving pictures media (animation and video)			SCANMIC
1.4.2	Interface appearance customizability	What proportion of interface elements can be customized in appearance to the user's satisfaction?	Conduct user test and observe user behavior. X= A / B A= Number of interface elements customized in appearance to user's satisfaction	This metric is generally used as one of experienced and justified.	ISO 9126, QUIM

			B= Number of interface elements that the user wishes to customize	
1.4.3	Content	Is textual content aligned with the software goals and does it provide optimal amount of credible information in appropriate way?		SCANMIC
1.4.3.1	Language	The quality of a language • Spelling, grammar • Crispness		SCANMIC
1.4.3.2	Scope	Subjective: • Breadth of subject coverage • Depth of subject coverage • Intrinsic value of information Objective: • Suitable language for audience. • Publication and press release • Archive of previously published materials		SCANMIC
1.4.3.3	Authority	Objective: • Name of text or document's author's • Positions or affiliations of text or document's authors • References or sources of text/document • Background information of institution/		SCANMIC

		organization/ owner of the site. i.e. name, address, phone number and email address • Copyright holder statement	
1.4.3.4	Currency	Objective • Resource date • Page revision date	SCANMIC
1.4.3.5	Uniqueness	Objective • Output/ print format as alternative to HTML format • Viewing format other than HTML, for example, PDF and slides • Choices of media type for information, for example, text only, audio or video.	SCANMIC
1.4.3.6	Linkage	Objective • Links to other relevant sites • Links to state and local branches • Links to supporting or sponsoring organizations.	SCANMIC, QUIM
1.4.3.7	Accuracy	Subjective: • High quality writing, good grammar and no spelling or typographical error. • Separation between informational and opinion context	SCANMIC

Product quality attributes - compliance

Compliance is described as: Internal compliance metrics assess adherence to standards, conventions, style guides or regulations.

No.	Name	Description	Measurement techniques	Notes	Source
1.5.1	Usability compliance	How completely does the software adhere to the standards, conventions, style guides or regulations relating to usability?	Specify required compliance items based on standards, conventions, style guides or regulations relating to usability. Design test cases in accordance with compliance items. Conduct functional testing for these test cases.	It may be useful to collect several measured values along time, to analyze the trend of increasingly satisfied compliance items and to determine whether they are fully satisfied or not.	ISO 9126
1.5.2	Security Compliance	Capability of the software product to protect information and data so that unauthorized persons or systems cannot read or modify them and authorized persons or systems are not denied access. [30] Compliance with security standards, conventions and guidelines [27]			QUIM, ISO 9126
1.5.3	Liability compliance	Liability of the software product vendors in case of fraudulent use of users' personal information			QUIM

Product quality attributes - effectiveness

Effectiveness of productivity is described as: The effectiveness with which users employ a product to carry out a task is defined as a function of two components, the quantity of the task attempted by the users, and the quality of the goals they achieve. [20]

No.	Name	Description	Measurement techniques	Notes	Source
1.6.1	Task effectiveness	What proportion of the goals of the task is achieved correctly?	User test. M1 = |1-ΣAi| Ai= proportional value of each missing or incorrect component in the task output	NOTE Each potential missing or incomplete component is given a weight Ai based on the extent to which it detracts from the value of the output to the business or user. (If the sum of the weights exceeds 1, the metric is normally set to 0, although this may indicate negative outcomes and potential safety issues.) (See for example G.3.1.1.) The scoring scheme is refined iteratively by applying it to a series of task outputs and adjusting the weights until the measures obtained are repeatable, reproducible and meaningful.	ISO 9126
1.6.1.1	Quality	What proportion of the time are the user performing productive actions?			MUSiC
1.6.1.1.1	Task completion	What proportion of the tasks is completed?	User test. X = A/B A = number of tasks completed B = total number of tasks attempted	This metric can be measured for one user or a group of users. If tasks can be partially completed the Task effectiveness metric should be used.	

1.6.1.1.2	Error frequency	What is the frequency of errors?	User test. X = A/T A = number of errors made by the user T= time or number of tasks	This metric is only appropriate for making comparisons if errors have equal importance, or are weighted.	ISO 9126, QUIM
1.6.1.2	Quantity				MUSiC

Product quality attributes - efficiency

Efficiency of productivity is described as: An external efficiency metric should be able to measure such attributes as the time consumption and resource utilization behavior of computer system including software during testing or operations**.**

