The Method and Metric of User Experience Evaluation: A Systematic Literature Review

Aulia Inan Nur
Faculty of Computer Science,
Universitas Indonesia
aulia.inan@ui.ac.id

Harry B. Santoso
Faculty of Computer Science,
Universitas Indonesia
harrybs@cs.ui.ac.id

Panca O. Hadi Putra
Faculty of Computer Science,
Universitas Indonesia
hadiputra@cs.ui.ac.id

ABSTRACT

With the growth of User Experience (UX) research field, researchers have developed various ways to implement UX evaluation method. These evaluation methods have different practice, evaluated application, and type of collected aspect. This article provides a systematic literature review on research papers from 2000 to 2019 related to UX evaluation, to better understand UX evaluation method and its implementation, what kind of application its applied to, and what type of collected metric. The result of this paper presents the most frequently used method is self-reported measurement that evaluates self-reported metric and issue-based metric and the least frequently used method is physiological measurement that evaluates emotion and stress metric.

CCS CONCEPTS

• **General and reference** → Document types; General literature;
• **Human-centered computing** → Human computer interaction (HCI); HCI design and evaluation methods.

KEYWORDS

User Experience, Measurement, Methods, Systematic Literature Review

ACM Reference Format:
Aulia Inan Nur, Harry B. Santoso, and Panca O. Hadi Putra. 2021. The Method and Metric of User Experience Evaluation: A Systematic Literature Review. In *2021 10th International Conference on Software and Computer Applications (ICSCA 2021), February 23–26, 2021, Kuala Lumpur, Malaysia.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3457784.3457832

1 INTRODUCTION

For the last two decades, User Experience (UX) has been a popular research field in the community of Human Computer Interaction (HCI) [1]. In early days, the community in HCI mostly focused in analyzing and evaluating behavioral goals, but Hassenzahl et al. [2] argued that HCI should not only focus on pragmatic aspect (behavioral goals) but also hedonic aspect of the interaction between product and user. As it is mentioned in ISO 9241-110:2020 [4], UX defined as: "combination of user 's perceptions and responses that result from the use and/or anticipated use of a system, product or service". So, it can be interpreted that UX involves user's feeling and thought when interacting with the product. While Alben [5] defined experience as what users feeling when they interact with the product, how users understand the ways to use the product and if it meets their goals. UX focus on user's feeling, thought and behavior when interacting with the product or system [2]. That is why UX has become a more popular research field in HCI than usability in recent studies. Usability mainly focused on task and performance [2, 3]. It has limitation comparing to UX that focused in task-related aspect and non-task related aspect.

Measuring user's feeling while interacting with a product, system or service is a significant challenge in UX evaluation. Measuring UX using self-reported metric is the traditional method and most common method in user experience evaluation method (UXEM) [7]. This method is highly subjective and thus dependent on user's interpretation. These subjective natures make a disadvantage in evaluating UX because they create bias in their result [6, 7]. And even if the questionnaire could be correctly interpreted by a user, sometimes the user does not give the answer that is true to their emotions. This also makes the self-reported measurement unable to acquire the true feeling of the user' emotional experience [7]. Observational measurement method is another method in UXEM that focused on objective value. This method observes user behavior while interacting with a product or system. But this method cannot observe user emotional state [7]. Because of this disadvantage, researcher has begun using modern technologies to help evaluate UX. The UXEM using modern technologies to measure the physiological states of the user is called psychophysiological measures method. The most common of this method include the use of galvanized skin response (GSR) to evaluate stress and arousal through skin activities, electroencephalography (EEG) to detect user emotional response by evaluating brain activities, electromyogram (EMG) to detect stress level or muscular activities, eye tracking to measure eye movement and visual attention [6-9].

With the growth of UX research field, researchers have developed various ways to implement UXEM. These evaluation methods have different practice, evaluated application, and type of collected metric. Arifin et al. [10] reviewed research papers related to UX component, evaluated application, measured metric, and UX aspect from 2010 to 2017. They categorized the metric based on the proposed standard metric developed by Albert et al. [17]. But they only reviewed research papers in augmented reality domain and the research paper did not review the method used for UX evaluation. Zarour et al. [1] reviewed research papers that related to UX aspect and dimension from 2005 to 2015. Their main purpose was to discover the relation between UX aspect and UX dimension.

ICSCA 2021, February 23–26, 2021, Kuala Lumpur, Malaysia

ACM ISBN 978-1-4503-8882-5/21/02…$15.00
https://doi.org/10.1145/3457784.3457832

Table 1: Exclusion and Inclusion Criteria

Inclusion Criteria	Exclusion Criteria
The article is journal or conference paper	Literature review paper
The article is published in English Language	Full paper not accessible
An empirical study	Duplicate paper
Answered all research question	Non-IT domain

They did not review the UX evaluation and the evaluated product. Rivero et al. [11] created a systematic mapping study of UX evaluation between 2010 and 2015. The systematic mapping study reviewed UXEM and its implementation, the evaluated product, and the type of collected data. Their research paper did not review the metric evaluated by UXEM, they only reviewed the type of collected data (e.g., qualitative or quantitative data). The research method used by Rivero et al. [11] is similar to the research method used by Vermeeren et al. [12] because Rivero et al. [11] research is a continuation from Vermeeren et al. research [12].

