TELKOMNIKA Telecommunication, Computing, Electronics and Control
Vol. 18, No. 5, October 2020, pp. 2787~2799
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v18i5.16061

A rigorous user needs experience evaluation method based on software quality standards

Mohammad Zarour
College of Computer and Information Sciences, Prince Sultan University, Kingdom of Saudi Arabia

Article Info

Article history:

Received Mar 13, 2020
Revised Apr 8, 2020
Accepted May 1, 2020

Keywords:

Evaluation theory
SQuaRE standards
User experience framework
User needs experience
UX evaluation method

ABSTRACT

Our lives, nowadays, are digital. We, as humans, are using software applications in all our life aspects to meet our daily objectives and fulfill our needs. Software solutions that comprise mobile apps are widely spread, users can select from hundreds of available software solutions that fit their needs. Accordingly, user needs are becoming intricate and the software organizations are competing high to satisfy user requirements and the desires for better quality. This competition is not about satisfying the functional requirements but also satisfy user experience as well. Accordingly, studying, measuring, and improving user experience is crucial for the success of any software product. 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 ISOsoftware quality standards 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 method. Linking there three disciplines led to systematically identify the necessary evaluation criteria to evaluate user needs experience.

Corresponding Author:

Mohammad Zarour,
College of Computer and Information Sciences (CCIS),
Prince Sultan University,
Riyadh, Kingdom of Saudi Arabia.
Email: mzarour@psu.edu.sa

1. INTRODUCTION

Our lives, nowadays, are totally digital. We, as humans, are using software applications in all our life aspects to meet our daily objectives and fulfill our needs. Software solutions that comprise mobile apps are widely spread, users can select from hundreds of available software solutions that fit their needs. Accordingly, user needs are becoming intricate and the software organizations are competing high to satisfy user requirements and the desires for better quality. This competition is not about satisfying the functional requirements but also satisfy user experience as well [1].

Recently, the domain of user experience (UX) gained my focus from both academia and industry where academic environments are trying to better comprehend, define, and formulate the concept of user experience. This effort has shaped various definitions of user experience [2]. In conclusion, a consensus in the various researches and practitioners communities in that UX "is more than just a product's usefulness and usability" [3], UX is the result of interaction with software, system or service, affected by a set of aspects in a "dynamic, context-dependent, and subjective" manner [2]. The user experience "attempts to include subjective attributes like, for instance, aesthetics, emotional, and social aspect in a design space which has

previously concerned with ease of use" [4]. So, for software applications to remain competitive and attractive for users, their UX should be evaluated high compared to other competitors in the market. But the question is how to evaluate UX?

The literature documents some researches to evaluate UX, see for example [5, 6]. Unfortunately, the main flaw in such research work is that it is constructed without considering the evaluation principles. i.e., none of the current research has used the evaluation theory as bases to develop the evaluation method. Such practice will provide more rigor and formal evaluation method. From this perspective, this paper evaluates user needs experience via constructing an evaluation method using three main disciplines the UX framework, the evaluation theory principles, and the **S**oftware Quality Requirement dEtermination (SQuaRE) standard documented in ISO/IEC 25022 and ISO/IEC 25023.

They remain part of this paper is organized as follows. Section 2 discusses the adopted research method Section 3, discusses the development of the evaluation method that includes presenting the mapping process between user experience aspects of UX framework and quality attributes of ISO 25000 series SQuaRE standards as well as the application of the key concepts of evaluation to design and develop the proposed user experience evaluation methods. Section 4, discusses the quality of the developed UX evaluation method. Section 5, presents a case study that applies the developed evaluation method to evaluate user needs experience of a mobile application owned by one of the main telecommunication companies in the Kingdom. Section 6, presents the conclusion of our study and future works.

2. RESEARCH METHOD

This research paper focusses on developing a user experience evaluation method. The following are used as input to develop such an evaluation method: the UX framework developed in [7], the evaluation theory concept [8, 9], and the ISO/IEC 25022 and ISO/IEC 25023 [10, 11].

2.1. UX framework

The UX framework, proposed in [7], has specified four UX dimensions, namely, Value, technology experience (TX), brand experience (BX), and user needs experience (NX) dimensions. These dimensions form the central part of the framework as illustrated in Figure 1. The framework illustrates, as well, the relationship between the UX dimensions and the user experience aspects that have a direct or indirect impact on the user experience. This framework has defined seven categories of UX aspects and several generic methods that can be used to measure user experience aspects [7]. Our research work focuses on one dimension, the user needs experience dimension, to develop user experience needs evaluation method.

