

## **The use of evaluation theory and square standards to develop user needs experience evaluation method**

**Mohammad Zarour**

College of Computer and Information Sciences (CCIS) Prince Sultan University, Riyadh  
Kingdom of Saudi Arabia

**Abstract. Introduction:** Technology is changing the way we used to live. Software applications including mobile applications are widely spread nowadays. The increased competitive advantage in application industry gradually led to increasing the challenges in dealing with growing user requirements as well as their demand for high quality. In order to survive in the highly competitive market, the software development companies should not only work to satisfy user requirements but also seek to satisfy user experience. Accordingly, studying, measuring and improving user experience is crucial for the success of any software product.

**Method:** This research focuses on evaluating user experience needs by developing user experience needs evaluation method based on three main disciplines: the user experience framework, the evaluation theory concept and the SQuaRE standards in ISO/IEC 25022 and ISO/IEC 25023. Although these disciplines are available in the literature, they are not linked together to complete the mosaic picture of developing a UX evaluation methods. Linking these three disciplines led to systematically identify the necessary evaluation criteria to evaluate user needs experience.

**Results:** A user experience needs evaluation method is developed in this paper and used to conduct a case study to evaluate user needs experience of a mobile application.

**Conclusion:** Using the evaluation theory, UX frameworks and the SQuaRE standard to develop UX evaluation method was very useful. The proposed user experience needs evaluation method is promising to evaluate various software applications including mobile applications. More empirical research is needed to evaluate the developed method.

**Keywords.** User needs experience; User experience needs evaluation method; Evaluation theory; User experience framework; SQuaRE standards.

### **1. Introduction**

Technology is changing the way we used to live. Today, people constantly use information systems almost everywhere to accomplish tasks and achieve their needs. Software applications including mobile applications are widely spread nowadays. People take advantage of the hundreds of existing software application to accomplish daily tasks. Increasing competitive advantage in application industry gradually led to increasing the challenges in dealing with growing user requirements as well as their demand for high quality. In order to survive in this highly competitive market, the software development companies should not only work to satisfy the user requirements but also seek to achieve a satisfying user experience [1].

Over the last years, the field of Human-Computer Interaction (HCI) becomes more interested in User Experience (UX). Recently, many conferences, workshops, forums and similar activities held aiming to better understand user experience and develop a universal definition of user experience. The outcome of this interest is a number of various viewpoint and definitions of user experience [2], Despite the diversity in definitions of user experience, all authors agree that “it is more than just a product's usefulness and usability” [3], User experience is the result of interaction with product, system or service, influenced by a set of aspects. Furthermore, most of researchers and participants agree that user experience “is dynamic, context-dependent, and subjective” [2]. The user experience concept has an interesting contribution in classical Human-Computer Interaction, the user experience “attempt to include subjective attributes like, for instance, aesthetics, emotional, and social aspect in a design space which has previously mainly concerned with ease of use” [4].

Consequently, the success or failure of a software product is affected by the users of that product. Accordingly, studying, measuring and improving user experience is crucial to the success of any software

product. Although various researches have been done in studying user experience and its evaluation methods, non-is found to be based on strong bases in regard to developing the evaluation method, for instance using the evaluation theory as bases to develop the evaluation method. From this perspective, we conducted our research, this research focuses on evaluating user needs experience by developing an evaluation method based on three main disciplines the UX framework, the evaluation theory concept and the SQuaRE standards in ISO/IEC 25022 and ISO/IEC 25023.

The paper is structured as follows. Section 2 discusses the background concepts of evaluation, user experience framework, the SQuaRE standards, and evaluation theory. Section 3, discusses the mapping process between user experience aspects of user experience framework and quality attributes of ISO 25000 series SQuaRE standards. Section 4 discusses the application of the key concepts of evaluation to design and develop the proposed user experience evaluation methods. Section 5, Presents the cases study applies the developed evaluation method to evaluate user needs experience of a mobile application own by one of the main telecommunication companies in the kingdom. Section 6, presents the conclusion of our study and future works.

## **2. Background**

To develop the UX evaluation method, three main disciplines are adopted: the user experience framework as stated by Alharbi [5], the evaluation theory concept [6][7] and the SQuaRE series of international standards in ISO/IEC 25022 and ISO/IEC 25023 [8], [9]. Those disciplines are briefly described in this section.

### **2.1. User experience framework**

Recently, research efforts have been undertaken to collect, consolidate and categorize user experience influence factors, attributes or aspects in a comprehensive framework. Those frameworks can be used as starting step for understanding how the user experience got effective and how to measure it in software applications include mobile applications.

Alharbi proposed a comprehensive framework based on systematic literature review. The proposed framework identified four main UX dimensions and the relationship between these dimensions and the user experience aspects that have a direct or indirect impact on the user experience. The user experience aspects that have been identified are categorized into seven categories and assigned to the corresponding user experience dimension. Furthermore, the UX framework identified fourteen generic methods that can be used to measure user experience aspects (see Figure 1) [5]. Our research work focuses on one dimension, the user needs experience dimension, to develop user experience needs evaluation method.

### **2.2. ISO 25000 series standard (SQuaRE)**

“The quality of a system is the degree to which the system satisfies the stated and implied needs of its various stakeholders, these stated and implied needs are represented in System and Software product Quality Requirements and Evaluation (SQuaRE) series of standards by quality models that categorize system quality into characteristics, which in some cases are further subdivided into sub-characteristics. It is important that the quality characteristics are specified, measured, and evaluated whenever possible using validated or widely accepted measures and measurement methods” [10]. Consequently, the SQuaRE series of standards defined three quality models can be used to identify relevant quality characteristics that can be further used to establish requirements, their criteria for satisfaction and the corresponding measures.



Figure 1: User Experience Framework [5]

**SQuaRE division.** The SQuaRE divisions consist of a series of international standards organized in divisions under the general title Software Product Quality Requirements and Evaluation. Figure 2 illustrates the organization of the SQuaRE series representing families of standards, further called Divisions.

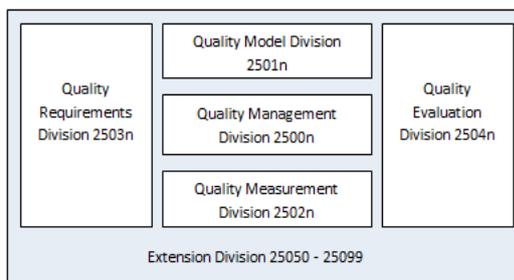


Figure 2: Organization of SQuaRE series of International Standards [10]

In this research context, we will adopt ISO/IEC 25000 series to develop evaluation criteria for the proposed user needs experience evaluation methods. Specifically the following parts of International Standards:

- (1) ISO/IEC 25010 – system and software quality models  
 ISO/IEC 25010 is a part of ISO/IEC 2501n - Quality Model Division, it defined two quality models: a quality in use model and a product quality model. The quality models together serve as a framework to ensure that all characteristics of quality are considered.
  - (a) a quality in use model: The quality in use model composed of five characteristics: effectiveness, efficiency, satisfaction, freedom from risk and context coverage. Which are further subdivided into sub-characteristics that related to outcomes of interaction when a product is used in a particular context of use. “This model is applicable to both computer systems in use and software products in use” [10].
  - (b) a product quality model: The product quality model composed of eight characteristics: functional suitability, performance efficiency, compatibility, usability, reliability, security, maintainability, and portability. Which are further subdivided into sub-characteristics that relate to static properties of software

and dynamic properties of the computer system. “This model is applicable to both computer systems and software products” [10].