Previous researches have not done a review of implementation UXEM and its metric. Arifin et al. [10] reviewed the evaluated product, UX aspect, and metric, but did not review the UXEM. Zarour et al. [1] reviewed UX aspect and UX dimension, but did not review the UXEM and evaluated product. Rivero et al. [11] reviewed the evaluated product, UXEM, and its implementation, but did not review the evaluated metric. This research paper is different from the previous research mentioned above. This paper reviews UXEM to find the relation between UXEM with the evaluated metric and the implementation of UXEM. This research paper focuses on UXEM published from 2000 to 2019. The result of this paper is expected to create an overview of the development of UXEM.

2 METHOD

To understand the development of UX evaluation, this research reviews papers published in Scopus, Science Direct IEEE Xplore, ACM Digital Lib, Emerald insight from 2000 to 2019. The inclusion and exclusion criteria used in this study are mentioned in Table 1

This paper main purpose is to review UXEM and how the method is implemented, the evaluated product, and the metric that measured using UXEM. The research questions that were used to collect relevant research papers are as follows:

- What User Experience Evaluation Method (UXEM) that was used and how the method was implemented?
- What kind of product that the method is applied to?
- What type of metric that was evaluated by the UXEM?

Keywords are extracted from the research question as the search term was used to collect relevant research papers. The keyword that was used for the search term is: (UX OR "User Experience" OR "User perception*") AND (Measurement OR Evaluation) AND (Metric OR Attribute OR Aspect OR Scale).

3 LITERATURE REVIEW

3.1 Theories of User Experience

In HCI, the term of usability and UX have a close relationship. ISO 9241-110:2020 [4] describe usability as "a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use". From that definition, it can be interpreted that usability focuses on the functionality of the product and whether or not the product has attained certain goals while the user uses a product. ISO 9241-110:2020 [4] also mention the definition of UX as "combination of user's perceptions and responses that result from the use and/or anticipated use of a system, product or service". The definition of UX is different than usability. While usability focuses on the goal-related interaction between user and product, UX focuses on user's response and perception that include emotion, preference, and behavior of the user.

There are various perceptions regarding the connection between UX and usability. Følstad et al. [13] present three viewpoints regarding the connection between UX and traditional usability. The first viewpoint is "UX encompass usability" which defined UX as a broad concept that contains usability and other concepts. The second viewpoint is "UX complements usability" which defined UX as usability that focuses on measuring non-goal related aspects. The last viewpoint is "UX is one of several components constituting usability" which defined UX as part of the usability concept [13]. Hassan et al. [14] performed a research that reviews the author's perception of UX and usability relationships.

Hassenzahl [15] defines a key element of UX from user's perspectives such as apparent product character and consequence. Apparent product character is the product feature perceived by an individual when they interact with the product. The product character perceived by the user consists of pragmatic attributes and hedonic attributes. Pragmatic attribute focuses on utility (i.e., relevant functionality of product) and usability (i.e., ways to access this functionality), while hedonic attribute focuses on a user physiological well-being. Hedonic attributes are divided into stimulation, identification, and evocation. Consequence is related to apparent product character as the user assessment of product features. The consequence could lead to the user emotional consequence e.g. satisfaction and pleasure or behavioral consequence e.g. time spent using the product [15].

Mahlke [16] proposes a framework called CUE-model (Component of User Experience). Based on CUE-model, UX component consist of instrumental qualities, non-instrumental qualities, and emotional reaction. Instrumental qualities focusses on utility (i.e., product functionality that fulfill user requirements) and usability (i.e., how well users utilize the product). While non-instrumental qualities focusses on aesthetic, symbolic value, and motivational aspect. Aesthetic aspect of non-instrumental qualities related to the appearance of the product. Symbolic value of non-instrumental qualities is subdivided into two dimension, communicative and associative aspects. Associative aspect relates to personal memories (i.e.,

what the user associate with product attributes) while communicative aspect relates to individual qualities (i.e., self-expressiveness of the user). Motivational aspect of non-instrumental qualities defined as product capability to motivate the user. The last component of CUE-model is emotional reaction. Emotional reaction is influenced by the perception of instrumental and non-instrumental qualities [16].