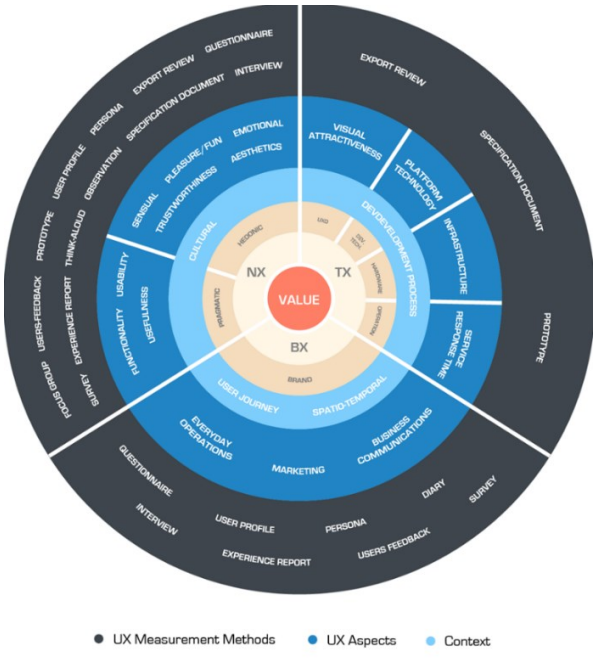

Figure 1. User experience framework [7]

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" [12].

The SQuaRE standard consists 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. ISO/IEC 25000 series is used in this paper to develop evaluation criteria for the proposed user needs experience evaluation methods as follows:

- ISO/IEC 25010 – system and software quality models

  This division defines both quality in use model (defines effectiveness, efficiency, satisfaction, freedom from risk and context coverage quality characteristics), and a product quality model that defines functional suitability, performance efficiency, compatibility, usability, reliability, security, maintainability, and portability quality characteristics.
- ISO/IEC 25022 – the measurement of quality in use

  This division provides a set of quality measures for measuring and evaluating quality in use [10].
- ISO/IEC 25023 – the measurement of system and software product quality

  This division provides a set of quality measures for specifying requirements, measuring, and evaluating the system/software product quality [11].

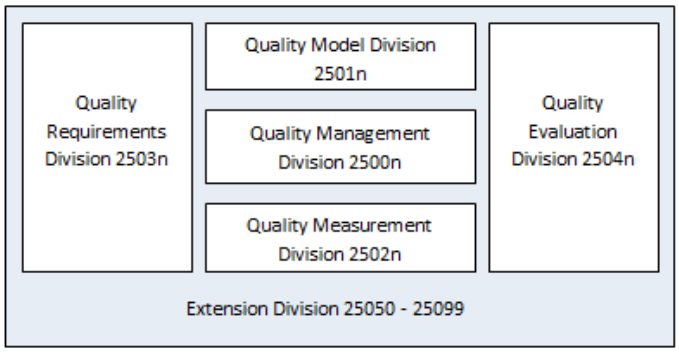

Figure 2. SQuaRE international standard's series [12]

2.3. Evaluation theory concept

Shadish, Cook, and Leviton [8] have reviewed the documented evaluation theories presented by seven well-known theorists and stated that "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". Scriven's theory of evaluation "attempts to clarify the logic behind evaluations" [9]. Figure 3 illustrates Scriven's six evaluation components which are discussed in detail as well in [13, 14].

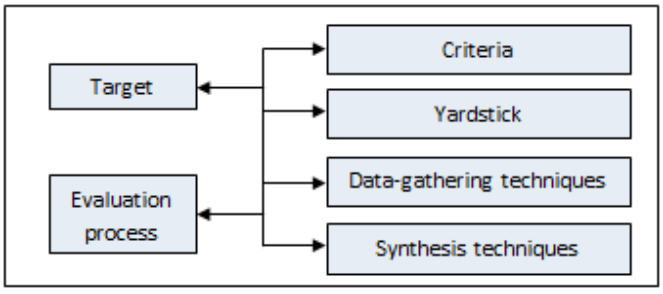

Figure 3. Components of an evaluation [9]

- The object of the evaluation is known as Target.
- The characteristics of the target are the criteria.
- Assessment techniques are the yardstick which is the standard or yardstick against which a real target is to be matched.
- Data-gathering techniques, these techniques should be defined and allocated to the corresponding evaluation criterion.
- Synthesis techniques, are used to judge the target and obtaining the results of the evaluation.
- The evaluation process, are the activities that should be executed to perform an evaluation.