(2) ISO/IEC 25022 – measurement of quality in use

ISO/IEC 25022 is a part of ISO/IEC 2502n - Quality Measurement Division. It provides a set of quality measures, including associated measurement functions for the characteristics and sub-characteristics in the ISO/IEC 25010 quality in use model, which can be used for specifying quality in use requirements and measuring and evaluating quality in use [8].

(3) ISO/IEC 25023 – measurement of system and software product quality

ISO/IEC 25023 is a part of ISO/IEC 2502n - Quality Measurement Division. It provides a set of quality measures, including associated measurement functions for the characteristics and sub-characteristics in the ISO/IEC 25010 product quality model, which can be used for specifying requirements, measuring and evaluating the system/software product quality [9].

### 2.3. Evaluation theory concept.

“Evaluation theory is a belief, policy, or procedure proposed or followed as the basis of action.” [11]. With reference to the discipline of evaluation, Shadish, Cook and Leviton [6] summarized the different evaluation theories presented by seven well-known theorists. “Scriven’s theory can be assumed to be at the highest level of abstraction as he described principles, concepts, and methods for any scenario of knowledge construction in evaluation”.

**Basic evaluation components.** Scriven’s theory of evaluation as mentioned in [6] “attempts to clarify the logic behind evaluations” [12]. Scriven’s identified six evaluation components involved in the most common type of evaluation. Figure 3 shows the six components of evaluation and each component is further explained as following [12], [13]:

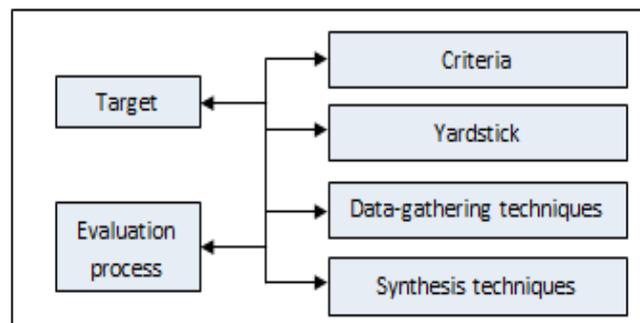
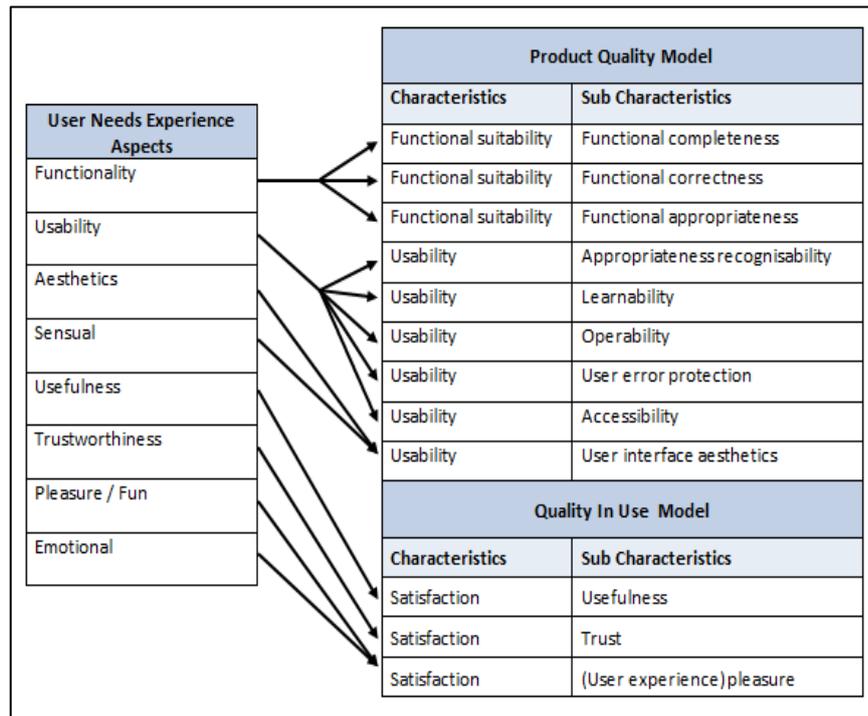


Figure 3: Components of an Evaluation [37]

- (1) Target, it is the object of the evaluation.
- (2) Criteria, the characteristics of the target.



(3) Standards/yardstick, assessment techniques, the ideal form of the target based on the optimality of defined

Figure 4: Mapping UX aspect with characteristics of SQuaRE standards

criteria makes up the standard or yardstick against which a real target is to be matched.

- (4) Assessment techniques, apart from building the yardstick, the potentially applicable data gathering techniques need to be identified and assigned to each evaluation criterion.
- (5) Synthesis techniques, techniques used to judge each criterion and, in general, to judge the target and obtaining the results of the evaluation.
- (6) Evaluation process, series of activities and tasks that are to be executed to perform an evaluation.

### 3. Mapping User Experience Aspects and ISO 25000 Series (SQuaRE) Standards

The user needs experience aspects of UX framework [5] define quality factors that should be met to satisfy users and meet their expectation. Hence, these aspects can be mapped to the software quality attributes of ISO 25000 (SQuaRE) standard [10]. The output of mapping process is a criteria tree used to evaluate user needs experience. Figure 4 illustrates the mapping of UX aspects with SQuaRE characteristics and sub-characteristics.

### 4. Develop User Experience Evaluation Method Based on Evaluation Theory

The six general components of evaluation theory can be applied in all kind of evaluation work [12]. Hence, we will adopt the six components of evaluation theory to develop user experience needs evaluation methods. Furthermore, we will adopt the result of mapping user needs experience aspects with SQuaRE characteristics and sub-characteristics (see Figure 4), to construct the core components of evaluation theory. The application of the three main disciplines used to develop the proposed user experience needs evaluation method is summarized in Figure 5. The application of these evaluation components to a particular evaluation involves, selecting the best-suited type of evaluation method considering mainly the target under analysis [13]. In the context of developing the user experience evaluation methods and based on House classification [14], the objective method named "control-oriented method" has been selected, the objective of this method is to ensure that the target is controlled and directed by the yardstick specified [15]. In this research context, the evaluation of user needs experience should be controlled and directed by the yardstick developed in this

research. The six evaluation components, shown in Figure 3, constitute the basis for the evaluation framework. The evaluators can develop user needs experience evaluation method by instantiating the evaluation framework. Further explanation of evaluation components discussed below.