No.	Name	Description	Measurement techniques	Notes	Source
1.7.1	Task proportions	What is the nature of actions user executes to perform a task			MUSiC
1.7.1.1	Productive actions	What proportion of the time are the user performing productive actions?	User test. X = Ta / Tb Ta = productive time = task time - help time - error time -	NOTE This metric requires detailed analysis of a videotape of the interaction (see Macleod M, Bowden R, Bevan N and Curson I (1997) The MUSiC Performance Measurement method Behaviour and Information Technology, 16, 279-293.)	MUSiC, ISO 9126:4

			search time Tb = task time	
1.7.1.2	Snag actions	The user or system performs an action that does not contribute directly or indirectly to the task output, and that cannot be categorized as a help or search action		MUSiC
1.7.1.2.1	Negation actions	User actions that completely cancel or negate previous user or system actions. They always cause cancelled actions.		MUSiC
1.7.1.2.2	Cancelled actions	User or system actions that are completely negated by the user or the system.		MUSiC
1.7.1.2.3	Rejected actions	User actions that are rejected or `ignored' by the system, and that consequently have no effect.		MUSiC
1.7.1.3	Search actions	The user explores the structure of the system displaying parts that are not currently accessed without activating any of the parts that are presented		MUSiC
1.7.1.4	Help actions	The user obtains information about the system, for example by: · referring to the on-line help,		MUSiC

		· reading an instruction manual, · looking at a reference card, · asking a supervisor or analyst for advice, · talking to an appointed expert on the telephone.		
1.7.2	Usefulness	Whether a software product enables users to solve real problems in an acceptable way. Usefulness implies that a software product has practical utility, which in part reflects how closely the product supports the user's own task model. Usefulness obviously depends on the features and functionality offered by the software product. It also reflects the knowledge and skill level of the users while performing some task (i.e., not just the software product is considered).		QUIM
1.7.3	Economic productivity/ Resource utilization	How cost effective is the user? [33]	User test. X = M1 / C M1 = task effectiveness C = total cost of the task	ISO 9126, QUIM
1.7.4	Relative user productivity	How efficient is a user compared to an expert?	User test. Relative user	ISO 9126

			efficiency X = A / B A = ordinary user‘s task efficiency B = expert user‘s task efficiency		
1.7.5	Task time/ Time behavior	How long does it take to complete a task? [33] Capability to consume appropriate task time when performing its function. [30]	User test. X = Ta Ta = task time	Task efficiency measures the proportion of the goal achieved for every unit of time. Efficiency increases with increasing effectiveness and reducing task time. It enables comparisons to be made, for example between fast error-prone interfaces and slow easy interfaces. If Task completion has been measured, task efficiency can be measured as Task completion/task time. This measures the proportion of users who were successful for every unit of time. A high value indicates a high proportion of successful users in a small amount of time.	ISO 9126, QUIM
1.7.6	Task efficiency	How efficient are the users?	User test. X = M1 / T M1 = task effectiveness T = task time		ISO 9126

Product quality attributes – satisfaction

Satisfaction is described as: an external metric to assess whether user is satisfied by using the software.

No.	Name	Description	Measurement techniques	Notes	Source
1.8.1	Overall satisfaction	How satisfied is the user?	User questionnaire.		ISO 9126
1.82	Satisfaction of features	How satisfied is the user with specific software features?			ISO 9126
1.8.3	Satisfaction of usage	What proportion of potential users chooses to use the system?			ISO 9126
1.8.4	Universality	Concerns whether a software product accommodates a diversity of users with different cultural backgrounds (e.g., local culture is considered).			QUIM

Product quality attributes – accessibility

Accessibility is described as: external metric to assess whether users can access the software and its functionality.

No.	Name	Description	Source	Measurement techniques	Notes
1.9.1	Loading time	Objective • Acceptable loading time	SCANMIC, QUIM		
1.9.2	Browser compatibility	Objective • Compatible contents for all main browsers • Compatible contents between different versions of same browser	SCANMIC		
1.9.3	Navigation	Subjective • Appropriate number of sections/categories of contents Objective • Menu/ list of contents in main page • Menu/list of contents in every page • Links to anywhere to anywhere within the site • Minimal number of links to arrive at a particular information • Use of both graphics and text based menu • Accurate and up to date links • Use of sitemap	SCANMIC		

1.9.4	Website accessibility	Objective • Links available in other relevant sites	SCANMIC

Appendix E – Context of use attributes

Context of use attributes – task characteristics

No.	Name	Description	Source
2.1.1	Task goal	What is the main objective of performing the task? For example: to obtain money from bank account as quickly and easily as possible, to type a letter with no mistakes in the minimum amount of time	[18, 33]
2.1.2	Task output	What are the outputs from the task? State the contents and medium of the output. For example, a complete letter with no mistakes, printed on paper, folded and sealed in a correctly addressed envelope.	[18, 33]
2.1.3	Choice	Can users choose whether or not to use the product to achieve their goals? For example, users can obtain money from the bank using the ATM, but during bank opening hours are also able to withdraw money over the counter	[18, 33]
2.1.4	Task side effects	Are there any adverse side effects that may occur as a result of carrying out this task? For example: User may save file and accidentally overwrite another existing file.	[18, 33]

