ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 9 (2022) No.1, pp.343-366
DOI:10.33168/LISS.2022.0121

Perceived Usability of Software Systems: A Framework-Driven Study

Abeer A. Al-Hassan [1], Latefah F. Alshammari [2], Bareeq A. AlGhannam[2], Haneen Alabdulrazzaq[2]

[1] College of Business Administration, Kuwait University, Kuwait
[2] Public Authority for Applied Education and Training, College of Business Studies, Kuwait

abeer.alhassan@ku.edu.kw

Abstract. The main objective of perceived usability studies is to develop better quality software that is both efficient and effective. Retrospective usability studies in the literature are rich in data that can be used by systems developers to achieve that purpose. However, some developers still fail to make use of users' past user experiences, and as a result some systems continue to include persistent flaws following updates. To address this problem, a conceptual framework of evaluation was developed from the usability evaluation literature. This research proposes a Perceived Usability Evaluation Framework to be consulted by evaluators during maintenance and system updates. To validate this framework, the researchers used evidence empirically collected from the Public Authority for Applied Education and Training (PAAET) online registration system. Results indicate that the framework provides a promising structure which can be followed by researchers, practitioners and systems developers when synthesizing patterns of dissatisfaction from previous system usability evaluations, and these syntheses can in turn can guide future system updates.

Keywords: Software usability, System Usability Evaluation Framework, perceived usability, standard usability, system evaluation

1. Introduction

During development, software systems undergo constant changes and updates. Most organizations implement systems to improve the quality, efficiency, and effectiveness of their software, and to reduce errors. In 2006, Hoegh, Nielsen, Overgaard, Pedersen & Stage recognized that feedback from usability studies significantly impacts developers' understanding of their systems, and found "results [which] indicate that observation of user tests facilitated a rich understanding of usability problems and created empathy with the users and their work " (Hoegh et al., 2006). Usability studies conducted over periods of time accumulate a wealth of data which developers can employ to successfully resolve problems in their systems. However, usability studies are not commonly employed by working organizations and systems developers, and this is evidenced by the vast failures of user satisfaction in many newly-updated versions of computer systems. This compelling lack of practice has recently been articulated in the literature: researchers have stated that the "knowledge and application of usability methods and practices have not yet found widespread adoption in practice (Feuersenger, Arndt & Riechers, 2020). As a result, these studies have only a minor impact on the development and updating of software systems.

Valid, reliable and easy-to-use usability tools are required to explore and address this issue. As such, and to assist in this area in an efficient and effective manner, this paper presents a conceptual framework derived from the usability evaluation literature. There is a serious need to learn from past usability studies, as patterns of satisfaction and/or dissatisfaction can imply the necessity of certain changes in the system. Furthermore, it is necessary to present usability processes in the form of easy-to-follow steps. The administration of usability evaluations requires awareness, guidance and appropriate tools. A convenient way to evaluate usability is by using standard evaluation methods. The use of standards in general is supported by the computing literature (Bevan, 2009; Quiñones & Rusu, 2017; Quiñones, Rusu & Rusu 2018). Standard questionnaires are one of the most common usability evaluation methods employewd in the literature (Lewis, 2002; Sauro & Lewis, 2012), and facilitate system comparison studies (Georgsson & Staggers, 2016). Lewis (1995) describes a method for the evaluation of subjective usability using Likert scale questionnaires. One such questionnaire is the Computer System Usability Questionnaire (CSUQ), which is used to evaluate perceived user satisfaction with computer systems (Lewis, 2018a; Lewis, 2018b).

This paper contributes to the research area of systems usability and explores the possibility of implementing and validating a framework that benefits from previous usability evaluations of a system; especially those collected over a period of time, from the perspective of the system's stakeholders, and by using standard evaluation tools. This research proposes a phase-wise framework for conducting usability studies, in the hope of regulating the practice of software usability

evaluations in an efficient and effective way.

To explore the validity of the proposed framework, CSUQ is employed and evidence is collected from the Public Authority for Applied Education and Training (PAAET) – an academic institution that acquired a student information registration system (Banner) in 2011/2012 which was updated in 2015/2016. CSUQ was administered to the stakeholders in 2011/2012 when the Banner system was first launched, and administered again in 2016/2017 after the system was updated. This kind of longitudinal study is important, but scarce in the literature and difficult to conduct (McLellan, Muddimer & Peres, 2012).

2. Literature Review

2.1. Academic Systems and Usability

Academic institutions undergo continuous changes which involve the facilitation of technology in the processes of their operations. Technology comes in many forms, from online services to mobile services. The aim is to enhance the ease-of-use, speed, efficiency and cost-effectiveness of these systems. This is accomplished by conducting continuous evaluations of the systems, which in turn contribute to the success of the software in usage terms (Lewis, 2018a; Moorthy, Ibrahim & Mahrin 2014).

2.2. Usability Evaluation

Software usability evaluations can be embedded as processes within or during the post- development phases of the software development life-cycle (Gediga, Hamborg & Düntsch, 1999), enabling developers to identify potential problems in the system. Moorthy et al. (2014) have developed a usability risk model that prioritizes usability risk assessment throughout the software development life-cycle, achieving improved systems. This provides feedback which enables better design; this feedback can be either summative or formative (Gediga et al., 1999). Furthermore, diverse methodologies and tools have been applied at various points during software development to evaluate usability (Bevan, Kirakowaski & Maissel, 1991; Scholtz, 2010; Sauro and Lewis, 2012; Moorthy et al., 2014; Rohrer, 2014; Hayat, Lock & Murry, 2015). Scholars have categorized usability evaluation in various ways, and it has been demonstrated over the years that usability evaluations follow a defined process (Norman, 1998; Assila, de Oliveira & Ezzedine, 2016).

One tool cannot be applied to all software usability studies for two main reasons. Firstly, the evaluation method must be inferred from the evaluation goal (Petrie & Bevan, 2009; Alghannam, Albustan, Al-Hassan & Albustan, 2017). Secondly, usability is constructed of a combination of attributes (Bevan et al., 1991) with multidimensional characteristics (Lewis, 1995). Different usability factors call for different measurement tools, which makes it challenging to decide which tool to use (Hornbæk, 2006).

Once a usability tool is chosen, it is important to note the environment of the evaluation, given that usability is the measurement of the unique interaction between product, user and circumstances (Bevan et al., 1991). Users' perception of the usability of a software system is influenced by their environment (Bevan et al., 1991; Hom, 1998) and context (Bevan et al., 1991; Bevan, 1995) and therefore cannot be measured separately from that environment (Petrie & Bevan, 2009; Bevan, Carter & Harcker, 2015).

2.3. Standard Usability Evaluation

Standards are widely used in the computing literature (Lewis, 1995; Bevan, 2009; Quiñones & Rusua, 2017; Quiñones et al., 2018). The sub-field of system usability also has its share of documented tools, processes and methodologies. Standardized tools provide reliable ways to measure users' perceptions of software usability (Petrie & Bevan, 2009). Lewis (1995) states that standardized measurement has advantages in terms of objectivity, quantification, communication, economy and scientific generalizability.