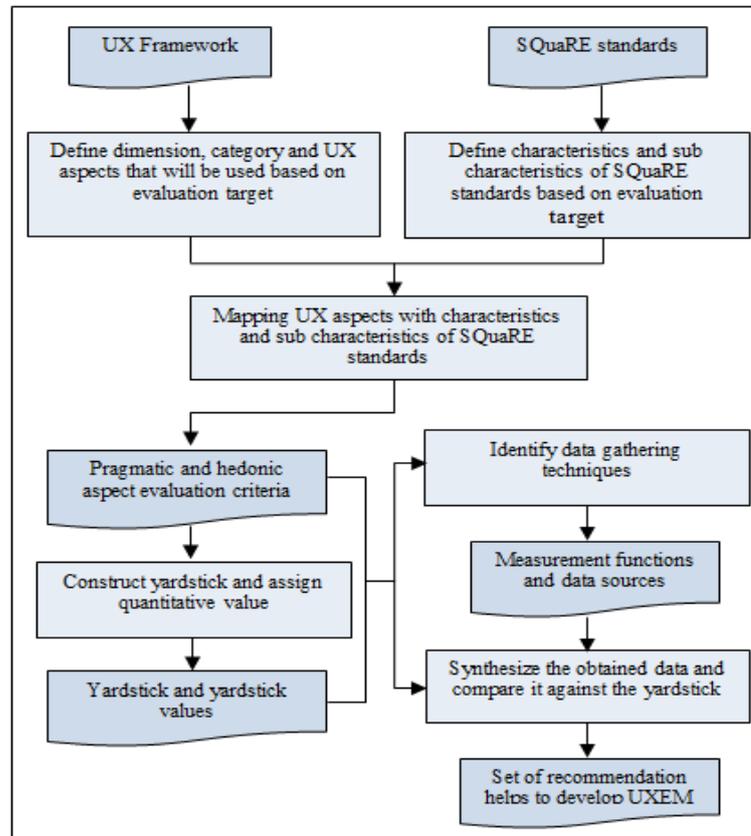


Figure 5: Develop user experience evaluation method

**2.4. Target.**

Target delimitation is the first essential step in any evaluation. “To be able to identify the criteria evaluation component, it is necessary to study and delimit the object under evaluation, which means identifying the factors to be considered” [13]. In this research context, the user needs experience is the target. The user needs experience focuses on the underlying user needs, both pragmatic and hedonic needs are the key to providing good systems that meet customers’ needs and thereby contribute to the business success [16]. Studying the user needs experience relies on studying the user needs aspects. The user needs aspects divided into two categories, the pragmatic aspects and the hedonic aspects[16]–[18]. Figure 6 shows the components of target evaluation criteria.

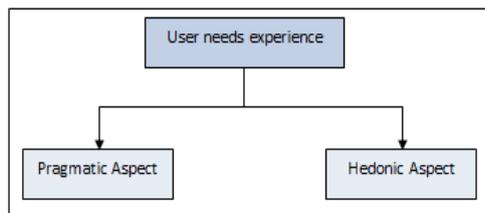


Figure 6: Components of target evaluation criteria

2.5. Criteria

“Defining the criteria is the second essential and critical step for developing a method of evaluation. Having ascertained and delimited the target, it is necessary to identify what characteristics of the target are of interest for evaluation purposes. These characteristics are referred to as evaluation criteria. Criteria elicitation can be made by either using an obligatory standard that contains implicitly the criteria to be applied in the evaluation or by using other diverse techniques for criteria elicitation” [13].

In this research context, the criteria will be developed using one approach: the approach is based on the needs assessments elicitation method. In this approach, the needs to evaluate user needs experience are analyzed. These needs are represented by a set of user needs aspects (pragmatic aspects, hedonic aspects) defined in the user experience framework [5]. The proposed evaluation criteria formed the basis of general evaluation criteria, which are used to evaluate user needs experience. Due to the fact that the general criteria and specific criteria are interrelated, “it is possible to draw a diagrammatic tree that contains all the general and specific criteria that are to be evaluated” [5]. In order to construct user need experience criteria tree, we adopted the result of mapping UX aspects with SQuARE characteristics and sub-characteristics in section 3. The process eventually led to construct pragmatic aspect evaluation criteria tree (see Figure 7) and hedonic aspect evaluation criteria tree (see Figure 8) based on international standards. This criteria tree is the basis for developing the evaluation yardstick in section 4.1.3.

(1) pragmatic aspect.

Pragmatic aspect refers to “the system’s perceived ability to support the achievement of tasks and focuses on the system’s actual usability in completing tasks, that are the ‘do-goals’ of the user” [19], [20]. These aspects can be measured using technical characteristics of the developed software which can be found in the technical reports. Using SQuARE terminology, these criteria are mainly evaluating software product quality related to UX (internal and external quality). The evaluation part that corresponds to these criteria requires technical document review. The user needs experience pragmatic aspects are divided into three general evaluation criteria each of which is divided into a set of specific evaluation criteria as shown in Figure 7. General evaluation criteria refer to “characteristics that cannot be assigned a value directly and require further decomposition to which the set of questions will be applied successively until specific criteria are obtained [13]. Specific evaluation criteria refer to “characteristics that can be assigned a value directly using a particular data-gathering technique” [13].