2.1.5	Task steps	Does the main tas consist of any steps or subsequentional tasts?	[18, 33]
2.1.6	Task frequency	How frequently is the task normally carried out? For example: Continuously throughout the day, three or four times a day, once a week etc.	[18, 33]
2.1.7	Task duration	How long does the task generally take the user? For example: Duration ranges between 20 and 35 minutes. In 90% of cases it takes between 25 and 30 minutes.	[18, 33]
2.1.8	Task flexibility	Do users have to follow a pre-defined order when carrying out the task? For example: users are not obliged to follow a pre-defined order, although they normally will due to force of habit	[18, 33]
2.1.9	Task dependencies	What information or resources are required by the users in order to perform the task? For example: an audio tape of dictation, a supply of paper and envelopes, etc. If there are any potential problems in the dependencies being satisfied, these should be noted here.	[18, 33]
2.1.10	Physical and mental demands		[18, 33]
2.1.10.1	Factors which make task demanding	Describe any factors that may make the task physically or mentally demanding.	[18, 33]

		For example; task requires complex split-second decisions to be made	
2.1.10.2	How demanding in comparison with others	How demanding is this task compared to the other tasks in the evaluation? For example; setting up a spreadsheet will be more mentally demanding than entering data onto the same spreadsheet	[18, 33]
2.1.11	Criticality of task output	How critical is the output of the task? Note here if the task output is critical in terms of safety, security or financial integrity. For example: writing software that is to be used to control aircraft in flight, or setting up a spreadsheet controlling the flow of large amounts of money	[18, 33]
2.1.12	Safety	To what extent is this task hazardous to the health or lives of the user or other individuals? For example: commissioning a gas burner which may explode if set incorrectly	[18, 33]
2.1.13	Linked tasks	Does the user normally carry out the task as part of a set procedure? If so, list the tasks that would normally precede or follow this task For example: bank staff processing a loan request must always carry out a credit check before processing the loan	[18, 33]

Context of use attributes – environment characteristics

No.	Name	Description	Source
2.2.1	Organizational Environment	The social or organizational environment in which the work is carried out will affect the way a job is done, the way a product is used, and consequently the usability of the product. This section is concerned with the structure, attitudes and culture of the user's organization. If the product is being used by an individual for his or her own purposes, parts of this section will not be relevant and can be ignored. If two or more user types have been identified for separate evaluation, then it maybe necessary to fill in this section for each of those types	[18, 33]
2.2.1.1	Structure	Here we ask questions about the nature of working relationships, and the flow of information between individuals in the organization	[18, 33]
2.2.1.1.1	Group working	Does the user do the task alone, or in collaboration with other individuals or groups of individuals? If the user collaborates with other individuals, specify their roles and their relationship with the user	[18, 33]
2.2.1.1.3	Assistance	Can assistance be obtained if the user has a problem? Assistance includes the immediate assistance from colleagues in the workplace, as well as assistance via an	[18, 33]

		internal or external telephone 'help line'	
2.2.1.1.4	Interruptions	How frequently is the user generally interrupted while carrying out the task? Describe the frequency and nature of the interruptions. For example, an average of three telephone interruptions per hour	[18, 33]
2.2.1.1.5	Management structure	Who has direct influence on the user's work in the organization? Describe the responsibilities of these individuals, and their relationship with the user. If the product is being used by an individual for his or her own purposes, this question will not be relevant	[18, 33]
2.2.1.1.6	Communications structure	How does information which is related to the user's task flow between individuals inside and outside the organization? Describe the main means of communication between colleagues and/or customers, and the relationships between these individuals. If the product is being used by an individual for his or her own purposes, this question will not be relevant	[18, 33]
2.2.1.1.7	Salary or payment	A nature of benefits the user‘s gain from performing a task	[18, 33]
2.2.1.2	Attitude & culture	This subsection explores the enduring aims, objectives, opinions and common practices demonstrated or espoused by the members of the organization within which the product is used.	[18, 33]

		If the product is being used by an individual for his or her own purposes, this section is not relevant	
2.2.1.2.1	Policy on IT use	What is the organization's policy on the introduction, acquisition and usage of Information Technology? For example: The organization is committed to computerizing all of its procedures over the next ten years. This question will not be relevant for non - IT products	[18, 33]
2.2.1.2.2	Organizational aims	What are the roles, objectives and goals of the user's organization? These may be addressed in an organization's 'mission statement'	[18, 33]
2.2.1.2.3	Industrial relations	What is the status of industrial relations within the company?	[18, 33]
2.2.1.3	Worker/user control	This subsection is concerned with the factors which affect productivity and quality. If the product is being used by an individual for his or her own purposes, this subsection may not be relevant	[18, 33]
2.2.1.3.1	Performance monitoring	How is the quality and speed of the user's work monitored and assessed? For example: Operators are continuously monitored for speed by computer link	[18, 33]
2.2.1.3.2	Performance feedback	How do users receive feedback about the quality and speed of their work? For example: Each week all workers are publicly informed of their productivity; staff have a six-monthly review where their work is discussed with line managers	[18, 33]