Several standardized usability tools take the form of questionnaires. Standard questionnaires can be used to quantify the subjective measurement of user experience (Sauro & Lewis, 2012), and practitioners use them to measure users' perceptions of satisfaction (Lewis, 1995). Standardized usability questionnaires can yield substantial results. One of the most widely used standard tools is the System Usability Scale (SUS) (Brooke, 1996; Brooke, 2013), which is used to evaluate the usability of various systems. The Computer System Usability Questionnaire (CSUQ) (Lewis, 1995) developed by IBM is another example of a standard usability questionnaire, in this case used specifically for computer systems. Other commonly used standard questionnaires include the Software Usability Measurement Inventory (SUMI) (Kirakowski & Corbett, 1993) and the Questionnaire for User Interaction Satisfaction (QUIS) (Chin, Diehl & Norman, 1988), which is a standard tool in the usability field.

Standardized questionnaires are psychometrically proven (Lewis, 1995). This means that they are: reliable, in that they can be measured consistently; valid, in that they measure the intended objective; and sensitive, as they are susceptible to low degrees of variance (Nunnally, 1978). Psychometric acceptance indicates that standard questionnaires can be used in confidence. Moreover, standardized questionnaires are well documented and thus easy to implement (AlGhannam, AlEssa & Almukhaizim, 2018), analyze, and compare. They provide a convenient way to acquire a better understanding of usability (Assila et al., 2016). They also make it easy to compare against benchmarks. Some scholars have called for setting benchmarks by promoting the use of standardized tools (AlGhannam, Alsuwaidi & Almayan, 2018). Once usability responses are collected using these tools, a framework that extrapolates crucial findings can then be used to guide practitioners and researchers through a series of easy-to-follow steps.

2.4. Usability Frameworks

A novel attempt to link a standard usability questionnaire with interactive image segmentation has been developed in Germany, in which a correlation between the System Usability Scale (SUS) and AttrakDiff-2 questionnaires were used and a framework for automation was proposed (Amrehn, Steidl, Kortekaas, Strumia, Weingarten, Kowarschik & Maier, 2019). Previous studies have also presented usability evaluation frameworks extracted from offline e-learning tutorials which cover "learning theories, evaluation, and practical aspects" (Nyang'or, DeVilliers & Ssemugabi, 2013). A recent study makes use of Nielsen's innovative Usability Testessen – a discount usability engineering approach in which evaluation is carried out by independent third parties (Feuersenger et al., 2020). An additional recent conceptual framework regulates the usability evaluation of disaster apps by analyzing online reviews (Tan, Prasanna, Stock, Doyle, Leonard. & Johnston, 2020). Apps for wearable devices are also supported by a usability framework which assists empirical assessment (Khakurel, Porras, Melkas, & Fu, 2020).

3. Methodology

3.1. Submitting (Use "Header 2" style)

The phase-wise framework for conducting effective and efficient perceived usability studies was derived from usability evaluation literature that is extensively linked with both information systems and evaluations (Raza, Siddiqui & Standing, 2019). The framework is comprised of several phases, as depicted in Figure 1.

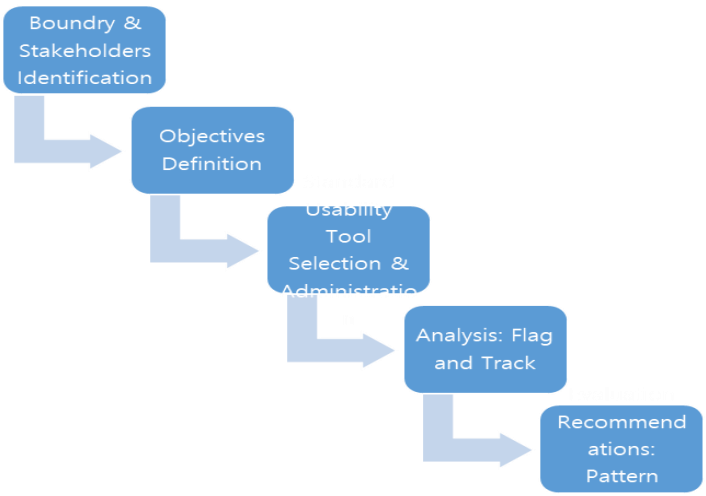

Figure 1: Perceived Usability Evaluation Framework

Phase 1: Boundary & Stakeholder Identification: In this phase, the boundary of the system under evaluation is identified, along with its prospective stakeholders.

Several tools can be utilized to identify the system's boundaries and stakeholders. The choice of tool can be left to the evaluator's discretion.

Phase 2: Definition of Objectives: The objectives of the usability evaluation are defined, which directs the evaluator to the appropriate standard usability tool. This phase is approached from a usability practice perspective, in which specific tools are used to evaluate specific objectives.

Phase 3: Standard Usability Tool Selection & Administration: The appropriate tool for meeting the defined objectives is chosen. These tools include CSUQ, SUS, UMUX, etc. Once a tool is selected, it is then administered to the studies which have been identified as within the boundary of the system, and then analyzed (Hartson, Andre & Williges, 2001; Lewis, 2018).

Phase 4: Analysis: Flag and Track: In this phase, previous concerns are flagged and tracked by stakeholders, and patterns are identified (Elmqvist & Yi, 2012).

Phase 5: Evaluation Recommendations & Pattern Encapsulation: This phase includes the encapsulation of patterns found in the selected studies, and the development of recommendations accordingly (SEBoK, 2021).

3.2. Validation Approach

In order to validate the proposed framework, the authors used PAAET's registration system as a case study. The population was a sample of students who had used the system in two separate academic years: 2011/2012, when the system was first launched; and 2016/2017, after the system was updated. Statistical analysis of the data was performed using SPSS version 22. Psychometric evaluation of the Arabic version of CSUQ is also presented. Reliability and validity measures were calculated to validate the Arabic version of CSUQ. Furthermore, a t-test was performed to measure the significance of any differences between the two samples. Finally, factor analysis was administered using one-way ANOVA, and questions 1-19 of the questionnaire were factored using the demographic item 'level of credits earned' as follows: items 1-8 to measure system usefulness; items 9-15 to measure information quality; and items 16-18 to measure interface quality. In addition, a comparison between the mean response scores of both samples was carried out. The authors also decoded the qualitative questions 20 and 21 of the questionnaire into tables and compared the answers to find any differences between them.

4. Framework Implementation/Application: Results, Analysis and Discussion

As previously stated, the objective of this research is to propose and implement a phase-wise framework for conducting an efficient and effective perceived usability study for any software system. In this section, the implementation of the Perceived Usability Evaluation Framework is demonstrated, following the phases described in Figure 1.

4.1. Phase 1: Boundary & Stakeholders Identification: Project Background (Participants and Setting)

The authors used the Soft Systems methodology CATWOE to identify the boundary of the system being studied (Checkland, 2000). The boundary was defined and limited to the current student registration system used by PAAET – Banner[1] – and its possible updates.

The IT department of PAAET were very helpful in guiding the authors when defining the boundary of the system. Figure 2 summarizes the data that was derived using CATWOE.