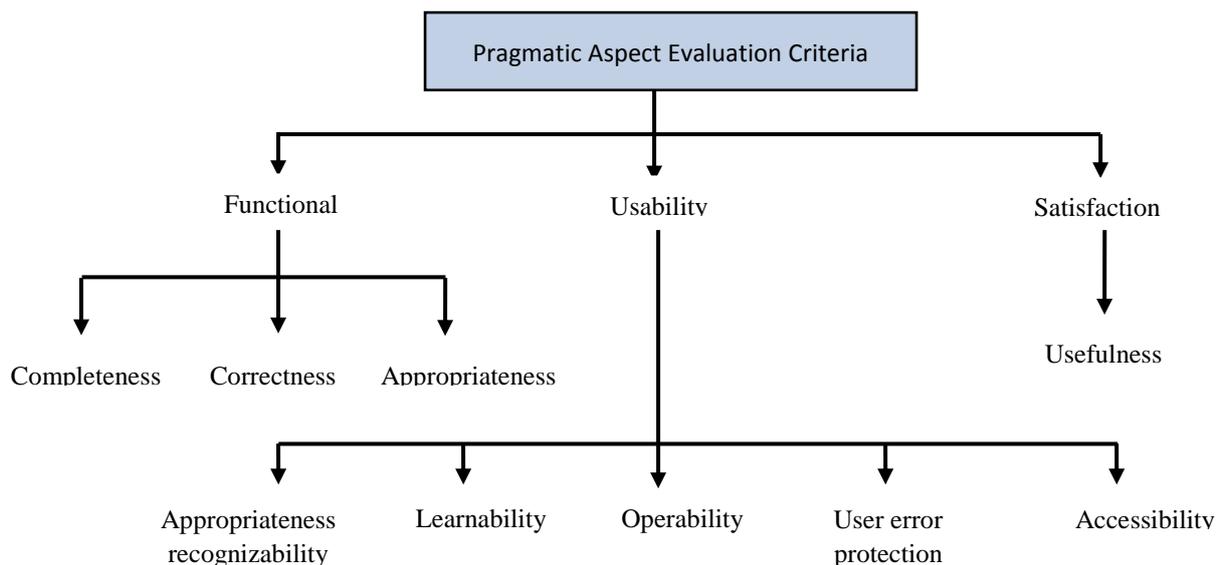


Figure 7: Pragmatic aspect evaluation criteria tree

Table 1 illustrates the specific definition of the evaluation criteria. The criteria tree and these definitions aid in the accurate understanding of the yardstick and this would help the evaluator know exactly what characteristics are to be analyzed.

Table 1: General and specific evaluation criteria description of pragmatic aspect [10]

General evaluation criteria	Specific evaluation criteria	Description
Functional suitability	---	Functional suitability is the “Degree to which a product or system provides functions that meet stated and implied needs when used under specified conditions”.
	Functional completeness	“Degree to which the set of functions covers all the specified tasks and user objectives”.
	Functional correctness	“Degree to which a product or system provides the correct results with the needed degree of precision”.
	Functional appropriateness	“Degree to which the functions facilitate the accomplishment of specified tasks and objectives”.
Usability	---	Usability is the “Extent to which a product or system can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use”.
	Appropriateness recognizability	“Degree to which users can recognize whether a product or system is appropriate for their needs”.
	Learnability	“Degree to which a product or system can be used by specified users to achieve specified goals of learning to use the product or system with effectiveness, efficiency, freedom from risk and satisfaction in a specified context of use”.
	Operability	“Degree to which a product or system has attributes that make it easy to operate and control”.
	User error protection	“Degree to which the system protects users against making errors”.
	Accessibility	“Degree to which products and systems can be used by people with the widest range of characteristics and capabilities to achieve a specified goal in a specified context of use”.
Satisfaction	---	“Degree to which user needs are satisfied when a product or system is used in a specified context of use”.
	Usefulness	Satisfaction is the “Degree to which a user is satisfied with their perceived achievement of pragmatic goals, including the results of use and the consequences of use”.

**RESEARCH ARTICLE - ENGINEERING TECHNOLOGY**

The description of the developed pragmatic criteria is given in Table 2.

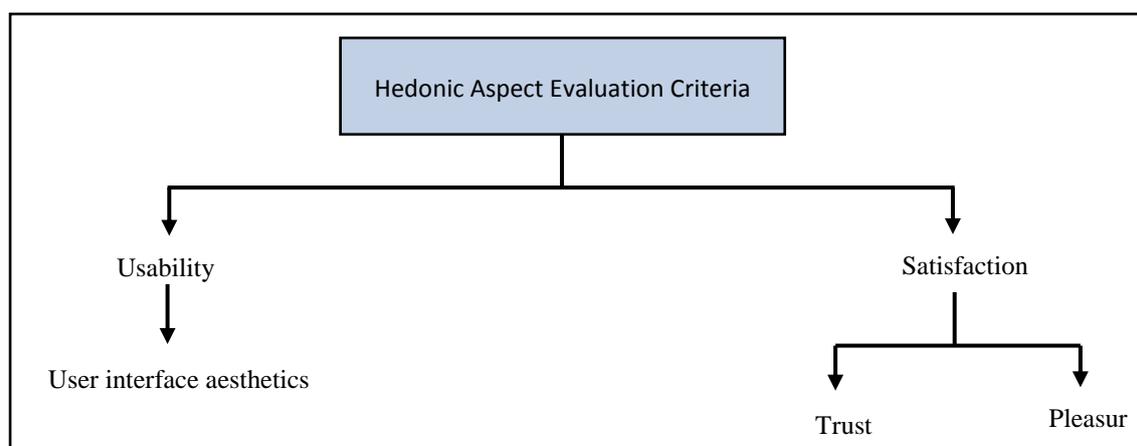
*Table 2: Description of pragmatic aspect evaluation criteria [8], [9]*

<b>ID</b>	<b>Yardstick</b>	<b>Description</b>
1	Functional coverage	Proportion of the specified functions has been implemented
2	Functional correctness	Proportion of functions provides the correct results
3	Functional appropriateness of usage objective	Proportion of the functions required by the user provides appropriate outcome to achieve a specific usage objective
4	Functional appropriateness of application	Proportion of the functions required by the users to achieve their objectives provides appropriate outcome
5	Description completeness	Proportion of usage scenarios is described in the application description or user documents
6	Demonstration coverage	Proportion of tasks has demonstration features for users to recognize the appropriateness
7	Entry point self-descriptiveness	Proportion of the commonly used landing pages on a website that explain the purpose of the website
8	User guidance completeness	Proportion of functions explained in sufficient detail in user documentation and/or help facility that enables users to apply the functions
9	Entry fields defaults	Proportion of entry fields that could have default values are automatically filled with default values
10	Error messages understandability	Proportion of error messages that state the reason why the error occurred and how to resolve it
11	Self-explanatory user interface	Proportion of information elements and steps presented to the user enable common tasks to be completed by a first-time user without prior study or training or seeking external assistance
12	Operational consistency	The extent to which the interactive tasks have a behavior and appearance that is consistent both within the task and across similar tasks
13	Message clarity	Proportion of messages from a system that convey the right outcome or instructions to the user
14	Functional customizability	Proportion of functions and operational procedures that a user can customize for his/her convenience
15	User Interface customizability	Proportion of user interface elements that can be customized in appearance
16	Monitoring capability	Proportion of function states that can be monitored during operation
17	Undo capability	Proportion of tasks that has a significant consequence provides an option for re-confirmation or undo capability
18	Understandable categorization of information	Proportion of software information that is organized in categories that are familiar to the intended users and convenient for their tasks
19	Appearance consistency	Proportion of user interfaces with similar items that have similar appearance
20	Input device support	The extent to which the tasks can be initiated by all appropriate input modalities (such as keyboard, mouse or voice)
21	Avoidance of user operation error	Portion of user actions and inputs that are protected against causing any system malfunction
22	User entry error correction	The extent to which the system provides suggested corrections for detected user entry errors with an identifiable cause
23	User error recoverability	Proportion of user errors that can be corrected or recovered by the system

24	Accessibility for users with disabilities	The extent to which the potential users with specific disabilities successfully use the system (with assistive technology if appropriate)
25	Supported languages adequacy	Proportion of supported languages
26	Satisfaction with features	User satisfaction of using specific system features
27	Discretionary usage	The proportion of potential users using a system or function
30	Feature utilization	The proportion of users using a particular feature
31	Proportion of users complaining	The proportion of users making complaints
32	Proportion of user complaints about a particular feature	The proportion of user’s complaints about a particular feature

(2) hedonic aspect

Hedonic aspect refers to “the system's perceived ability to support the user’s achievement of ‘be-goals’[19], [20], such as being happy, or satisfied with a focus on self”. The user needs experience hedonic aspect is divided into two general evaluation criteria each of which is divided into a set of specific evaluation criteria as shown in Figure 8. The evaluation part that corresponds to these criteria requires collecting users’ satisfaction using the questionnaire as a data gathering tool. Using SQuaRE terminology, these criteria are mainly evaluating quality in use.