3. DEVELOPING UX EVALUATION METHOD

At this point, and after explaining the different disciplines used to develop the UX evaluation method, we need to define the evaluation criteria tree to evaluate the user needs experience of the proposed evaluation method. The evaluation criteria are defined by mapping the quality factors defined as part of the user needs experience aspects presented in [7] to the software quality attributes of ISO 25000 (SQuaRE) standard [12]. Figure 4 demonstrates the mapping of user experience aspects with ISO 25000 characteristics and sub-characteristics.

The evaluation theory components apply to any evaluation [13]. These components are used in this research work to develop UX needs evaluation method. Interestingly, the mapping results depicted in Figure 4 will be used as the central evaluation criteria. The procedure to develop the proposed UX needs evaluation method is briefed in Figure 5. In this context, the "control-oriented method" of House classification [15] is used to ensure that the target is controlled by the yardstick specified [16]. The evaluators can develop user needs experience evaluation method by instantiating the evaluation framework depicted in Figure 3.

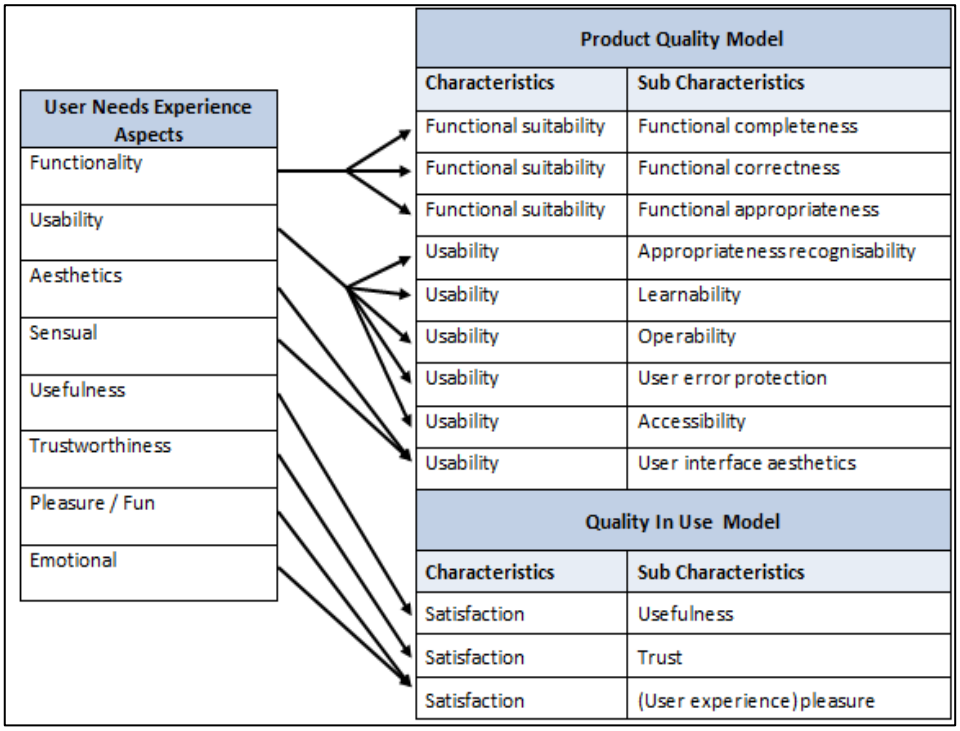

Figure 4. User expereince aspect and ISO 25000 standard mappings

3.1. Target

"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" [14]. In our case, the user needs experience is the target. The user needs experience focuses on both pragmatic and hedonic needs [17-19] that are the key to providing good systems that meet customers' needs and thereby contribute to the business success [17]. Figure 6 shows the components of the target evaluation criteria.