The model proposed by Hassenzahl [15] and Mahlke [16] is similar. Mahlke CUE-model [16] consists of instrumental qualities that are reflected in pragmatic aspect of Hassenzahl's UX model [15] and non-instrumental qualities that are reflected in hedonic aspect of Hassenzahl's UX model [15]. Both values affect the consequence of interaction e.g. user's emotion. From the review of Hassan et al. [14] and the model proposed by Mahkle [16] and Hassenzahl [15], it could be interpreted that usability is a subset of UX. Usability included in instrumental qualities (pragmatic attribute) of UX component focuses on product functionality and goal-related aspect. While aspects that is not related to the goal of the product (e.g., aesthetic aspect) were included in non-instrumental qualities (hedonic attribute).

3.2 User Experience Evaluations Method (UXEM)

There are many methods of UX evaluation that have been developed. Hussain et al. [7] categorized UXEM into three categories: (1) Self-reported measurement method, (2) Observational measurement method, and (3) Physiological measurement method. Self-reported measurement method is the most common method. In this method, users report their perspective and experience about the product, system or service without any need of expert intervention. The second measurement method is observational measurement method. In observational method, there is a need for expert intervention. Unlike in self-reported measurement method, observational measurement method requires an expert to observe the user while interacting with the product. The last method is physiological measurement method. Physiological measurement method is used to measure the biological information of a user. The biological information reflects how the user feel during experiencing the product [7].

3.3 User Experience Metric

Metric is a value that is measured to indicate a certain phenomenon or thing for example temperature, weight, height, time, and many others. In UX domain, metric is also used as an indicator that is measured to gauge the experience of the user while interacting with a product, system, or service. With UX metric, the researcher can discover a user's feeling and experience with the evaluated product or discover an improvement in the product [17].

Albert and Tullis [17] present four types of metric that are used in UX domain. The first metric is performance metric. This metric is used to measure user's behavior while interacting with a product or system for example user journey (i.e., user behavior) to reach certain web pages (i.e., user goal). The most common performance metric is the task success, error, or time to take to reach the goal. The second metric is issue-based metric. This metric is used to measure issues or problems that the user encounters while interacting with the product. The third metric is self-reported metric. Self-reported metric is used to measure the user's perception or opinion of a product, system, or service. The last metric is behavioral and physiological metric. This metric is used to measure user's body behavior or reaction when interacting with the product. There is a possibility that there is a difference in user's thought and what user's really feels. Behavior and physiological metric show what user's really feels, that is why this metric is useful to gauge UX in objective perspective.

Hussain et al. [7] categorized UX metric into 3 types of metric: interaction metric, self-reported metric, and emotion and stress metric. The definition of self-reported metric is similar between Hussain et al. [7] and Albert and Tullis [17], which is to record user's perception or opinion of product, system, or service. Interaction metric divided into behavior metric and performance metrics. The definition of performance metric is the same as performance metric defined by Albert and Tullis [17], but behavior metric is grouped within behavior and physiological metric instead with performance metric. Hussain et al. [7] performance metric focus on the value of the user's capability while using the product for example task time, task success, or error. While behavior metric focus on the value of common interaction between a user and product e.g. page/screen, event such as user click. The definition of emotion and stress metric is similar to the definition of behavioral and physiological metric by Albert and Tullis [17]. Hussain et al. [7] did not mention issue-based metric, a metric that indicates issue or problem that the user encountered.

4 RESULT

This section presents the result of our analysis of the research paper. There are 3.231 papers (1.826 papers from Scopus, 284 papers from Science Direct, 484 papers from IEEE Xplore, 618 papers from ACM Digital Lib, and 19 papers from Emerald insight) that were found based on the search query. There are 128 papers that fulfilled all inclusion and exclusion criteria. From 128 papers, only 61 papers will be reviewed. 67 papers were excluded due to it having quality value less than 2.5 in quality assessment.

This section describes the user UXEM that has been implemented in the reviewed paper and what is the evaluated product and evaluated metric. The result of the review is described in Table 2. Table 2 illustrate users experience evaluation method and how to implement the experiment (data source, the location of the experiment, and period of evaluation), evaluated product, and evaluated metric and collected data type. The result of the literature review is described in the next paragraph to answer the research question.

4.1 What User Experience Evaluation Method (UXEM) That Used and How the Method Is Used?

UXEM is categorized into three types of method: Self-reported measurement, observational measurement and physiological measurement [7]. To improve the result of UX evaluation, few researchers usually apply a triangulation method which means that they used more than one method and sometimes from different types of methods [17]. This study found that the most frequently implemented method is self-reported method (95%) with 17 papers that use self-reported measurement and observational measurement

(e.g., [23, 27, 29, 30, 34, 36, 46, 49, 51, 54, 55, 57, 59, 62, 67, 68, 73]), 4 papers that use self-reported measurement and physiological measurement (e.g., [18, 22, 41, 52]), and 33 papers that only use self-report measurement (e.g., [19–21, 24–26, 28, 31–33, 35, 37–40, 42, 45, 47, 48, 50, 53, 60–64, 66, 69–71, 74–77]). The second measurement method is observational measurement (37%) and the third measurement method is physiological measurement (14%). There are 2 papers that only use observational method (e.g., [43, 56]), 1 paper that only uses physiological method (e.g., [65]) and 4 papers that use all measurement methods (e.g., [22, 44, 72, 78]).