Customers	• Students
Actors	• Dean of Student Affirs, System Administrators, System Developers, IT Department, ,management
Transformation	• System and Features updates
World View	• The College wants to bullet proof the regisrtation process for better registration experience for students.
Owners	• PAAET and IT Department
Environmental	• System Scalability, Financial Constratint due to set governement funding

Figure 2 - CATWOE + Stakeholder analysis integrated.

4.2. Phase 2: Definition of Objectives

In phase 2, as previously mentioned, the goal is to set the proper path for evaluators to ask the proper questions to guide them in selecting the appropriate usability tool. In order to test the validity of this phase, the authors held discussions with the IT department and management to gain an understanding of their objectives. A summary of their objectives is listed below, in the form of the questions that the evaluators were interested in:

(1) Are students satisfied with the online registration system?

(2) To what degree do students find the system useful?

(3) To what extent are students satisfied with the information quality of the system?

[1]**Banner** by Ellucian is a Higher Education Enterprise Resource Planning Information System that integrates many add-on system components to perform various online tasks to serve students, academic staff and administrative staff. The Banner online registration system was first launched at PAAET colleges in the academic year 2011/2012.

(4) To what degree are students satisfied with the interface quality of the system?
(5) Are there any significant differences between students' perceptions in the academic years 2011/2012 and 2016/2017?
(6) Are there any repeated concerns from stakeholders from the past evaluation?

4.3. Phase 3: Usability Tool Selection & Administration

Phase 3 concerns the selection of the usability tool. As mentioned in the literature review, researchers and system evaluators have many different usability tools to choose from. Depending on the objective, some tools may be better suited than others. Once phase 2 is completed, the system usability objectives are defined, and specific questions are identified, finding the proper tool becomes an easier task.

CSUQ was chosen in this case because it covers three essential factors: (1) System Usefulness, (2) Information Quality, and (3) Interface Quality. These three factors are in line with the objectives defined in phase 2 of the framework. The authors chose CSUQ as the appropriate usability tool on the basis of the questions raised by the IT department. CSUQ is designed to investigate user satisfaction with a system. It is also one of the few questionnaires that allows for the collection of quantitative data in addition to qualitative data.

4.3.1 Computer System Usability Questionnaire (CSUQ)

CSUQ was used to collect data on the perceived usability of Banner, as described in Section 3. The tool is comprised of 21 questions and has both quantitative and qualitative elements. The quantitative element consists of 19 questions that respondents answer using a 7-point Likert scale which ranges from "strongly agree" (1) to "strongly disagree" (7) and includes an additional "not applicable - N/A" option. The last two questions – 20 and 21 – allow for the collection of qualitative user feedback on the positive and negative features of the software being assessed (Lewis, 1995).

Since the majority of PAAET's students lack a command of the English language and the curricula are taught mostly in Arabic, the English version of CSUQ had to be translated to Arabic. The translation was performed by the authors, who are bilingual and have a background in computer engineering and management information systems. The authors tried to use simple words in the translation to keep the statements as clear as possible. Table 1 depicts the original English questionnaire, along with its counterpart in Arabic.

Table 1. The Original English CSUQ* and the counterpart of each question in Arabic

The Original English Version		The translated Arabic Version	
1	Overall, I am satisfied with how easy it is to use this system	بصورة إجمالية، أنا راضي عن مدى سهولة استخدام النظام.	1
2	It was simple to use this system	كان استخدام النظام بسيطا.	2
3	I can effectively complete my	أستطيع القيام بجميع المهام التي أحتاجها من هذا	3

	work using this system	النظام بكفاءة.	
4	I am able to complete my work quickly using this system	لدي القدرة على القيام بجميع المهام التي أحتاجها من هذا النظام بسرعة.	4
5	I am able to efficiently complete my work using this system	لدي القدرة على القيام بجميع المهام التي أحتاجها من هذا النظام بفعالية.	5
6	I feel comfortable using this system	أشعر بالراحة من استخدام هذا النظام.	6
7	It was easy to learn to use this system	كان تعلم استخدام هذا النظام سهلا.	7
8	I believe I became productive quickly using this system	أنا متيقن من قدرتي على أن أنجز بسرعة باستخدام هذا النظام.	8
9	The system gives error messages that clearly tell me how to fix problems	النظام يعطي رسالة تبين الخطأ وتشرح بوضوح كيفية إصلاح هذا الخطأ.	9
10	Whenever I make a mistake using the system, I recover easily	عندما أخطئ باستخدام النظام، أستطيع تجاوز هذا الـ بسرعة وسهولة.	10
11	The information (such as online help, on-screen messages, and other documentation) provided with this system is clear	المعلومات المزودة مع النظام (مثل المساعدة أثناء الـ بالمهام المطلوبة،الرسائل على الشاشة وغيرها) واضحة.	11
12	It is easy to find the information I needed	من السهل الحصول على المعلومة التي أحتاجها.	12
13	The information provided for the system is easy to understand	المعلومات المزودة بالنظام كانت سهلة الفهم.	13
14	The information is effective in helping me complete the tasks and scenarios	المعلومات كانت فعالة في مساعدتي لإنهاء المهام المطلوبة.	14
15	The organization of information on the system screens is clear	تنظيم عرض المعلومات في شاشات النظام كان واضحا.	15
16	The interface of this system is pleasant	شاشات النظام تبعث في النفس السعادة.	16
17	I like using the interface of this system	أحب استخدام شاشات هذا النظام.	17
18	This system has all the functions and capabilities I expect it to have	أتوقع أن يحتوي النظام على جميع الوظائف والقدرات المطلوبة منه.	18
19	Overall, I am satisfied with this system	راضي عن أداء النظام إجمالا.	19
20	Please list the three things you liked most about this system software.	الرجاء أذكر أكثر ثلاثة أمور تحبها في هذا النظام.	20
21	Please list the three things you liked least about this system software.	الرجاء أذكر أقل ثلاثة أمور تحبها في هذا النظام.	21
*Source Lewis 1995			

4.3.2 CSUQ Validation

For validation, the authors calculated the Cronbach alpha to measure reliability and the Pearson coefficient to measure validity.

4.4. Phase 4: Results and Analysis (Flag and Track): Quantitative Results and Analysis

4.4.1 CSUQ Validation

A total of 122 students participated in the research in the academic year 2011/2012, and 115 participated in the research during the academic year 2016/2017.

The distribution of students from the sample, based on credits earned, is shown in Figure 3.