2.2.1.3.3	Pacing	How is the rate at which users carry out work controlled? For example: For banking staff, there is customer queue pressure at busy periods; for factory staff, work is paced by the speed of the conveyor belt	[18, 33]
2.2.2	Technical environment	This section is concerned with the technical environment in which the product is used. If two or more user types have been identified for separate evaluation, then it may be necessary to fill in this section for each of those types	[18, 33]
2.2.2.1	Hardware		[18, 33]
2.2.2.1.1	Hardware required to run the product	What hardware is needed to run the product? Examples of hardware are items like the processor, storage devices, input and output devices, networks, gateways, other user equipment	[18, 33]
2.2.2.1.2	Hardware likely to be encountered when using the product	List other hardware usually associated with the product and its user interface environment. For example, when using a personal computer, users will often need to produce output on a printer	[18, 33]
2.2.2.2	Software		[18, 33]
2.2.2.2.1	Software required to run the product	What software is needed to run the product?	[18, 33]

		This may include the operating system or user interface environment. For example, WINDOWS™ may be required to run a particular application	
2.2.2.2.2	Software likely to be encountered when using the product	What software is likely to be encountered when using the product? List other applications usually associated with the product and its user interface environment	[18, 33]
2.2.2.3	Reference materials	What reference materials are provided to help the user learn about the technical environment? For example, manuals on how to operate Windows 3.0 or Apple Macintosh System 7.0. Please note, this does not refer to the instructional materials for the product. These will be listed in the product description	[18, 33]
2.2.2.4	Other equipment	Equipment that is not directly a part of product however user may encounter it during use of the product.	[18, 33]
2.2.3	Physical environment	This section is concerned with the physical environment of the user and product. In many cases a product will be intended for use in a physical environment similar to the standard office working conditions found in Europe (for example, conforming to ISO 9241). In this case, you need put only 'SO' as your answer. Where a feature of the physical environment is non-standard you will need to provide as accurate a description as possible.	[18, 33]

		If two or more user types have been identified for separate evaluation, then it may be necessary to fill in this section for each of those types	
2.2.3.1	Workplace conditions	Here we attempt to identify the physical conditions of the workplace, or the place where the product will be used. If the environment in which the product is used is a Standard Office all sub characteristics may be skipped.	[18, 33]
2.2.3.1.1	Atmospheric conditions	What are the atmospheric conditions of the workplace? If the product is used outdoors then this refers to the weather conditions, otherwise it will refer to the condition of the atmosphere which exists inside buildings such as air quality, speed, humidity etc.	[18, 33]
2.2.3.1.2	Auditory environment	What are the auditory conditions of the workplace? List all types of noise or sound, in particular sounds which would limit interpersonal communication, cause stress or annoyance to the user, or affect the user's perception of sounds relevant to the task	
2.2.3.1.3	Thermal environment	What are the thermal conditions of the workplace? Describe the temperature of the workplace and the heating and air conditioning facilities	[18, 33]
2.2.3.1.4	Visual environment	What are the visual conditions of the workplace?	[18, 33]

		Describe the strength and locations of light sources including natural light. Describe the degree of control the user would have over light conditions including use of blinds etc	
2.2.3.1.5	Environmental instability	Is the workplace physically unstable in any way? e.g., as a result of vibration or any other motion of the workplace	[18, 33]
2.2.3.2	Workplace design	Here we are concerned with the location and design of the workplace, the layout of furniture, and the posture user adopted whilst using the product.	[18, 33]
2.2.3.2.1	Space and furniture	What are the size, layout, and furnishings of the workplace? Include items such as desks, screens, cabling, printers etc.	[18, 33]
2.2.3.2.2	User posture	What posture does the user generally adopt when using the product? For example; standing looking down at a display (height 1.5m)	[18, 33]
2.2.3.2.3	Location		[18, 33]
2.2.3.2.3.1	Location of the product	Where is the product located in relation to the workplace? How is the product located in relation to the furniture of the workplace and the usual working position of the user?	[18, 33]
2.2.3.2.3.2	Location of the workplace	Where is the workplace located? How close is this location to the target area of influence, resources, fellow work colleagues, customers, and	[18, 33]

		the user's home?	
2.2.3.3	Workplace safety	This section inquires about the conditions of the workplace or surrounding environments which may affect the user's health and safety, and require the use of protective clothing or equipment.	[18, 33]
2.2.3.3.1	Health hazards	Are there any conditions of the workplace, or surrounding environment, which may affect the user's physical well being? Include conditions which may affect the user's physical well being in the short term (e.g. by accidents) as well as in the long term (e.g. gradual hearing loss).	[18, 33]
2.2.3.3.2	Protective clothing and equipment	Describe any protective clothing or safety equipment the user is required to wear when in the workplace. This includes such things as clothes or equipment which protects the user from the effects of high or low temperatures. For example : gloves, steel toe-capped boots, face mask	[18, 33]

Context of use attributes – job characteristics

Job characteristics are concerned with details about the jobs carried out by users, i.e. collections of tasks. If the product is not being used in a work environment, then this subsection will not be relevant.