*Figure 8: Hedonic aspect evaluation criteria tree*

Table 3 illustrates the specific definition of the evaluation criteria. The criteria tree and these definitions aid in the accurate understanding of the hedonic yardstick and this would help the evaluator know exactly what characteristics are to be analyzed.

*Table 3: General and specific evaluation criteria description of Hedonic aspect [10]*

General evaluation criteria	Specific evaluation criteria	Description
Usability	---	Usability is the “Extent to which a product or system can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use”.
	User interface aesthetics	“Degree to which the user interface enables pleasing and satisfying interaction for the user”.
Satisfaction	---	Satisfaction is the “Degree to which user needs are satisfied when a product or system is used in a specified context of use”.
	Trust	“Degree to which a user or other stakeholder has confidence that a product or system will behave as intended”.
	Pleasure	“Degree to which a user obtains pleasure from fulfilling their personal needs”.

The description of the developed hedonic evaluation criteria is given in Table 4.

*Table 4: Description of hedonic aspect evaluation criteria [8], [9]*

ID	Yardstick	Description
1	Appearance aesthetics of user interfaces	The extent to which the user interfaces and the overall design aesthetically pleasing in appearance
2	User trust	The extent to which the user trusts the system
3	User pleasure	The extent to which the user obtains pleasure compared to the average for this type of system

## 2.6. Yardstick.

The yardstick is “the description of the target and the criteria tree developed in the previous two steps are the basis for developing the yardstick. All yardsticks must contain the specifications, requirements, descriptions, or values for each criterion considered” [13]. In this research context, the yardstick is built based on pragmatic aspect evaluation criteria and hedonic aspect evaluation criteria developed in section 4.1.2. The yardstick, in this case, will consist of the different criteria that should be taken into consideration during the evaluating of user needs experience. Hence, the synthesis technique would verify criterion-by-criterion to ensure that each criterion has been considered in the evaluation. Furthermore, for the pragmatic criteria, a quantitative value has been assigned to each evaluation criteria based on measurement function given in SQuaRE standards, specifically in ISO/IEC 25022 and ISO/IEC 25023 [8], [9]. Which are used to measure the assigned yardstick for each evaluation criteria? The assigned quantitative value ranging from 0.0 to 1.0, the closer to 1.0 is better. A sample of the quantitative value assigned for the evaluation criteria is presented in Table 5.

Table 5: Quantitative Values of the Pragmatic Aspect Evaluation Criteria

General Criteria	Specific Criteria	Yardstick	Yardstick Values	Data gathering Techniques (DGT)
Functional suitability	Functional completeness	Functional coverage	$0 \leq X \leq 1$ The closer to 1.0 is the better.	DGT used is measurement. The functional coverage is measured as: $X = 1 - A / B$ A = Number of functions missing B = Number of functions specified Data can be collected from: <ul style="list-style-type: none"> <li>• Requirement specification document</li> <li>• Design specification document</li> <li>• User manual document</li> <li>• Test report</li> </ul>
	Functional correctness	Functional correctness	$0 \leq X \leq 1$ The closer to 1.0 is the better.	DGT used is measurement. The functional correctness is measured as: $X = 1 - A / B$ A = Number of functions that are incorrect B = Number of functions considered Data can be collected from: <ul style="list-style-type: none"> <li>• Requirement specification document.</li> <li>• Design specification document</li> <li>• User manual document</li> <li>• Test report</li> </ul>
Usability	Appropriateness recognisability	Description completeness	$0 \leq X \leq 1$ The closer to 1.0 is the better.	DGT used is measurement. The description completeness of system is measured as: $X = A / B$ A= Number of usage scenarios described in the application description or user documents B= Number of usage scenarios of the product. Data can be collected from: <ul style="list-style-type: none"> <li>• User manual document</li> <li>• application description</li> <li>• Operation (test) report</li> </ul>

For the hedonic criteria, a questionnaire has been developed as an evaluation tool to cover the various hedonic criteria. The questionnaire consists of four parts: usefulness part, pleasure part, user interface aesthetics part and trust part. Each part consists of a set of positive and negative statements designed to evaluate user satisfaction. The answers are based on a Likert scale and a weight has been assigned as follows: 1 for strongly disagree, 2 for disagree, 3 for neither agree nor disagree, 4 for agree and 5 for strongly agree. This proposed evaluation method is aligned with the evaluation framework discussed in [21] in the sense that we defined the UX evaluation criteria related to the product quality and quality in use based on the corresponding defined measures in the SQuaRE standard. A sample of the evaluation tool is presented in Table 6. The full version of the evaluation tool is available at <https://www.surveymonkey.com/r/CHMPTHM>

*Table 6: Sample of the evaluation tool*

<b>A. Usefulness Part:</b> This part includes a list of statements for evaluating user satisfaction with their perceived achievement of a pragmatic goal include the result of use and consequence of use. Please rate your agreement with these statements by using the following scales.					
Question	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
1. Overall, I am satisfied with how easy it is to use this application	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. It is simple to use this application	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. I can effectively complete my work using this application	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. I am able to complete my work quickly using this application	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. I am able to complete my work in the best possible manner with the least waste of time and effort using this application	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>B. Pleasure Part:</b> This part includes a list of statements for evaluating user pleasure from fulfilling personal needs. Please rate your agreement with these statements by using the following scales.					
1. I am interested in using this application	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. I feel distressed/sad while using this application	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. I always feel excited while using this application	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. This application makes me feel upset when using it	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**2.7. Data gathering techniques.**

Here, the potentially applicable data-gathering techniques should be identified. These techniques are used to obtain necessary information to judge the target. The main data-gathering techniques used in most evaluations in the software engineering field can be classed in three groups [13]: the measurement techniques, the assignation techniques, and Opinion techniques. In this research context, the measurement and assignation techniques are used. For each criterion, we assigned measurement function as measurement data-gathering technique, the measurement function used to combine the quality measure elements for each criterion to produce the quality measure (yardstick). Consequently, assignation data gathering technique is assigned to obtain data, which is used to generate the numerical values of quality measure elements, to judge the user needs experience (target) with the next component (synthesis techniques).