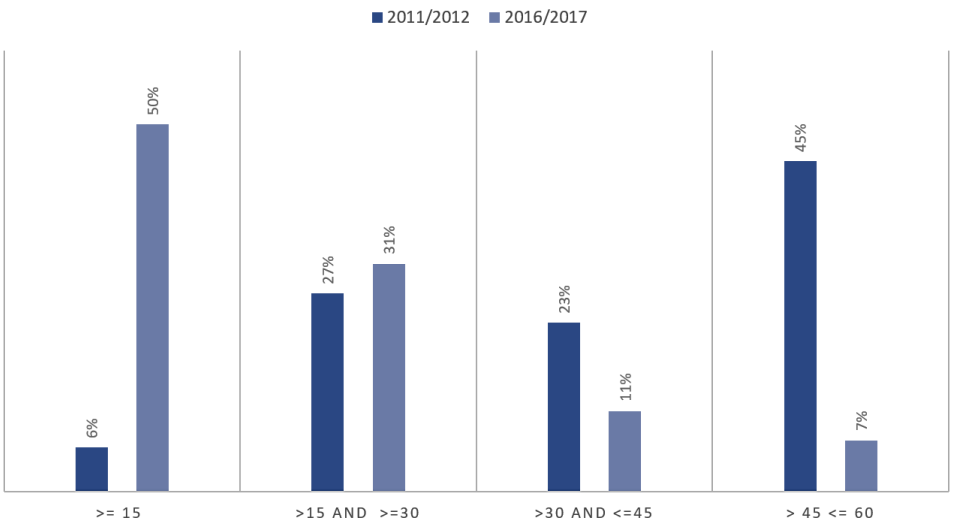

Figure 3. Percentages of level of student (credit earned)

4.4.2 Psychometric Analysis and Mean Distributions

Table 2 shows a comparison of CSUQ statistical results from 2011/2012 and from 2016/2017. The reliability of the questionnaire, as measured by Cronbach's alpha coefficient, was estimated as 0.96 and 0.94 for the academic years 2011/2012 and 2016/2017, respectively. These results indicate that the reliability of the Arabic-language CSUQ research tool is within the acceptable range, i.e. Cronbach's alpha coefficient was greater than 0.70 in each case. Nunally stated in 1978 "that the recommended minimum sample size for psychometric analysis (specifically for factor analysis) is at least five participants for each item". For the 19-item CSUQ, the minimum sample size would therefore be 95.

Table 2. Comparison of 2011/2012 CSUQ and 2016/2017 CSUQ statistic results

	2011/2012	2016/2017

Sample no.	122	115
Q no.	19	19
Reliability	0.96	0.94
Validity	.585-.855	0.545 – 0.786
Mean	4.31	4.36
Std	2.150	1.755

The t-test resulted in a p-value of 0.42, demonstrating that there were no statistically significant differences between the data for the years 2011/2012 and 2016/2017. In the academic years 2011/2012 and 2016/2017 the overall means of the questionnaire responses were 4.31 and 4.36, respectively. Figures 4 and 5 show the sorted mean values for the responses to each of the 19 items in CSUQ for the years 2011/2012 and 2016/2017, respectively. Figure 6 compares each question, and flags those which reflected low levels of satisfaction with the system. Although there was a five year gap between the two samples, the researchers did verify that the system was updated once during that period. The two evaluations were conducted basically on the same system, Banner; however, changes in the degree of users' maturity and their familiarity and confidence in using technology and online systems between 2011 and 2016 should be taken into consideration.

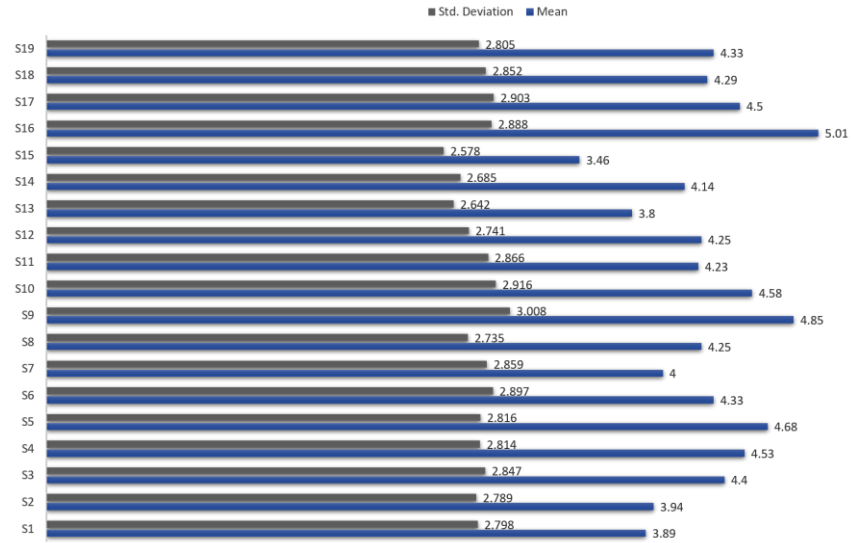

Figure 4. CSUQ mean and standard deviations in 2011/2012

4.4.3 Flag and Track

In an attempt to flag and track student satisfaction, the authors first looked at the p-value derived by the t-test to see if there was a significant difference between the two samples. Since the difference was not significant, the authors took a closer look at (1) the differences of the means of each question sorted by preferences (Figures 4 and 5) and (2) a comparison of the mean value of responses to each question (Figure 6) in order to flag any concerns.

4.4.3.1 Sorted Mean Comparison

Although the average mean values of responses from both samples were close (4.31 and 4.36) and the t-test showed no significant differences between them, the authors took a closer look at the means of the sorted data. The results of this analysis (Figures 4 and 5) show that the most agreed-upon items in the academic year 2011/2012 were 16, 9 and 5, and the least agreed-upon were items 15, 13 and 1. On the other hand, for the academic year 2016/2017, the authors found that the most agreed-upon items had changed to items 15, 14 and 11, and the least agreed-upon items to 16, 9 and 17.