No.	**Name**	**Description**	**Source**
2.3.1	Job function	What is the purpose of the user's work? List the main objectives and responsibilities of the job, as carried out by the user	[35]
2.3.2	Job history		[35]

2.3.2.1	How long employed	Typically, how long have users been employed by the organisation?	[35]
2.3.2.2	How long in current job	How long have users been doing their current job in this organisation?	[35]
2.3.3	Hours of work or operation		[35]
2.3.3.1	Hours of work	What hours do users work? Provide details about the hours of work of the user, including shift work, irregular hours, home working hours, etc.	[35]
2.3.3.2	Hours of using product	What hours do users spend using the product? Provide details about when the product will be used; for example, the product is used throughout the shift which can either be early, i.e. 0500-1300 hrs or late i.e. 1300 to 2200 hrs. Workers alternate between weeks on early and late shifts.	[35]
2.3.4	Job flexibility	Can users decide how to approach the job, organise their time and carry out tasks?	[35]

Context of use attributes – temporal context

No.	**Name**	**Description**	**Source**
2.4	Temporal context	Temporal context as a UX component refers to the time period that the user is able to dedicate for the system given the context restrictions, e.g. finding out which bus to take before missing it. In case of multitasking, the period dedicated to the system is split to many pieces, e.g. browsing while waiting for a bus, continuing the session in the bus, and later at home.	[50]

Appendix F – List of user attributes

No.	Name	Description	Source
3.1.	Training	This includes formal training as well as less formal methods such as open learning packages, video instruction or training manuals. State the amount of training users have received in each of the following areas:	[18, 33]
3.1.1	Training in tasks supported by product main functions	In performing tasks supported by the product's specific functions, manually or with any automated system.	[18, 33]
3.1.1.2	Training in using the product main functions	In using the product itself to perform the specific functions, as listed in the Product Report	[18, 33]
3.1.1.3	Training in using other products with similar main functions	In using other products to perform similar functions.	[18, 33]
3.1.1.4	Training in using products with the same interface style or operating system	For computer-based products only: In using the same operating system or environment, or other products based on it. For example, a one day course of instruction in using WINDOWS™	[18, 33]

3.1.2	Experience		[18, 33]
3.1.2.1	Product experience	How much practical experience have users had in using the product for its main functions? IMPORTANT: you should refer at this point to your completed Product Report. List the practical experience users have of using the product for its main functions, as listed in the Product Report. e.g. • Function 1: Daily use • Function 2: No experience • Function 3: Used less than once a month	[18, 33]
3.1.2.2	Experience with products with similar functions	How much practical experience have these users had in using other products performing similar functions? List for each main function.	[18, 33]
3.1.2.3	Experience with same interface style or operating system	For computer-based products only: state how much practical experience users have in using the operating system or environment on which the product is based. For example, for a UNIX™-based product, state experience with other UNIX™-based applications; for a WINDOWS™ - based product state experience with other WINDOWS™-based applications	[18, 33]
3.1.3	Training and experience in	How much practical experience does this group of users have in performing, either manually or with any automated system, the tasks that this product supports?	[18, 33]

	business processes	For example, for a financial package, how much experience do these users have of the accounting procedures performed using each of the product's main functions? Without experience of accounting procedures, it may be difficult to use some functions. Or for an ATM (cashpoint machine), when a task is withdrawing money from a bank account, if users are experienced in withdrawing money over the counter, but not in using the cashpoint, their experience with the product will be low, but task experience considerably higher.	
3.1.4	Task knowledge	Level of user's knowledhe related to the task	[18, 33]
3.1.5	Organizational knowledge	Level of user's knowledhe on the organization structure	[18, 33]
3.1.6	Background knowledge	Is there any general background knowledge which is indirectly relevant to the users' performance of tasks with the product? Background knowledge is knowledge which is not directly connected to the product, the task, or IT, but which user's may have due to membership of a social, cultural, organizational, regional, national or religious group. An example of background knowledge could be that company telephone operators are not on duty after 6.00pm.	[18, 33]
3.1.7	Qualifications	What range and distribution of qualifications might members of this user group typically have? Include formal and informal qualifications; e.g., degrees, apprenticeships	[18, 33]
3.1.8	Input device skills	What input device skills do they possess?	[18, 33]

		For example: regular user of mouse; touch typing (60 to 90 wpm), fast two finger typing or slow 'hunt and peck; familiarity with a touch screen, etc.	
3.1.9	Language skills	State any deficiencies users may have in the language in which the product and its documentation have been written	[18, 33]
3.2	Physical attributes	This subsection is concerned with the physical characteristics of the user type	[18, 33]
3.2.1	Age		[18, 33]
3.2.1.1	Age range	What is the age range of the user type? For example: Age ranges between 16-70 years.	
3.2.1.2	Typical age	If appropriate, state the typical age of this user group	[18, 33]
3.2.2	Gender	What is the male/female distribution of the user type? For example: 10% male, 90% female	[18, 33]
3.2.3	Physical capabilities and limitations	Describe any physical limitations or disabilities of the user type. This includes general physical limitations - such as reach distances, as well as physical disabilities. Examples of such disabilities are short sightedness, color blindness, loss of hearing, loss of limbs, reduced psychomotor capabilities	[18, 33]
3.3	Mental attributes	This section asks about the mental characteristics of this type of intended user, including their intellectual abilities and motivations	[18, 33]
3.3.1	Intellectual abilities		[18, 33]
3.3.1.1	Distinctive abilities	Do the users possess any distinctive intellectual abilities?	[18, 33]