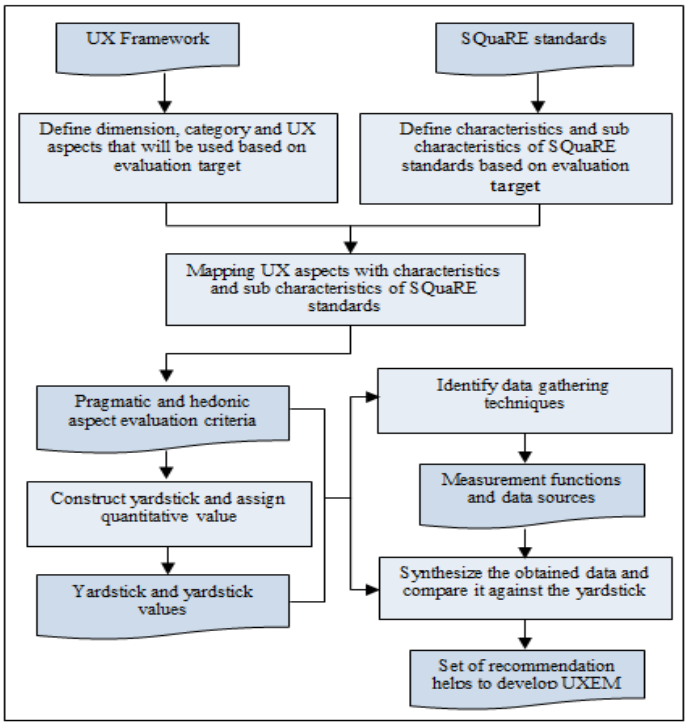

Figure 5. Develop user experience evaluation method

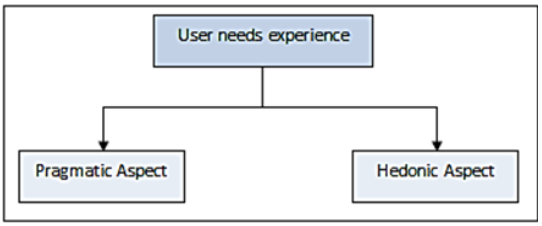

Figure 6. Components of target evaluation criteria

3.2. Criteria

After defining the target, we need to specify what are the target's characteristics that are of interest for the evaluation purposes. Such characteristics represent the evaluation criteria. The technique used for criteria elicitation is based on the needs' assessments elicitation method [14]. In this approach, the needs are analyzed and represented by a set of user needs aspects (pragmatic aspects, hedonic aspects) defined in the user experience framework [7]. To construct users' need experience criteria tree, we adopted the result of mapping user experience aspects with ISO25000 standard's characteristics.

In the end, the mapping process directed the construction of the pragmatic aspect evaluation criteria tree as shown in Figure 7, and hedonic aspect evaluation criteria tree as shown in Figure 8. This criteria tree is the basis for developing the evaluation yardstick.

a. 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" [20, 21]. These aspects can be measured using technical characteristics of the developed software which can be found in the technical reports. Using ISO 25000 terminology, these criteria are mainly evaluating software product quality related to user experience 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 are 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 [14]. Specific evaluation criteria refer to "characteristics that can be assigned a value directly using a particular data-gathering technique" [14].

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. The description of the developed pragmatic criteria is given in Table 2. These criteria are extracted from ISO25022 and ISO 25023 [10, 11] which refer to the measurement of quality in use, and system and software product quality respectively.

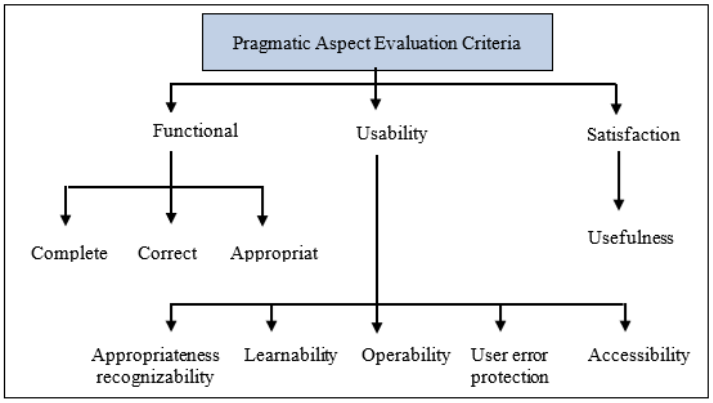

Figure 7. Pragmatic aspect evaluation criteria tree

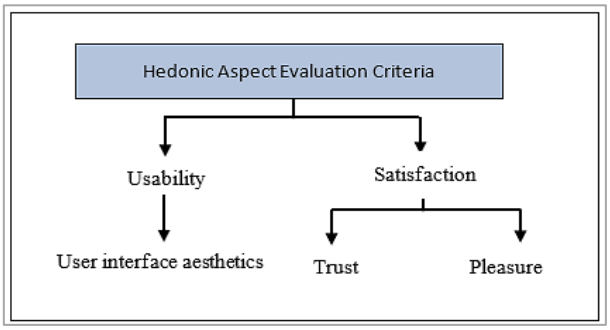

Figure 8. Hedonic aspect evaluation criteria tree

Table 1. General and specific evaluation criteria description of pragmatic aspect [12]