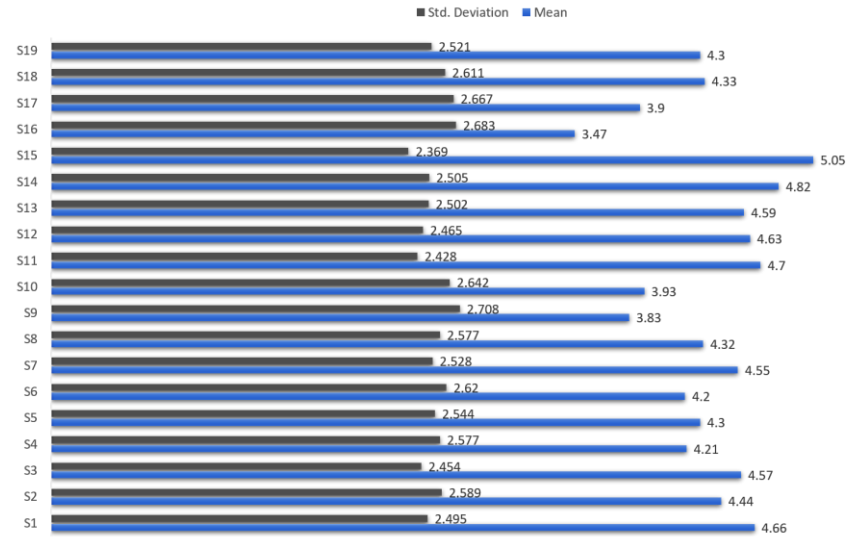

Figure 5. CSUQ mean and standard deviations in 2016/2017

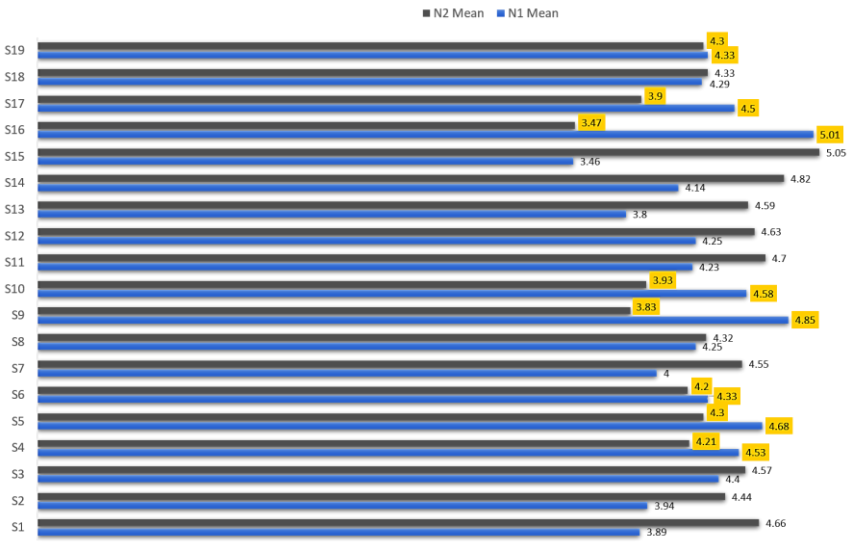

Figure 6. Comparison of Mean Values for 2011/2012 (N1) and 2016/2017 (N2)

This shift in the dominating constructs may be related either to a change in the system or to a change in student culture. When the issue was discussed with the Director of the Computer Department at PAAET, it was indicated that the system had not undergone major changes. This could mean that student culture changed in the five-year period between the samples. With the internet, online systems and social media dominating the scene, students are now more accustomed to handling services online, and are more experienced with software systems in general.

A notable observation that can be made from the data is that item 16 (*"The interface of this system is pleasant")* was the most agreed-upon statement in 2011/2012 but was the least agreed-upon statement in 2016/2017. This can be attributed to the fact that in 2011/2012 students had not been as exposed to information systems and the use of technology in general as their younger counterparts, five years later. Furthermore, the students of 2016/2017 could have been more demanding by nature, as they were exposed to a greater number of different interfaces. This may be due to the increase in smartphone ownership, engagement with many online interfaces to buy products or stream movies, and the increase in use of social media. The 2011/2012 batch were not as exposed to technology as the 2016/2017 batch was, so their point of reference was completely different.

Another finding was that item 9 (*"The system gives error messages that clearly tell me how to fix problems")* also shifted from being the most agreed-upon statement in 2011/2012 to the least agreed-upon in 2016/2017. This could be attributed to the fact that the older batch naturally needed assistance, so following prompts was important for them; whereas the newer generation reacted differently, expecting quick responses, more user-friendly systems and better-quality information.

Item 15 (*"The organization of information on the system screens is clear*") was one of the least agreed-upon statements in 2011/2012 but one of the most agreed-upon in 2016/2017. Again, this could be due to the fact that the later sample had more experience in using information systems and was skilled in finding elements on a webpage. The 2011/2012 student sample had not been exposed to technology to the same extent as the 2016/2017 sample, so for them navigating a system might be frustrating or simply challenging.

For the 2011/2012 sample, item 5 (*"I am able to efficiently complete my work using this system"* was among the most-agreed upon statements – this indicates that the students were satisfied in moving to an online system – particularly given that this was their first exposure to an online system in this context – which relieved them from the need to drive to college and wait in long lines to register. Among the statements which this batch agreed with the least were items 1 and 13 – *"Overall, I am satisfied with how easy it is to use this system"* and *"The information provided for the system is easy to understand"*, respectively. This could be due to this group's limited or non-existent exposure to software systems in general. It was their first encounter: they had no prior training in the system. It would be natural for them to experience some difficulty when using the system or in understanding some of the features that the system provides.

The 2016/2017 batch also agreed to a great extent with items 14 and 11, which were, respectively, *"The information is effective in helping me complete the tasks and scenarios"* and *"The information (such as online help, on-screen messages, and other documentation) provided with this system is clear".* It seems that the 2016/2017 students were very comfortable with interacting with the system and viewed the system as effective and helpful. On the other hand, this sample did not highly agree with the statement in item 17 – *"I like using the interface of this system".* This could be due to the younger generation's exposure to other information systems, which may have led them to have high expectations for information systems.

Both samples had equal measures of agreement on the statements in items 3 (*"I can effectively complete my work using this system")*, 6 (*"I feel comfortable using this system")*, 8 (*"I believe I became productive quickly using this system")* and 19 (*"Overall, I am satisfied with this system")*.

4.4.3.2 Comparison of Means for each Question

In order to facilitate flagging and tracking, the authors viewed the data from different angles. Figure 6 compares the means of responses to each item. At first glance, the average means of both samples might lead one to think that there was no relative difference between sample 1 (4.31) and sample 2 (4.36). As the researchers took a closer look, they found that out of the 19 items, 10 items indicated a small increase in satisfaction, with most of these increases falling within less than 1 point on the Likert scale. There was one exception: responses to *"The organization of the information on the system screen is clear*" increased from 3.04 in 2011/2012 to 5.05 in 2016/2017. The remaining eight items indicated that the 2016/2017 batch was less satisfied with the system than the 2011/2012 batch. These items flag concerns, and system developers need to investigate the reasons for this decrease in satisfaction. These concerns were expressed by low ratings for statements such as "*I am able to complete my work using this system"*, "*I am able to efficiently complete work quickly using this system" ,"I feel comfortable using this system"*, *"It was easy to lean to use this system",* and "|*I believe I became productive quickly using this system*". The three items in question demanded more investigation so that developers could be certain of the issues which required attention. Were these issues technical, and related to the system, or were they cultural? The 2011/2012 sample included eight students who had earned less than 15 credits – indicating that this was their first semester – whereas the 2016/2017 sample contained 58 students who had earned less than 15 credits. One might conclude that the freshman group of the students may not have been familiar with the registration process, and thus took a long time to complete it.

Items 11-15 all indicated that the later batch of students was less satisfied with the system than the earlier batch. These items were all related to the organization, display and location of information, and to on-screen messages. These concerns should all be considered, as the 2016/2017 batch were by nature more experienced in using computers and software – these concerns should be addressed in the next update. System developers need to administer a usability study prior to the next update and also continue to flag and track prior concerns if they are persistent.

Low rates of agreement with item 18 (*"This system has all the functions and capabilities I expect it to have"*) were also a concern, and developers need to investigate which functions were lacking and what the next generation of students is expecting.

4.4.4 Factor Analysis

Factor analysis was applied to questionnaire results from both samples. Questionnaire items 1-19 were factored using a one-way ANOVA test according to students' earned credits, as follows: items 1-8 measured system usefulness; items 9-15 measured information quality; and items 16-18 measured interface quality. Tables 3 and 4 show the means and significance levels of the results.