**2.8. Synthesis techniques.**

Synthesis techniques are used to synthesize all the data and information obtained after applying the data-gathering techniques and for comparison against the yardstick to judge the target and obtain the results of the evaluation [13]. Usually, two types of synthesis techniques are used, single value techniques and multiple values techniques [13]. The selection of the synthesis techniques depends on the preceding components. In

this research context, the multiple values technique is used where criteria grouping and datum-by-datum comparison with the yardstick is applied. Consequently, a set of recommendation obtain based on evaluation results helps to develop the evaluation target, hence the user needs experience evaluation methods.

## 2.9. Evaluation process.

The evaluation process is a series of activities and tasks that are executed to perform an evaluation. All the previous components are necessary to describe and design an evaluation method, but it is the evaluation process that describes the list of activities to perform and when to use the previous elements in practice. The evaluation process describes three main phases, the planning or preparation phase, examination phase and decision making phase[13], these phases match the three major points through which an evaluation passes. In this research context, the activities associated with each phase are shown in Figure 9. In the planning phase, the target should be analyzed first. This analysis is needed to get more information about the target in order to design the components in the next steps. In the last stage, all the activities and resources required for conducting the evaluation should be prepared. In the examination phase, the evaluator should apply the data gathering techniques to collect the data and verify the completeness of collected data. Finally, in the decision-making phase, the evaluator should apply the synthesis technique to compare the data collected from the preceding phase with the yardstick. This comparison would show the weakness points in the evaluated target and be able to suggest improvements in the final evaluation report. The evaluation process should be documented, this documentation would be useful for comparisons with the results obtaining in future evaluations of the same or similar targets” [19].

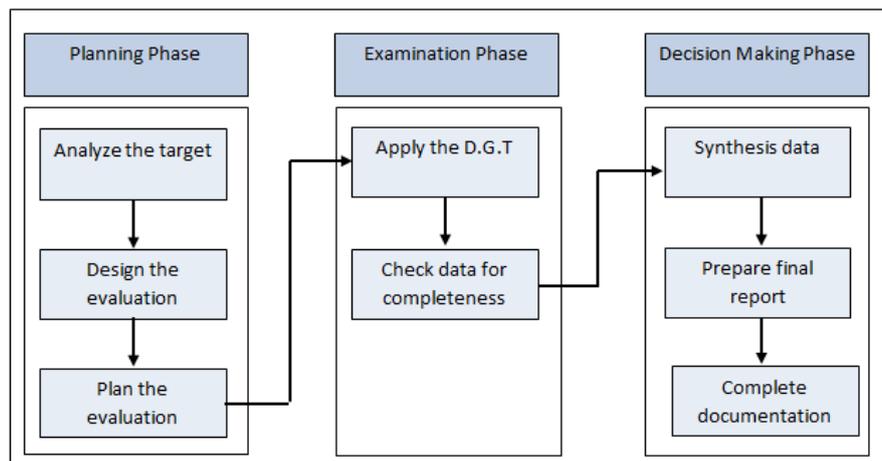


Figure 9: Main sub-processes of the proposed evaluation process

## 5. Quality of the developed UX evaluation method

Measuring UX needs using the developed evaluation method is a way to quantify the phenomenon under study which is the UX in our case. Such phenomenon is an abstract concept usually known as a theoretical construct which is available in various domains that include health and social sciences [22]. “Using tests or instruments that are valid and reliable to measure such constructs is a crucial component of research quality” [22].

In this research context, we used two verification approaches to validate the developed evaluation method as follows:

### 3.1. Content Validity

Content validity is concerned with determining how well the items developed to measure a concept of interest are adequate and representative of all the items that might measure that concept. Determining whether a measure or tool adequately covers a content area or adequately represents a concept is difficult to be quantified using statistical tests. Hence, content validity usually depends on the judgment of experts in the concept domain.

Accordingly, two user experience experts have been asked to review and rate the developed evaluation tool and answer a short survey of 17 questions about the clarity and suitability of the evaluation tool. The answer for the questions are based on a Likert scale of three scales (agree, partially agree and disagree). To measure the degree of concordance between the two raters, the inter-rater reliability test is calculated using Cohen’s kappa [23]. The calculated Cohen’s kappa is given in Table 7. It can be seen that the agreement level between the two raters (the kappa coefficient  $k$ ) is 73%. According to the kappa divisions defined in [24], the agreement level ranges between 0.61-.80 is considered to having a substantial level of agreement. This means that the two raters agree to an acceptable level on the suitability of the developed UX evaluation tool.

A more accurate vision of the suitability of the developed tool can be achieved if more experts rate the evaluation tool, but unfortunately, no other experts are found to agree on rating the evaluation tool. Maybe more experts should be contacted in the future for further improvements.

Table 7: Cohen’s Kappa Calculations

		Expert-2				
Exper-1	Answers	1	2	3	Total	%
	1	6	1	0	7	50%
	2	0	6	0	6	43%
	3	1	0	0	1	7%
	Total	7	7	0	14	
	%	0.5	0.5	0		
Probability Of agreement: P(a)						86%
Probability of agreement by Chance: P(e)						46%
Cohen’s Kappa (k)						73%

### 3.2. Internal analyses of the evaluation method

It is the process to explore the reliability and validity of developed evaluation method by evaluating the main internal disciplines, that used to develop the evaluation method. As we mentioned in section 2, we used three main disciplines to develop the proposed user experience needs evaluation method: the user experience framework as stated by [5], the evaluation theory concept [6], [25] and the SQuARE standards [8], [9], using those disciplines contributed to constructing strong bases to develop the evaluation method. Such strong bases lead to developing reliable and valid US evaluation method. The strength of the disciplines attributed to the following reasons:

- a. The user experience framework is a framework developed based on conducted a systematic literature review and analysis of extracted data from primary studies. Furthermore, the framework can be used as strategic guidelines for anyone interested in using the user experience activities in the organization.
- b. The evaluation theory is concept has been used in past decades and applied in various fields include the software engineering field. Many researchers in the software engineering field used the evaluation theory concept as bases to develop and evaluate different frameworks, methods and models for instance [12], [15], [26]. Furthermore, the evaluation theory can be applied in all kind of evaluation work.
- c. The SQuARE standards include set of international standards developed through technical committees established by the respective organization to deal with particular fields of technical activity. The

SQuaRE standards represent set of valid quality requirements, developed based on a set of quality characteristics and measures used to ensure high-quality software.

Moreover, to statistically judge the consistency of the evaluation items that constitute the evaluation method, we calculated the Cronbach's alpha coefficient after collecting answers of the participants in the case study discussed in section 6. The details of the Cronbach's alpha calculations and interpretations are discussed at the end of the next section.