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	"The degree to which the set of functions covers all the specified tasks and user objectives".
	Functional correctness	"The degree to which a product or system provides the correct results with the needed degree of precision".
	Functional appropriateness	"The 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	"The degree to which users can recognize whether a product or system is appropriate for their needs".
	Learnability	"The 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	"The degree to which a product or system has attributes that make it easy to operate and control".
	User error protection	"The degree to which the system protects users against making errors".
	Accessibility	"The 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	"The 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".

Table 2. Description of pragmatic aspect evaluation criteria [10, 11]

ID	Yardstick	Description
1	Functional coverage	The proportion of the specified functions has been implemented
2	Functional correctness	The proportion of functions provides the correct results
3	Functional appropriateness of usage objective	The proportion of the functions required by the user provides an appropriate outcome to achieve a specific usage objective
4	Functional appropriateness of the application	The proportion of the functions required by the users to achieve their objectives provides appropriate outcome
5	Description completeness	The proportion of usage scenarios is described in the application description or user documents
6	Demonstration coverage	The proportion of tasks has demonstration features for users to recognize the appropriateness
7	Entry point self-descriptiveness	The proportion of the commonly used landing pages on a website that explain the purpose of the website
8	User guidance completeness	The 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	The proportion of entry fields that could have default values are automatically filled with default values
10	Error messages understandability	The proportion of error messages that state the reason why the error occurred and how to resolve it
11	Self-explanatory user interface	The 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 behavior and appearance that is consistent both within the task and across similar tasks
13	Message clarity	The proportion of messages from a system that convey the right outcome or instructions to the user
14	Functional customizability	The proportion of functions and operational procedures that a user can customize for his/her convenience
15	User Interface customizability	The proportion of user interface elements that can be customized in appearance
16	Monitoring capability	The proportion of function states that can be monitored during operation
17	Undo capability	The proportion of tasks that has a significant consequence provides an option for re-confirmation or undo capability
18	Understandable categorization of information	The proportion of software information that is organized in categories that are familiar to the intended users and convenient for their tasks
19	Appearance consistency	The proportion of user interfaces with similar items that have a 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	The 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	The 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	The 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	The proportion of users complaining	The proportion of users making complaints
32	The proportion of user complaints about a particular feature	The proportion of user's complaints about a particular feature

b. Hedonic aspect

The hedonic aspect refers to "the system's perceived ability to support the user's achievement of 'be-goals'[20, 21], such as being happy, or satisfied with a focus on the 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 ISO 25000 terminology, these criteria are mainly evaluating quality in use. Table 3 illustrates the specific definitions of the evaluation criteria. The criteria tree and these definitions would help in the understanding of the hedonic yardstick and this would help the evaluators know exactly what characteristics are to be analyzed.

3.3. 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" [14]. Both the pragmatic aspect evaluation criteria and

hedonic aspect evaluation criteria, developed in section 3.2, are used to develop the evaluation method in this paper. The description of the developed hedonic evaluation criteria is given in Table 4.

So, the synthesis technique will be used to verify criterion-by-criterion that each criterion has been considered in the evaluation. Furthermore, for the pragmatic criteria, a quantitative value has been assigned to each evaluation criterion based on measurement function given in SQuaRE standards, specifically in ISO/IEC 25022 and ISO/IEC 25023 [10, 11]. 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.

One the other hand, the hedonic criteria are evaluated via a questionnaire. The questionnaire consists of four sections: usefulness, pleasure, user interface aesthetics, and trust. Each section 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 [22] in the sense that we defined the user experience evaluation criteria related to the product quality and quality in use based on the corresponding defined measures of the ISO25000 series standard. The full version of the evaluation tool is available at https://www.surveymonkey.com/r/CHMPTHM.

Table 3. General and specific evaluation criteria description of Hedonic aspect [12]

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".

Table 4. Description of hedonic aspect evaluation criteria [10, 11]

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

3.4. Data gathering techniques

Data gathering techniques are used to obtain the 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 [14]: the measurement techniques, the assignation techniques, and Opinion techniques. In this paper, 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, the 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).