Table 3. One-way ANOVA according to credits for the academic year 2011/2012

		N	Mean	Sig
System Usefulness	less than 15	8	2.78	0.31
	from 16 -30	33	3.56	
	from 31 - 45	53	3.99	
	more than 45	28	4.35	
	Total	122	3.88	
Information Quality	less than 15	8	2.61	0.06
	from 16 -30	33	3.76	
	from 31 - 45	53	4.50	
	more than 45	28	4.32	
	Total	122	4.13	
Interface Quality	less than 15	8	3.35	0.20
	from 16 -30	33	3.81	
	from 31 - 45	53	4.20	
	more than 45	28	4.85	
	Total	122	4.19	
Total	less than 15	8	2.89	0.16
	from 16 -30	33	3.71	
	from 31 - 45	53	4.23	
	more than 45	28	4.49	
	Total	122	4.06	

Table 4. One-way ANOVA According to for the academic year 2016/2017

		N	Mean	Sig
System Usefulness	less than 15	58	2.30	0.43
	from 16-30	36	2.33	
	from 31-45	13	2.23	

	more than 46	8	2.21	
	Total	**115**	**2.29**	
Information Quality	less than 15	58	4.55	0.08
	from 16-30	36	4.07	
	from 31-45	13	4.23	
	more than 46	8	5.80	
	Total	**115**	**4.45**	
Interface Quality	less than 15	58	4.35	0.34
	from 16-30	36	3.71	
	from 31-45	13	3.37	
	more than 46	8	3.81	
	Total	**115**	**4.00**	
Total	less than 15	58	4.51	0.17
	from 16-30	36	4.00	
	from 31-45	13	4.05	
	more than 46	8	5.38	
	Total	**115**	**4.36**	

Table 3 shows that, on a scale of 1 (strongly disagree) to 7 (strongly agree), average overall satisfaction with the system as measured by CSUQ was 4.06, indicating that users (N = 122) were somewhat satisfied with the system in 2011/2012. Mean CSUQ factor scores were: System Usefulness (M = 3.88), Information Quality (M = 4.13), and Interface Quality (M = 4.19). For the academic year 2016/2017, Table 4 shows that users (N=115) were also somewhat satisfied with the system, with an overall CSUQ mean of 4.36. The mean CSUQ factor scores for 2016/2017 were as follows: System Usefulness (M = 2.29), a low score which indicates that the students did not find the system very useful, Information Quality (M = 4.45) and Interface Quality (M = 4.0). Both samples gave System Usefulness the lowest mean score. Overall, there was no significant difference between the responses of the two samples. The final two items of CSUQ (items 20 and 21) are of a qualitative nature. They are both open ended questions that ask the users to list three aspects that they like about the system and three that they dislike. The results and analysis of these questions can be found in the Appendix.

4.5. Phase 5: Evaluation Recommendations: Pattern Encapsulation

4.5.1 Discussion

Both sets of qualitative data were consistent with each other: there were no significant differences between the 2011/2012 and 2016/2017 data. No-one chose “strongly agree” or “agree” on the Likert scale – most of the data fell into the “somewhat agree” or “somewhat disagree” categories, with a mean of 4.31 on the scale for the 2011/2012 sample, and a mean of 4.36 for the 2016/2017 sample.

The result of a t-test comparison ($p = 0.42$) of the 2011/2012 data and the 2016/2017 data demonstrated that there was no significant difference between the means of all scores (4.31 and 4.36, respectively), which may indicate that students were satisfied with the system to a certain degree. A closer investigation of item scores found consistency between the samples when comparing the mean distributions and factor analyses of both batches. The standard deviation values for each batch (2.150 for 2011/2012 and 1.755 for 2016/2017) were relatively close, and this indicates that the collected responses from the students were distributed in an acceptable manner.

However, after taking a closer look at the mean scores, we were able to flag eight areas in which students in the later batch were less satisfied with the system than students in the earlier batch. System developers should examine these areas. After updating a system, one would hope that user satisfaction would increase. One factor which may have reduced the satisfaction rate is the students' relative level of experience of using IT, which was higher for the 2016/2017 sample.

In general, one hopes that updating a system will increase user satisfaction. In this case it did not do so; this is clearly an issue which needs to be flagged, and systems developers should look into all the reported concerns and decide whether they represent systems issues or issues that should be raised with the college administration, such as the possible introduction of an orientation program.

Analysis of the qualitative data also indicated that students reported problems such as "the system is complicated" and "the system hangs", and that they demanded additional functions such as "registration of field training". Flagging these issues will allow the system developers to tackle them in a new upgrade. The new upgrade could simplify the steps involved in registration, since both batches felt that the process was too complicated. However, problems with the system hanging may represent a technical issue of scalability, and this should also be addressed. In addition to the these two points, system developers might also need to communicate the students' concerns about the unavailability of a function to allow registration for field training to the Dean of Student Affairs, and find out whether this function was left out for a reason or whether they should they consider adding such a feature to the system.

5. Conclusion and Recommendations

This research contributes to the field of systems usability by emphasizing the importance of conducting regular usability evaluations of perceived satisfaction using a phase-wise framework. This Usability Evaluation Framework used in this research consists of five phases: Phase 1 (Boundary & Stakeholder Identification); Phase 2 (Definition of Objectives); Phase 3 (Standard Usability Tool Selection & Administration); Phase 4 (Analysis: Flag and Track); and finally Phase 5 (Evaluation, Recommendations & Pattern Encapsulation).

The authors feel that the results indicate that this framework provides a promising structure to be followed by researchers, practitioners and systems developers when synthesizing patterns of dissatisfaction from previous usability evaluations of a system, and that this can, in turn, guide future system updates. It would also be interesting to substitute machine learning tools for the statistical analysis tools and compare the results with those obtained in this research.

The authors applied each phase of the model proposed for validation to the case of the Banner registration system used by PAAET. Phase 1 of the framework emphasizes the inclusion of the stakeholders and the identification of the scope and boundary of the system. In this research, the authors used CATWOE to identify the boundary and stakeholders of the system that the update was targeting.

The framework also emphasizes that the choice of a tool should be based on the objective of the updates. This will guide system developers to choose the appropriate tool for their purpose. In Phase 2, the IT department shared their objectives, and questions were drafted to enable the developers to choose the appropriate tool to help achieve those objectives.

Phase 3 was concerned with tool selection. Once the authors gathered the required data, an appropriate tool for answering the questions raised by the IT department was selected. In this case, the authors chose to make use of CSUQ. CSUQ consists of a set of quantitative questions – which are analyzed using the distribution of response means as well as by factor analysis – and a set of qualitative open-ended questions. Both sets of questions allowed the authors to investigate the satisfaction or dissatisfaction of students with PAAET's online registration system over two academic years, 2011/2012 and 2016/2017.