## 6. Case study

A case study has been conducted based on the developed user experience needs evaluation method, to measure and evaluate user needs experience of a mobile app of one of the domestic telecommunications companies. The mobile application helps users to gain control of their telecommunication services account, which will allow them to view their bills and manage their services. The mobile application has been used by wide range of customers, and the company needs to measure and evaluate user experience for this mobile application, in order to improve the design and development of a mobile application, and support good user experience will enhance user satisfaction and loyalty. The evaluation process has been conducted in four phases:

### (1) determining the sample size

The sample size of any study is determined based on the population size, confidence level and margin of error. Population size is the total number of people in the group that we are trying to reach with our questionnaire, in this case, study the population is students and staff in the university ( $N \leq 4994$ ). The confidence level is 90% and the margin error is 6%, hence, the representative sample size is  $n = 184$ .

### (2) conducting the evaluation

The evaluation process has two main sub-tasks:

#### (a) answering the questionnaire

184 individuals participated in answering the questionnaire. The questionnaire was disseminated through the university email and Moodle website. The result of the questionnaire is shown in Table 8.

Table 8: The satisfied and dissatisfied percentages of user needs experience evaluation criteria

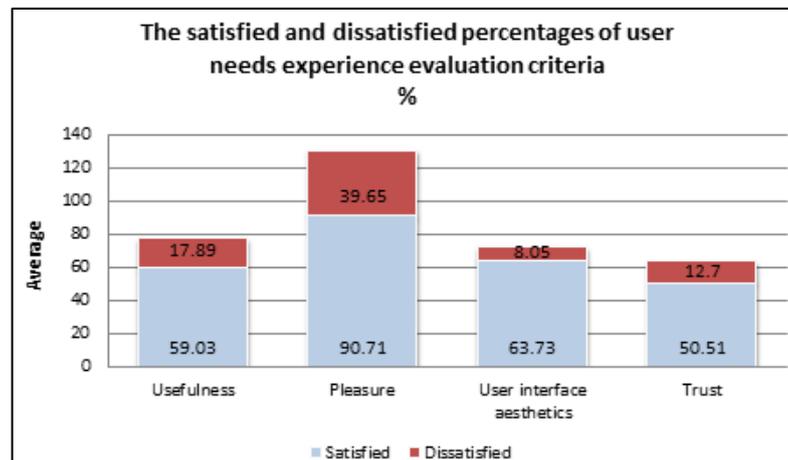
Criteria	Satisfied	Dissatisfied
Usefulness	59.03 %	17.89%
Pleasure	90.71%	39.65 %
User interface aesthetics	63.73%	8.05 %
Trust	50.51%	12.7%

#### (b) reviewing the documents

We asked the telecommunication company to allow us to review the technical documents related to the mobile app. we are evaluating such as the requirement specification document, design specification document, user manual document, test report, evaluation report, application description, user monitoring record, measure user behavior report, automated data collection report. Unfortunately, the company did not respond to our requests. This can be linked to cultural issues that need to be studied in a separate research. Hence, our case study focused on evaluating the user satisfaction only (via the survey).

### (3) analyzing the questionnaire results

The questionnaire results presented as a percentage. For each question, a weight has been assigned: 1 for strongly disagree, 2 for disagree, 3 for neither agree nor disagree, 4 for agree and 5 for strongly agree. The satisfied percentage is achieved by summing up the percentage of strongly agree and agree on answers for the positive sentences and by summing up the percentage of strongly disagree and disagree answers for the negative sentences. Similarly, the unsatisfied user percentages are achieved by summing up the number of strongly disagree and disagree answers for the positive sentences and the percentage of strongly agree and agree on answers for the negative sentences. The results of applying the questionnaire on the evaluation participants have been depicted in Figure 10.



**Figure 105: The satisfied and dissatisfied percentages of user needs experience evaluation criteria**

- (4) calculating the results of reviewing the documents (no done in this case study as we do not have access to the technical documents)

The result of reviewing technical documents is a set of measurement and technical data extracted based on answering the evaluation questions. The task of the evaluator is to calculate the quality measures by using the extracted measurement data and applying measurement functions based on the developed evaluation method.

- (5) applying synthesis technique to synthesize all the data and information obtained after applying the developed evaluation methods, the synthesis technique would verify criterion-by-criterion to ensure that each criterion has been considered in the evaluation, and for comparison against the ideal yardstick value to judge the target and obtain the results of the evaluation. This comparison represents the evaluation results, the comparison would show the weakness points in the evaluated target. Consequently, a set of recommendations obtains based on evaluation results helps to develop the evaluation target, hence the user experience needs evaluation methods.

- (6) the evaluation results

The results of comparison between the actual and ideal yardstick values illustrated that mobile application fulfilled partially the user needs experience criteria. These fulfilled needs are considered as strength points. These strengths points can be summarized as:

- a) Usefulness: 59.03% of participants are satisfied with mobile application features, and agreed that mobile application achieved their perceived pragmatic goals, including the results of use and the consequences of use.
  - The application helps the users to complete their tasks in the best possible manner with the least waste of time and effort.
  - The application is easy and simple to use, that's lead to help the users to learn how to use the application quickly.

- The application provides the users with clear, understandable and effective information help them to handle the application in a better way and finds the information they need easily.
- The application has all the functions and the capabilities that meet the users' expectations and satisfaction.
- b) Pleasure: 90.71% of participants are pleased, and agreed that mobile application achieved user pleasure from fulfilling personal needs.
  - The application designed in a way that does not make the user feel upset, nervous, guilty, embarrassed and scared when using it.
  - The application is user-friendly, the application designed in a way makes the user feel friendly when using it.
  - The application designed in a way makes the user interesting in using it.
- c) User interface aesthetics: 63.73% of participants are satisfied with the user interface of a mobile application, and agreed that the user interface of a mobile application and overall design aesthetically pleasing in appearance.
  - The application interface designed in a skillful way.
  - The application interface is pleasant.
  - The characters on interface written in clear and simple terms.
  - The application interface contains an appropriate amount of content organized and presented in a beautiful way.
  - The application interface has consistent elements and appropriate color, graphics, and images.
- d) Trust: 50.51% of participants are satisfied and trust mobile application, and agreed that mobile application is confidence and behave as intended.
  - The application achieved the user confidence.
  - The application has correct and predictable execution.
  - The application and its data are not corrupted.
  - The application performed the intended services without failure for a specified period of time.

The evaluation criteria that gained below 50% of the criteria are considered as weakness points. Based on the evaluation results, that mobile application has several weakness points. These weakness points can be summarized as:

- a) The application does not have all the functions and the capabilities that meet the users' expectations and satisfaction. More work is needed to collect and achieve user requirements.
- b) The application is not designed in a way that makes the user feel excited, strong, proud, alert, inspired and active when using it.
- c) The application does not always behave in an understanding manner, more work is needed to enhance understandability.

The Cronbach's alpha coefficients were calculated for the four subscales. The Cronbach's alpha for the four subscales is shown in Table 9. All the Cronbach's alpha values were well within the acceptable range [79]. This indicates the consistency and high interrelatedness of the questionnaire items.