3.5. 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 [14]. Usually, two types of synthesis techniques are used, single value techniques, and multiple values techniques [14]. The selection of the synthesis techniques depends on the preceding components. In this research context, the multiple value technique is used where criteria grouping and datum-by-datum comparison with the yardstick is applied. Consequently, a set of recommendations obtain based on evaluation results helps to develop the evaluation target, hence the user needs experience evaluation methods.

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 <= X <= 1 The closer to 1.0 is better.	The functional coverage function X is: X = 1 – A / B A = Number of functions missed B = Number of functions specified Data can be collected from: • Requirement specification document • Design specification document • User manual document • Test report
	Functional correctness	Functional correctness	0 <= X <= 1 The closer to 1.0 is better.	The functional correctness function X is: X = 1 – A / B A = Number of incorrect functions B = Number of functions considered Data can be collected from: • Requirement specification document. • Design specification document • User manual document • Test report
Usability	Appropriateness recognizability	Description completeness	0 <= X <= 1 The closer to 1.0 is the better.	The description completeness of system function X is: 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: • User manual document • application description • Operation (test) report

3.6. 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[14], 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 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 weak 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" [20].

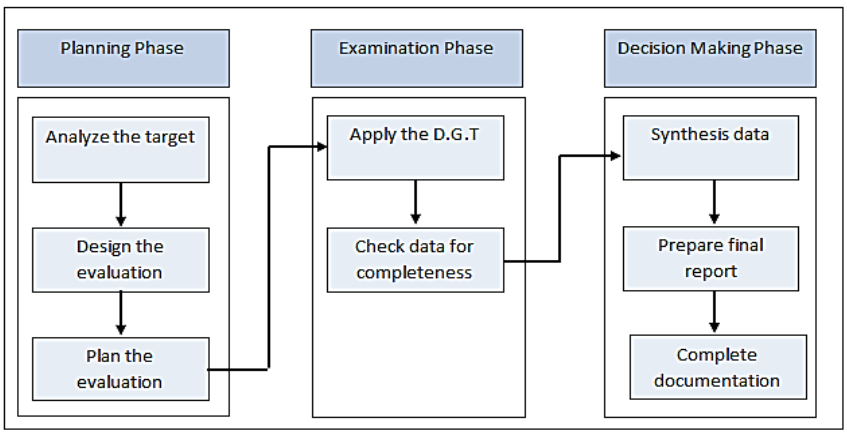

Figure 9. Main sub-processes of the proposed evaluation process

4. 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 a phenomenon is an abstract concept usually known as a theoretical construct that is available in various domains that include health and social sciences [23]. "Using tests or instruments that are valid and reliable to measure such constructs is a crucial component of research quality" [23]. In this research context, we used two verification approaches to validate the developed evaluation method as follows:

4.1. 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 answers to 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 [24]. The calculated Cohen's kappa is given in Table 6. 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 [25], the agreement level ranges between 0.61-.80 is considered as 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 6. Cohen's Kappa calculations

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

4.2. Evaluation method analysis

The evaluation method's reliability and validity are analyzed by evaluating the main internal disciplines, that used to develop the it. Three disciplines are used to develop the proposed evaluation method, namely, the user experience framework as described in [7], the evaluation theory concepts [8, 26] and the ISO25000 series standard [10, 11], The created evaluation method has made use of these disciplines as solid bases for its development. Such solid bases lead to developing reliable and valid user experience evaluation method. The solid nature of these disciplines is accredited to the following motives:

a. The UX framework is grounded on 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. Note that some new UX aspects are recently documented in the literature and apply to certain domains or exploring some new concepts that are beyond the UX domain, for instance, Shin et al. discussed algorithmic experience that comes beyond UX [27], Shin, as well, discussed the concept of immersion in augmented reality games and developed a model to predict UX of augmented reality games [28]. Such domain-specific aspects are excluded from this general UX evaluation method.
b. The evaluation theory has been adopted and applied in various arenas including 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, see for instance [13, 16, 29]. Furthermore, the evaluation theory can be applied to all kinds of evaluation work.

c. The ISO25000 series standard include a set of international standards developed through technical committees established by the respective organization to deal with particular fields of technical activity. The ISO25000 series standard represent a 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 5. The details of the Cronbach's alpha calculations and interpretations are discussed at the end of the next section.