Once the required data was available, analysis began and phase 4 was in progress. After thorough analysis, the authors were able to conclude that this phase is highly critical for any system update. It was also recognized that it is very important to track data over time: this allows system evaluators and system developers to flag repeated concerns and determine the cause of any dissatisfaction with a system. Analysis of the quality of the data for each sample showed that the students in each sample were consistent in terms of both qualitative and quantitative data. The data from both samples indicate that there is significant room for improvement in the system, particularly in the area of the quality of the user interface and the usefulness of the information provided by the system. The data also indicate that in some areas students simply lacked knowledge of how to use the system.

Once the system developers flagged repeated concerns, they were able to determine which concerns could be addressed in the new update. The evaluation also flagged some technical issues, as well as some concerns related to administrative policy. Problems that are system-related may be addressed by the developers, technical issues may be flagged to the IT department, and other policy-

related issues may be brought to the attention of management. These policy-related issues include the possibility of providing training in the system to users.

Further research may be carried out using a different method to gather data, such as qualitative focus groups. A questionnaire followed by focus groups might provide a more didactic tool that could result in a better understanding of usability. In addition, it may be beneficial to explore other usability tools, such as SUS or UMUX, and compare their results with those of CSUQ.

The framework was tested on only one case study. It would be interesting to apply the Perceived Usability Evaluation Framework to different systems and determine whether this systematic approach is useful for systems developers. Another limitation of this study is that the frequency of usability studies on systems depends on their environments – for example, on the frequency of updates to the system.

6. Appendix: Qualitative Results

Items 20 and 21 of the CSUQ were open-ended questions that facilitated the collection of qualitative data. Item 20 asked the participants to list the three things they liked most about the system, and item 21 asked the participants to list the three things they liked least about the system. Of the 122 respondents in 2011/2012, only 25 participated in the open-ended part of the CSUQ, whereas 46 of the 115 total respondents in 2016/2017 participated in the open-ended part of the CSUQ and shared their opinion.

The authors manually clustered the qualitative data into categories or themes that were repeated by the students. The categories were as follows: ease of use, system usefulness, information quality, interface quality and technical issues. The authors omitted many comments by students that were irrelevant to the study. Some examples of omitted data were comments such as “in early registration, we are only allowed to register for 12 credits”, “not enough sections offered”, or “faculty monopoly of courses”. Tables 5 and 6 reflect the aggregated student remarks for the years 2011/2012 and 2016/2017, respectively.

Table 5. Result of participants for the academic year 2011/2012

Themes	Liked (Satisfied)	Disliked (Unsatisfied)
Ease of Use	Login Process (4)	Complicated (11)
System Usefulness	Easier Process (25) Fast to Register (5)	
Information Quality	Shows GPA (3) Shows Academic Records (8) Displays Closed Sections (5)	No feature to register for field training (1)

Interface Quality	Clear (15)	Not Clear (2)
Technical		Slow (11) System Hangs (8)

Note: Numbers in parenthesis represent the number of respondents

Table 6. Result of participants for the academic year 2016/2017

Themes	Liked (Satisfied)	Disliked (Unsatisfied)
Ease of Use	Easy to Register (32)	Complicated (8)
System Usefulness	Register anywhere anytime (5)	
Information Quality	Help textbox (9) Shows Academic Records (7) Displays Closed Sections (5)	
Interface Quality	Clear (5)	Colors (3) Layout (1) Fonts are very small (1)
Technical	Register using Mobile (4)	Slow (36) System Hangs (20)

Note: Numbers in parenthesis represent the number of respondents

The data shown in Table 5 indicates that all 25 of the students in the 2011/2012 sample who responded to the open-ended section felt that the online registration process was easier than the offline process, and five felt that fast registration was one of the things that they liked most in the system. Eleven students, however, found the Banner system complicated, and two commented that PAAET should go back to the offline system. Fifteen students thought the interface was clear. No student expressed any dissatisfaction with the interface. The students were satisfied with the features that the system offered, which included the display of academic records, grade point averages, and closed sections.

The data shown in Table 6 indicates that 32 of the students in the 2016/2017 sample felt that the system was easy to use, and only eight found it complicated. Students were happy that they were able to register from anywhere by using the online system. As was the case in the 2011/2012 sample, the 2016/2017 students were happy that they could view academic records and that the system displayed closed sections. Five students agreed with the 2011/2012 batch that the interface was clear, but also expressed their dissatisfaction with the color, layout and font size of the interface.

There were two common remarks on the technology that are worth noting. Firstly, 11 participants in 2011/2012 and 36 in 2016/2017 expressed their dissatisfaction with the slow speed of the network. Secondly, eight students in 2011/2012 and 20 students in 2016/2017 said that the system “hangs”.

Acknowledgements

We would like to thank Mr. Siaq Natlus for his review and edit of the document. Also, we would like to thank the Public Authority for Applied Education and Training (PAAET), College of Business Studies faculty, who administered the questionnaire in their classes.

References

AlGhannam, B., Albustan, S., Al-Hassan, A. & Albustan, L. (2017), Towards a Standard Arabic System Usability Scale (A-SUS): Psychometric Evaluation Using Communication Disorder App. *International Journal of Human-Computer Interaction*, 34(9), 799-804. Available: https://doi.org/10.1080/10447318.2017.1388099, [Accessed 30 December 2018].

AlGhannam, B., AlEssa, E. & Almukhaizim, M. (2018), Mobile First Companies: A Case of App Usability in Kuwait. *International Journal of Computer Science and Information Security*, 16(3), 105-117.

AlGhannam, B., Alsuwaidi, M. & Almayan, W. (2018), Perceived Usability using Arabic System Usability Scale (A-SUS): Student Perspective of Smart PAAET App. *International Journal of Computer Science and Information Security*, 16(7), 66-78.

Amrehn, M., Steidl, S., Kortekaas, R., Strumia, M., Weingarten, M., Kowarschik, M. & Maier, A. (2019), A Semi-Automated Usability Evaluation Framework for Interactive Image Segmentation Systems. *International Journal of Biomedical Imaging*, 2019, 1-22. 1464592

Assila, A., de Oliveira, K. M. & Ezzedine, H. (2016), Standardized Usability Questionnaires: Features and Quality Focus, Electronic *Journal of Computer Science and Information Technology*, 6(1), 15-30.

Bevan, N., Kirakowaski, J. & Maissel, J. (September, 1991), What is Usability? *Proceedings of the 4th international conference on HCI*, Stuttgart, Germany.

Bevan, N. (1995), Usability is quality of use, *Proceedings of The 6th international conference on Human Computer Interaction*, Yokohama. Anzai and Ogawa (eds), Elsevier.

Bevan, N. (2009), International standards for usability should be more widely used. *Journal of Usability Studie*s, 4(3), 106-113.

Bevan, N.,, Carter J., & Harker S. (2015), "ISO 9241-11 Revised: What Have we Learnt About Usability Since 1998?" In 17th International Conference, HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part I, edited by Masaaki Kurosu, 143–151. Springer. DOI: 10.1007/978-3-319-20901-2_13

Brooke, J. (1996), SUS: A 'Quick and Dirty' Usability Scale. In P.W. Jordan, B. Thomas, B.A. Weerdmeester and I.L. McClelland (Eds.), *Usability Evaluation in Industry*, 189-194. London: Taylor and Francis.