3.3.1.2	Specific mental disabilities	Do the users have any specific relevant mental disabilities?	[18, 33]
3.3.2	User experience	A person's perceptions and responses that result from the use or anticipated use of a product, system or service	
3.3.2.1	Visceral quality	the visceral level of emotion that is emotional responses formed by physical senses such as ―looks nice,” ―feel cold,” etc.	[50-51]
3.3.2.2	Cognitive quality	the behavioral level of emotion that is formed from cognitive processes such as ―easy to use,” “simple to use,” ―hard to figure out,” etc.	[50-51]
3.3.2.3	Usefulness quality	an indirect quality related to their overall needs rather than a certain specific type of emotional response, such as ―fulfills what I need,” “practical”, “does what I want it to do,” etc.	[50-51]
3.3.2.4	Reflective and social quality	the reflective level of emotion such as ―it is a trend,” “creates an artificial world,” “everyone has these products,” etc.	[50-51]
3.3.3	Motivations	How positive or negative are the attitudes which the users display? (give reasons where helpful): For example: highly satisfying work despite low rates of pay, proud of products produced, suspicious that	[18, 33]

		the introduction of IT will lead to loss of jobs, lack of trust with higher management.	
3.3.3.1	Attitude to job and task	Users group attitude towards their job and task	[18, 33]
3.1.3.2.2	Attitude to the product	Users group attitude towards the product	[18, 33]
3.3.3.2.3	Attitude to information technology	Users group attitude towards information technology	[18, 33]
3.3.3.2.4	Employees attitude to employing organization	Users group attitude towards employing organization.	[18, 33]
3.3.4	Emotional attributes		[50-51]
3.3.4.1	Needs	User‘s tacit needs from a product	[50]
3.3.4.2	Emotional state	User‘s mood	[50]
3.3.4.3	Experiences	User‘s emotional experiences related to a product.	[50]
3.3.4.4	Expectations	Users expectations from a product	[50]

References

1. J. Boerg, 2008, A New standard for Quality Requirements, IEEE Software March/April 2008
2. B. Behkamal, M. Kahani, M. K. Akbari, 2009, Customizing ISO 9126 quality model for evaluation of B2B applications, Information and Software Technology 51 (2009) 599–609
3. Robert J. K. Jacob and Keith S. Karn, Eye Tracking in Human–Computer Interaction and Usability Research: Ready to Deliver the Promises
4. Victor R. Basili, 2002, Expermentation in Software Engineering, Experimental Software Engineering Group, Institute for Advanced Computer Studies, Department of Computer Science, University of Maryland and Fraunhofer Center for Experimental Software Engineering -Maryland
5. J. Biolchini, P. Gomes Mian, A. C. C. Natali, G. H. Traverssos, 2005, Systematic Review in Software Engineering, Technical Report EX 679/05
6. Y. C. Cavalcanti, Experimentation in Software Engineering, presentation slides
7. I. Douglas, 2006, Collaborative International Usability Testing: Moving from document-based Reporting to Information Object Sharing, IEEE International Conference on Global Software Engineering (ICGSE'06)
8. J.J. Garret, The Elements of User Experience, presentation slides
9. V. Roto, K. Väänänen-Vainio-Mattila, E. Law and A. Vermeeren, 2009, User Experience Evaluation Methods in Product Development (UXEM‘09), T. Gross et al. (Eds.): INTERACT 2009, Part II, LNCS 5727, pp. 981–982, 2009.
10. Zhu Wang, Xingshe Zhou, Haipeng Wang, Hongbo Ni, Ruijuan Wu, 2008, A Quantitative Evaluation Model of Group User Experience, IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application
11. B. Kitchenham, S. Charters, D. Budgen, P. Brereton, M. Turner, S. Linkman, M. Jørgensen, E. Mendes, G. Visaggio, 2007, Guidelines for performing Systematic Literature Reviews in Software Engineering, Version 2.3, EBSE Technical Report, EBSE-2007-01
12. J. Tan, 2009, FOUUX – A Framework for Usability & User Experience, Master Thesis in Blekinge Institute of Technology, School of Computing, MSE-2009-23
13. S. Wiedenbeck, S. Davis, 1997, The influence of interaction style and experience on user perceptions of software packages, Int. J. Human-Computer Studies (1997) 46, 563-588