5. CASE STUDY

A mobile app for a domestic telecommunication organization has been evaluated using the developed evaluation method. The app helps users to gain control of their accounts and supported services such as viewing and paying bills add and remove services, etc. The app is used by lots of customers, and the organization desires to measure and evaluate its mobile app. User satisfaction. By doing so, the organization will be able to improve the design of its mobile app as well as support better services. Hence, supporting good user experience will enhance their user satisfaction and loyalty.

The evaluation process includes two main phases or steps; the first one is to survey the users'rating of the various evaluation criteria. The other activity is to review the technical documents of the mobile app. Unfortunately, the organizations did not collaborate with us in having access to the needed resources in this regard and that forced us content with the first activity only. This revealed a gap in collaboration between academia and the businesses in the local market which has a cultural background and deserves more research.

Concerning the survey activity, students and staff in the university (N <= 4994) are used as a populace of this study. The confidence level is 90% and the margin error is 6%, so, the representative sample size is n = 184. The survey was distributed through the university email and Moodle website. The outcomes of the questionnaire are given in Table 7.

Table 7. User Needs Experience Evaluation Outcomes

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%

The synthesis technique is used to produce all the data obtained via the survey and to confirm criterion-by-criterion that each criterion has been measured during the assessment. The resuled data are compared against the ideal yardstick values to judge the target and obtain the results of the evaluation. The survey findings are shown in Figure 10.

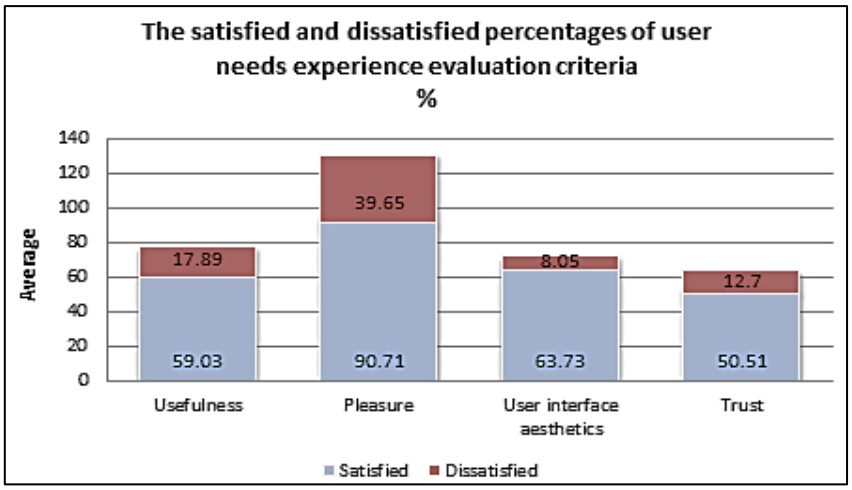

Figure 10. Percentages of the evaluation results

The actual and ideal yardstick values comparison illustrates that the mobile app fulfills partially the user needs experience criteria. These fulfilled needs are considered as strong points. These points can be summarized as:

- Usefulness: 59% of participants are satisfied with the mobile app features; the mobile app achieved its apparent pragmatic goals.
- Pleasure: 91% of participants are pleased, and agreed that the mobile app achieved user pleasure and fulfilled personal desires in this regard.
- UI aesthetics: 64% of participants are satisfied with the user interface of the mobile app; the mobile app interfaces' overall appearance is aesthetically pleasing.
- Trust: 51% of participants are satisfied and trust the mobile app; the mobile app is assured and behave as planned.

The evaluation criteria that gained below 50% of the criteria are considered as weak points. Hence, the mobile app has several weak points. These weak points can be summarized as:

- Not all the functions and capabilities that meet the users' expectations and satisfaction are present in the mobile app. Missed user requirements should be gathered and implemented properly.
- The mobile app does not make the user feels excited, inspired, and active when using it.
- The mobile app does not always behave in an understanding manner, more work is needed to enhance understandability.

We have calculated the Cronbach's alpha coefficients for the four subscales as shown in Table 8. The resulted Cronbach's alpha values were inside the acceptable ranges [29]. This indicates an acceptable level of reliability of the survey items. As a summary, the conducted case study showed that the developed user experience needs evaluation method 'Questionnaire' can measure and evaluate part of pragmatic aspect evaluation criteria and hedonic aspect evaluation criteria of user need experience.