Brooke, J. (2013), SUS: A Retrospective, *Journal of Usability Studies*, 8(2), 29-40.

Checkland P. (2000), Soft Systems Methodology: A Thirty Year Retrospective, *Systems Research and Behavioral Science*, 17, S11-S58.

Chin, J.P., Diehl, V. & Norman, K. (1988), Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface, Proceedings of *the SIGCHI conference on Human factors in computing System*, 213-218.

Feuersenger, H., Arndt, H. & Riechers, K. (2020), *The International Conference on Human-Computer Interaction*, 97-112.

Gediga, G., Hamborg, K. & Düntsch, I. (1999), The IsoMetrics Usability Inventory: An Operationalization of ISO 9241-10 Supporting Summative and Formative Evaluation of Software Systems. *Behavior and Information Technology*, 18(3), 151- 164.

Georgsson, M. & Staggers, N. (2016), Quantifying Usability: an Evaluation of a Diabetes mHealth System on Effectiveness, Efficiency, and Satisfaction Metrics with Associated User Characteristics. *The Journal of the American Medical Informatics Association*, 23(1), 5–11.

Hartson, H., Andre T. & Williges, R. (2001), Criteria for Evaluating Usability Evaluation Methods, *International Journal of Human-Computer Interaction*, 13(4), 373-410. DOI: 10.1207/S15327590IJHC1304_03

Hayat, H., Lock, R. & Murray, I. (2015), Measuring Software Usability. In Proceedings of *Software Quality Management Conference.* BCS. Retrieved from: https://dspace.lboro.ac.uk/2134/18275. Accessed [1 September 2016].

Hoegh, R., Nielsen, C., Overgaard, M., Pedersen, M. & Stage, J. (2006), The Impact of Usability Reports and User Test Observations on Developers' Understanding of Usability Data: An Exploratory Study, *International Journal of Human–Computer Interaction*, 21(2), 173-196, DOI: 10.1207/s15327590ijhc2102_4

Hom, J. (1998), The Usability Methods Toolbox h
Handbook. Retrieved from http://usability.jameshom.com/. Accessed [Jan 30th 2020]

Hornbæk, K. (2006), Current Practice in Measuring Usability: Challenges to Usability Studies and Research. *International Journal of Human-Computer Studies*, 64(2), 79-102. Retrieved from https://doi.org/10.1016/j.ijhcs.2005.06.002. Accessed [25 November 2017]

Khakurel, J., Porras, J., Melkas, H. & Fu, B. (2020), A Comprehensive Framework of Usability Issues Related to the Wearable Devices. *In: Paiva S., Paul S. (eds) Convergence of ICT and Smart Devices for Emerging Applications. EAI/Springer Innovations in Communication and Computing.* Springer, Cham. https://doi.org/10.1007/978-3-030-41368-2_2

Kirakowski, J. & Corbett, M. (1993), SUMI: the Software Usability Measurement Inventory. *British Journal of Educational Technology*, 24(3), 210-212. Available: doi:10.1111/j.1467-8535.1993.tb00076.x.

Lewis, J. (1995), IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction*, 7(1), 57-78.

Lewis, J. (2002), Psychometric Evaluation of PSSUQ Using Data from Five Years of Usability Studies, *International Journal of Human-Computer Interaction*, 14(3&4), 463-488.

Lewis, J. (2018a). The System Usability Scale: Past, Present, and Future. *International Journal of Human–Computer Interaction*, 35(7), 577-590. doi:10.1080/10447318.2018.1455307.

Lewis, J. (2018b), Measuring Perceived Usability: The CSUQ, SUS, and UMUX. *International Journal of Human-Computer Interaction*, 34(12), 1148-1156.

McLellan, S., Muddimer, A. & Peres, S. (2012), The effect of experience on system usability scale ratings, *Journal of Usability Studies*, 7(2), 56-67.

Moorthy, J., Ibrahim, S. & Mahrin, M. (2014), Developing Usable Software Product Using Usability Risk Assessment Model. *International Journal of Digital Information and Wireless Communications (IJDIWC)*, 4(1), 95-102. The Society of Digital Information and Wireless Communications, (ISSN: 2225-658X).

Elmqvist N. & Yi J. S. (2012), Patterns for Visualization Evaluation. Proceedings of the 2012 BELIV Workshop: Beyond Time and Errors - Novel Evaluation Methods for Visualization. Association for Computing Machinery, New York, NY, USA.

Norman, D. (1998). *The invisible computer*. Cambridge, Mass: MIT Press.

Nunnally, J. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.

Nyang'or, J., DeVilliers, R. & Ssemugabi, S. (2013), A Framework for Usability Evaluation of an Offline e-learning Tutorial and its Application in Usability Testing. *World Conference on Educational Multimedia, Hypermedia and Telecommunications (EDMEDIA) 2013*.Victoria, Canada.

Petrie, H. & Bevan, N. (2009), *The evaluation of accessibility, usability and user experience. The Universal Access Handbook*. C Stepanidis (ed), CRC Press. Retrieved from, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.461.1537&rep=rep1&type=pdf. Accessed [28 August, 2018].

Quiñones, D. & Rusu, C. (2017), How to Develop Usability Heuristics: A Systematic Literature Review. *Computer Standards and Interfaces*, 53(C), 89-122.
Quiñones, D., Rusu, C. & Rusu, V. (2018), A methodology to develop usability/user experience heuristics. *Computer Standards and Interfaces*, 59, 109-129. https://doi.org/10.1016/j.csi.2018.03.002

Raza, S., Siddiqui, A. & Standing, c. (2019), Exploring Systemic Problems in IS Adoption Using Critical Systems Heuristics, *Systemic Practice and Action Research*, 32, 125-153, https://doi.org/10.1007/s11213-018-9467-6

Rohrer, C. (2014), When to Use Which User-Experience Research Methods. Nielson Norman Group. Retrieved from https://www.nngroup.com/articles/which-ux-research-methods/. Accessed [30 March 2017].

Sauro, J. & Lewis, J. (2012), Quantifying the User Experience: practical statistics for user research. Waltham, MA: Morgan Kaufmann.

Scholtz, J. (2010). Usability Evaluation. National Institutes of Standards and Technology. Retrieved from:
https://pdfs.semanticscholar.org/8dec/cec5ace9235878e6aab06c3cd54f7b33a2ce.pdf
Accessed [Jan. 30th 2020].

SEBoK Editorial Board. 2021, The Guide to the Systems Engineering Body of Knowledge (SEBoK), v. 2.4, R.J. Cloutier (Editor in Chief). Hoboken, NJ: The Trustees of the Stevens Institute of Technology. Accessed Nov. 01 2020]. www.sebokwiki.org.

Tan, M., Prasanna, R., Stock, K., Doyle, E., Leonard, G. & Johnston, D. (2020), Modified Usability Framework for Disaster Apps: A Qualitative Thematic Analysis of User Reviews. *International Journal of Disaster Risk Science*, available at: www.ijdrs.com, https://doi.org/10.1007/s13753-020-00282,