14. J. Nielsen, 1994, Usability Inspection Methods, CID‘94- ―Celebrating Interdependence”, conference paper
15. J. W. Creswell, 2003, Research Design, Sage Publications
16. J. Nielsen, 2005, Ten usability heuristics, ISSN 1548-5552
17. J. Hom, 2003, Heuristic Evaluation
18. N. Bevan, M. Macleod, 1994, Usability measurement in context, Behaviour and Information Technology, 13, 132-145 (1994)
19. Blandford, Ann E. , Hyde, Joanne K. , Green, Thomas R. G. and Connell, Iain(2008) 'Scoping Analytical Usability Evaluation Methods: A Case Study', Human-Computer Interaction, 23: 3, 278 — 327
20. Macleod, Miles , Bowden, Rosemary , Bevan, Nigel andCurson, Ian(1997) 'The MUSiC measurement method', Behaviour & Information Technology, 16: 4, 279 — 293
21. S. Winter, S. Wagner and F. Deissenboeck, A Comprehensive Model of Usability, Institut für Informatik, Technische Universität München
22. E. Frøkjær, M. Hertzum, K. Hornbæk, 2000, Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated?, Proceedings of the ACM CHI 2000 Conference on Human Factors in Computing Systems (The Hague, The Netherlands, April 1-6, 2000), pp. 345-352. ACM Press, New York. Preprint version
23. J. Bendek, T.Miner, 2002, Measuring Desirability: New methods for evaluating desirability in a usability lab setting, Usability Professionals' Association, 2002 Conference Proceedings.
24. P. Biernacki, D. Waldorf, Snowball Sampling: Problems and Techniques of Chain Referral Sampling, Sociological methods & research, Vol. 10 No. 2. November 1981 141-163
25. N. Bevan, 1995, Human – Computer Interaction Standards, Proceedings of the 6th International Conference on Human Computer Interaction, Yokohama, July 1995. Anzai & Ogawa (eds), Elsevier.
26. N. Bevan, International Standards for HCI and Usability, International Journal of Human Computer Studies, 55(4), 533-552
27. ISO/IEC 9126, Reference number ISO/IEC 9126-1:2001(E)
28. D. D. Heckathorn, 1997, Respondent – Driven Sampling: A New Approach to the Study of Hidden Populations, Social Problems, Vol. 44, No. 2 (May, 1997), 174-199
29. J. Faugier, M. Sargeant, 1997, Sampling hard to reach hidden populations, Journal of Advanced Nursing, 1997, 26, 790–797

30. A. Seffah , M. Donyaee, R. B. Kline, H. K. Padda, 2006, Usability measurement and metrics:A consolidated model, Software Qual J (2006) 14: 159–178, DOI 10.1007/s11219-006-7600-8
31. A. Seffah, N. Kececi, M. Donyaee, 2001, QUIM: A Framework for Quantifying Usability Metrics in Software Quality Models, 07695-1287-9/01, IEEE
32. S. Hassan, F. Li, Identifying Web Usability Criteria: The ‗Scanmic' Model
33. ISO/IEC TR 9126-4:2004(E)
34. M. Maguire, Context of Use within usability activities, 2001, Int. J. Human-Computer Studies (2001) 55, 453}483 doi:10.1006/ijhc.2001.0486
35. C. Thomas, N. Bevan, 1996, Usability Context Analysis: A Practical Guide, Serco Usability Services
36. ISO/IEC TR 9126-2:2003(E)
37. Sung H. Han, Myung Hwan Yun, Kwang-Jae Kim, Jiyoung Kwahk, 2000, Evaluation of product usability: development and validation of usability dimensions and design elements based on empirical models, International Journal of Industrial Ergonomics 26 (2000) 477}488
38. C. Wohlin, P. Runneson, M. Host, M. C. Ohlsson, B. Regnell, A. Wesslen, 2000, Experimentation in software engineering , Kluwer Academic Publishers Norwell, MA, USA ©2000 ISBN:0-7923-8682-5
39. M. V. Zelkowitz, D. Wallace, 1997, Experimental validation in software engineering, Information and Software Technology 39 (1997) 735-743
40. R. Kivetz, O. Urminsky, Y. Zheng, 2006, The Goal-Gradient Hypothesis Resurrected: Purchase Acceleration, Illusionary Goal Progress, and Customer Retention, Journal of Marketing ResearchVol. XLIII (February 2006), 39–58
41. Hartson, H. Rex , Andre, Terence S. and Williges, Robert C.(2001) 'Criteria For Evaluating Usability Evaluation Methods', International Journal of Human-Computer Interaction, 13: 4, 373 — 410
42. E. Baauw, M. M. Bekker, and W. Barendregt, A Structured Expert Evaluation Method for the Evaluation of Children's Computer Games
43. Bent Flyvbjerg, 2006, Five Misunderstandings About Case-Study Research, DOI: 2006 12: 219 Qualitative Inquiry, 10.1177/1077800405284363
44. R. Fidel, 1984, The Case Study Method: A Case Study, LISR 6, 273-288 (1984)
45. M. V. Zelkowitz, D.R. Wallace, 1998, Experiment Models for Validating Technology, Computing Practices, May 1998