Table 8. 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

6. CONCLUSION AND FUTURE WORKS

This paper has discussed the development of user experience needs evaluation method grounded on user experience framework that defined the UX main quality factors, ISO25000 series standard which is a well-defined standard that defines the software product characteristics and is used to map the UX quality factors to the well-defined software quality characteristics, and the evaluation theory that defines the main components of any evaluation method and is used as a guideline to develop the new evaluation method. The findings documented in this research paper subsidizes to the evaluation of user needs the experience of software applications. Such evaluation methods are essential to help organizations welling to evaluate their software applications to provide a better user experience that would boost user satisfaction and loyalty. So, organizations can use the evaluation method to guide their software development.

The developed evaluation method is used in a case study to evaluate one of the commonly used local mobile apps. The users' evaluation has been collected via a survey. The conducted case study was faced by some limitations mainly ignoring our requests made to the mobile app. owners to participate and share technical documents necessary for the evaluation. Unfortunately, the mobile app owner/organization did not reply to our appeals to analyze a set of technical documents to extract data necessary to calculate the various measures. This can be connected to cultural issues that need further research and investigation. The forthcoming research work will emphasize more on showing more experimental case studies on developed evaluation methods. Moreover, the same method implemented in this paper will be used to develop UX evaluation methods for the other dimensions reported in the UX framework, this includes brand experience, technology experience, and context dimensions.

REFERENCES

[1] J. Tan, K. Rönkkö, and C. Gencel, "A framework for software usability & user experience measurement in 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] Q. Yu, et al., "A Hybrid User Experience Evaluation Method for Mobile Games," *IEEE Access*, vol. 6, pp. 49067-49079, Jul. 2018.
[6] W. Nakamura, E. Oliveira, and T. Conte, "TUXEL: A Technique for User eXperience Evaluation in e-Learning," *Anais dos Workshops do VII Congresso Brasileiro de Informática na Educação (CBIE 2018)*, 2018.
[7] M. S. Alharbi, "User Experience Dimensions, Aspects and Measures: Systematic Literature Review," Prince Sultan University, KSA, 2016.
[8] L. Shadish, W., Cook, T.,& Leviton, "Foundations of Program Evaluation. Theories of Practice," *Newbury Park, CA: Sage Publications*, 1993.
[9] M. Scriven, "Types of Evaluation and Types of Evaluator," *Am. J. Eval.*, vol. 17, no. 2, pp. 151-161, Jun. 1996.
[10] 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.
[11] 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.
[12] 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.
[13] A. Farooq, R. R. D. Prof, A. S. Prof, and F. Belli, "An Evaluation Framework for Software Test Processes," Dissertation genehmigt durch die Fakultät für Informatikder Otto-von-Guericke-Universität Magdeburg, 2009.
[14] M. Lopez, "An evaluation theory perspective of the Architecture Tradeoff Analysis Method (ATAM) 's evaluation process and criteria, as well as its data-gathering and synthesis techniques, and more," *Software Engineering Institute,* 2000.
[15] E. House, "Evaluating With Validity," *Sage Pubns*, 1980.
[16] M. Zarour, "Methods to evaluate lightweight software process assessment methods based on evaluation theory and engineering design principles," Thesis for: PhD, Universite du Quebec, 2009.
[17] V. Roto, "User experience from product creation perspective," *Towar. a UX Manif. Work.*, 2007.
[18] 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.
[19] 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.
[20] 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.
[21] P. Lew and L. Olsina, "Relating User Experience with MobileApp Quality Evaluation and Design," *International Conference on Web Engineering,* pp. 253-268, 2013.
[22] H. Nakai, N. Tsuda, K. Honda, H. Washizaki, and Y. Fukazawa, "A SQuaRE-based software quality evaluation framework and its case study," *2016 IEEE Region 10 Conference (TENCON)*, pp. 3704-3707, 2016.
[23] 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.
[24] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37-46, 1960.
[25] J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, vol. 33, no. 1, pp. 159-74, Mar. 1977.
[26] M. Scriven, "Evaluation Thesaurus," *SAGE Publication,* 1981.
[27] D. Shin, B. Zhong, and F. A. Biocca, "Beyond user experience: What constitutes algorithmic experiences?," *Int. J. Inf. Manage.*, vol. 52, 2020.
[28] D. Shin, "How does immersion work in augmented reality games? A user-centric view of immersion and engagement," *Inf. Commun. Soc.*, vol. 22, no. 9, pp. 1212-1229, 2019.
[29] 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.