**Table 9: Reliability statistics**

Scale	Responses	Cronbach's Alpha
Usefulness	184	0.995
Pleasure	138	0.994
User interface aesthetics	117	0.990
Trust	115	0.988

As a summary, the conducted case study showed that the developed user experience needs evaluation method 'Questionnaire' is able to measure and evaluate part of pragmatic aspect evaluation criteria and hedonic aspect evaluation criteria of user need experience.

## 7. Conclusion and future works

The purpose of this research is developing user experience needs evaluation method based on well-defined bases. Consequently, the research effort has been undertaken to develop user needs experience evaluation method based on three main disciplines the UX framework, the evaluation theory concept, and the SQuaRE standards. Various evaluation criteria have been defined in the context of user needs experience which is the result of the mapping between the user needs experience framework and the SQuaRE series of standards. The result of this research contributes to the evaluation of user needs experience of software applications. Supporting good user experience will enhance user satisfaction and loyalty. Hence, companies can use the evaluation method to guide their software development.

The conducted case study revealed the main limitation related the willingness and commitment of the mobile app. owners to participate and share technical documents necessary for the evaluation. Unfortunately, the company did not respond to our requests, to review a set of technical documents to extract data necessary to calculate the various measures. This can be linked to cultural issues that need to be studied in a separate research.

The future work that can be pursued can be summarized as follows:

- More experimental work on all the developed user experience needs evaluation method, to assure its applicability, this includes more inter-rater evaluations.
- Further improvements to the proposed user experience need evaluation method are possible as experience grows with use, the collaboration with more researchers and practitioners to evaluate the proposed evaluation method and get their feedbacks would help improve the proposed evaluation method.
- For each user experience evaluation method, a template and general guideline are required to assure that the questions and statements used in this method are covering all the related aspects.
- Apply the same methodology adopted in this research to develop user experience evaluation methods for the other user experience dimensions mentioned in the UX framework, such as brand experience dimension, technology experience dimension and context dimension.
- Encourage the software development companies to use the proposed user experience evaluation method, to enhance their user experience understanding and develop better applications that meet customers' expectations

## References

- [1] J. Tan, K. Rönkkö, and C. Gencel, "A framework for software usability & user experience measurement in the mobile industry," *Proc. - Jt. Conf. 23rd Int. Work. Softw. Meas. 8th Int. Conf. Softw. Process Prod. Meas. IWSM-MENSURA 2013*, pp. 156–164, 2013.
- [2] E. L.-C. Law, V. Roto, M. Hassenzahl, A. P. O. S. Vermeeren, and J. Kort, "Understanding, scoping and defining user experience," *Proc. 27th Int. Conf. Hum. factors Comput. Syst. - CHI 09*, no. April 2016, p. 719, 2009.
- [3] K. Väänänen-Vainio-Mattila, V. Roto, and M. Hassenzahl, "Towards practical user experience evaluation methods," *EL-C. Law, N. Bevan, G. Christou, M. Springett & M. L{á}rusd{ó}ttir Meaningful Meas. Valid Useful User Exp. Meas.*, no. October, pp. 19–22, 2008.
- [4] C. Ardito, M. Costabile, and R. Lanzilotti, "Towards the evaluation of UX," *Towar. a UX*, 2007.
- [5] M. S. Alharbi, "User Experience Dimensions, Aspects and Measures: Systematic Literature Review," Prince Sultan University, KSA, 2016.
- [6] L. Shadish, W., Cook, T., & Leviton, *Foundations of Program Evaluation. Theories of Practice*. Newbury Park, CA: Sage Publications, 1993.
- [7] M. Scriven, "Types of Evaluation and Types of Evaluator," *Am. J. Eval.*, vol. 17, no. 2, pp. 151–161, Jun. 1996.
- [8] ISO/IEC JTC 1/SC 7, "ISO/IEC 25022: Systems and software engineering -- Systems and software quality requirements and evaluation (SQuaRE) -- Measurement of quality in use," 2016.
- [9] ISO/IEC JTC 1/SC 7, "ISO/IEC 25023: Systems and software engineering -- Systems and software

- Quality Requirements and Evaluation (SQuaRE) -- Measurement of system and software product quality,” 2016.
- [10] ISO/IEC JTC 1/SC 7, “ISO/IEC 25010: Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models,” 2011.
- [11] M. Webster, *Merriam-Webster's collegiate dictionary*. Springfield, MA: Merriam-Webster, 2003.
- [12] A. Farooq, R. R. D. Prof, A. S. Prof, and F. Belli, “An Evaluation Framework for Software Test Processes,” no. September 2009.
- [13] M. Lopez, “An evaluation theory perspective of the Architecture Tradeoff Analysis Method (ATAM),” no. September 2000.
- [14] E. House, *Evaluating With Validity*. Sage Pubns, 1980.
- [15] M. Zarour, “Methods to evaluate lightweight software process assessment methods based on evaluation theory and engineering design principles,” 2009.
- [16] V. Roto, “User experience from product creation perspective,” *Towar. a UX Manif. Work.*, no. Roto, pp. 1–4, 2007.
- [17] A. Gross and S. Bongartz, “Why do I like it?: Investigating the product-specificity of user experience,” *Proc. 7th Nord. Conf. Comput. Interact. Nord.*, pp. 322–330, 2012.
- [18] N. Fronemann and M. Peissner, “User experience concept exploration,” *Proc. 8th Nord. Conf. Human-Computer Interact. Fun, Fast, Found. - Nord. '14*, pp. 727–736, 2014.
- [19] M. Hassenzahl, “User experience (UX): Towards an experiential perspective on product quality,” *Proc. 20th Int. Conf. Assoc. Francoph. d'Interaction Homme-Machine - IHM '08*, pp. 11–15, 2008.
- [20] P. Lew and L. Olsina, “Relating User Experience with MobileApp Quality Evaluation and Design,” pp. 253–268, 2013.
- [21] H. Nakai, N. Tsuda, K. Honda, H. Washizaki, and Y. Fukazawa, “A SQuaRE-based software quality evaluation framework and its case study,” in *2016 IEEE Region 10 Conference (TENCON)*, 2016, pp. 3704–3707.
- [22] C. L. Kimberlin and A. G. Winterstein, “Validity and reliability of measurement instruments used in research,” *Am. J. Heal. Pharm.*, vol. 65, no. 23, pp. 2276–2284, 2008.
- [23] J. Cohen, “A Coefficient of Agreement for Nominal Scales,” *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, Apr. 1960.
- [24] J. R. Landis and G. G. Koch, “The Measurement of Observer Agreement for Categorical Data,” *Biometrics*, vol. 33, no. 1, p. 159, Mar. 1977.
- [25] M. Scriven, *Evaluation Thesaurus*. 1981.
- [26] M. Qasaimeh and A. Abran, “Design and Assessment for Agile Auditing Model: The Case of ISO 9001 Traceability Requirements,” *Lect. Notes Softw. Eng.*, vol. 3, no. 3, pp. 189–194, 2015.