46. S. Soy, 1997, The Case Study as a Research Method, Uses and Users of Information, LIS 391D.1 - Spring 1997
47. W. Tellis, 1997, Application of a Case Study Methodology, The Qualitative Report, Volume 3, Number 3, September, 1997
48. B. Kitchenham, L. Pickard, 1995, Case Studies for Method and Tool Evaluation, IEEE SOFTWARE, JULY 1995
49. M. Csikszentmihalyi, Harper and Row, 1990, FLOW: The Psychology of Optimal Experience
50. V. Roto, User Experience Building Blocks, Nokia Research Center
51. Youn-kyung Lim, Justin Donaldson, Heekyoung Jung , Breanne Kunz, David Royer, Shruti Ramalingam, Sindhia Thirumaran, Erik Stoltermam, Emotional Experience and Interaction Design
52. D. Swallow, 2004, Smart Phone; Confused User: Theory and methods for understanding the user experience of smartphones, Master thesis, Department of Computer Science at the University of York.
53. E. Law, J. Kort, V. Roto, M. Hassnzahl, Arnold P.O.S. Vermeeren, Towards a Shared Definition of User Experience, Proceedings · Special Interest Groups April 5-10, 2008 · Florence, Italy
54. K. Halvorson, 2010, Intentional Communication: Expanding our Definition of User Experience Design, Interactions May - June 2010
55. M. Obrist, V. Roto, K. Väänänen-Vainio-Mattila, User Experience Evaluation – Do You Know Which Method to Use?, CHI 2009 ~ Special Interest Groups April 4-9, 2009 ~ Boston, MA, USA
56. N. Bevan, What is the difference between the purpose of usability and user experience evaluation methods?, Professional Usability Services
57. Why Standards Are Important, IHS Whitepaper, ©IHS Inc. www.ihs.com
58. M. Skoglund and P. Runeson, Reference-based search strategies in systematic reviews
59. F. Jambon, B. Meillon, User Experience Evaluation in the Wild, CHI 2009 ~ Spotlight on Works in Progress ~ Session 1 April 4-9, 2009 ~ Boston, MA, USA
60. K. Beckman, S. Khajenoori, N. Coultier, N. R. Mead, 1997, Collaborations: Closing the Industry–Academia Gap, IEEE SOFTWARE, NOVEMBER/DECEMBER 1997
61. T. Gorschek, Software Process Assessment & Improvement in Industrial Requirements Engineering, Blekinge Institute of Technology Licentiate Series No 2004:07, ISSN 1650-2140, ISBN 91-7295-041-2

62. Y. Mashiko, V. R. Basili, Using the GQM Paradigm to Investigate Influential Factors for Software process Improvement, J. Systems software 1997; 36; 17-32
63. Website: http://www.interfaces.com/blog/2010/09/the-history-of-usability/, accessed 2011.10.22

Such views conclude with a system in which product, context of use and user influences each other. See the following figure:

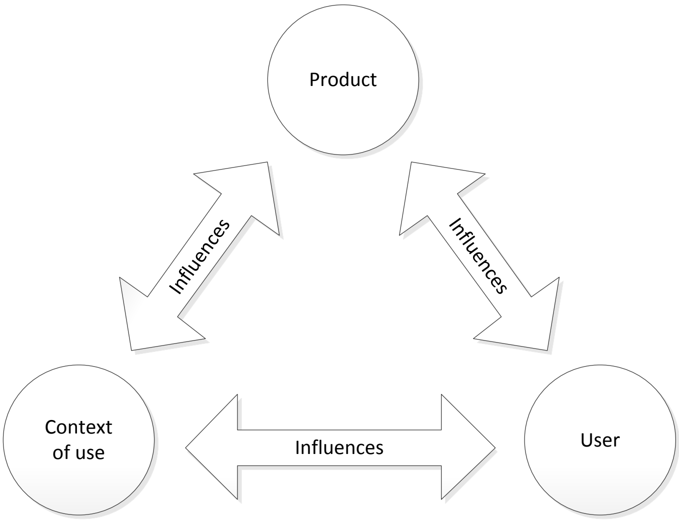

Figure 12, connection between usability, user experience and context of use

The perception of product features is influenced by context of use. The product influences user by satisfying his needs and expectations. User influences context of use by adding his part to temporal context (and maybe other contexts as well).

The other way around, context of use influence user and its perception of a product. The product alters the context.

Such view on the connection between usability and user experience allows putting both concepts in the single evaluation model and apply similar evaluation methods on both.

4.2 Level II – usability and user experience evaluation guidelines

Second level of the model contains guidelines on how to perform usability and user experience evaluation. Such guidelines combined with structure of usability and user experience from Level II allows performing practical usability and user experience evaluation.

Further sections describe and motivate adaptation process. It is assumed that the reader is familiar with concepts of empirical investigation methods. For the purpose of readability some widely known aspects of empirical investigation described here in detail. For more detailed information please refer to the references.

Usability and user experience is a complex concept and can be evaluated by assessing its influencing factors (attributes). [36, 20] However, complete assessment of all usability and user experience attributes is too complex and resource demanding task to be practically applicable.