Most of the research did not mention the method of sampling used in the experiment. The most frequently used sampling is convenience sampling with total 10 papers (e.g., [23, 25, 27, 33, 38, 45, 47, 59, 75, 76]). The other sampling method is purposive sampling (5 papers e.g. [28, 44, 56, 61, 63]), random sampling (4 papers e.g. [35, 39, 40, 61]), quota sampling (1 paper e.g. [26]), snowball sampling (1 paper e.g. [33]), and stratified sampling (1 paper e.g. [73]).

The most frequently period of evaluation for implementing self-reported measurement is after usage. There are 44 papers that implement self-reported measurement after a user used a product (e.g., [18–20, 22, 26], [28-35], [37, 38, 40–42, 44–46, 49], [51-55], [57, 59], [60-63], [67-76], [78]). The second is during usage with 12 papers in total (e.g., [21, 23, 30, 32, 34, 36, 44, 48, 55, 58, 59, 61]). Long term period of evaluation is implemented with 6 papers (e.g., [24, 25, 27, 33, 36, 64, 77]) in self-reported measurement method. The least frequently implemented period of evaluation for self-reported measurement is before usage with total 5 papers (e.g., [28, 29, 31, 60, 72]). There are 4 papers (e.g., [39, 47, 50, 66]) that did not explain the period of evaluation for self-reported measurement. Self-reported measurement was most frequently implemented in controlled environment (lab) with 24 papers in total (e.g., [18, 19, 22, 23, 26, 30, 35–37, 42, 48, 49, 51–53, 55, 59, 60, 62, 63, 69, 71, 72, 78]). The next most frequently implemented is online or automatically with total 13 papers (e.g., [24, 25, 27, 34, 36, 49, 64, 66, 73–77]). The least frequently implemented location for self-reported measurement is in non-controlled environment (field) with 6 papers in total (e.g., [20, 33, 45, 46, 54, 68]). There are 15 papers (e.g., [21, 29, 31, 32, 38–40, 44, 47, 50, 57, 58, 61, 67, 70]) that did not explain the location of self-report measurement.

As for the period of evaluation of observational measurement method, the most frequently period of evaluation for implementing observational measurement is during usage with total 20 papers (e.g., [22, 23, 29, 30, 34, 43, 44, 46, 49, 51, 54–57, 59, 62, 67, 72, 73, 78]). There are 2 papers that implement observational measurement in long-term period of evaluation (e.g., [27, 36]). For location of experiment for observational measurement, there are 13 papers that implement observational measurement in controlled environment (lab) e.g. [22, 23, 30, 49, 51, 55, 56, 59, 62, 68, 72, 73, 78]. The next most frequent, observational measurement is conducted over the internet (online) with 5 papers in total e.g. [27, 34, 36, 43, 49]. There are 3 papers (e.g., [46, 54, 68]) implement observational measurement in non-controlled environment (field), while there are 4 papers (e.g., [29, 44, 57, 67]) that did not explain the location of observational measurement method.

For physiological measurement, the only period of evaluation is during usage. There are 2 papers (e.g., [41, 44]) that did not explain the location of evaluation for physiological measurement, the location of physiological measurement is only implemented in controlled environment (lab). This restriction is possibly caused by the necessary equipment needed to measure bioinformatic data of the user. Therefore, researchers cannot implement physiological measurement for long-term period of evaluation or outside of controlled environment.

4.2 What Kind of Product the Method Is Applied to?

For the evaluated product there are many different products that the evaluated method applied to. Most frequently paper evaluated product within educational domain (17 papers e.g. [8, 18, 34, 35, 38, 39, 41, 44, 45, 49, 59, 61, 63, 71, 73, 74, 76]), for example educational institution website, e-Learning system, Learning Management System (LMS), mobile application for education, educational desktop game, educational Virtual Reality (VR) system, VR serious game and student information system. The second most frequently paper is product in health domain (15 papers e.g. [21, 24, 25, 27, 30, 32, 36, 37, 50, 51, 55, 64, 67, 75, 77]), for example exercise game, m-Health application, Health Information System (HIS), bioinformatic software and Health Information Exchange (HIE). There are 6 papers (e.g., [23, 26, 28, 29, 56, 70, 72]) that evaluate products in entertainment domain for example mobile game application, VR game, and Brain–Computer Interface (BCI) game. For evaluated products in travel and guide domain there are 6 papers in total (e.g., [21, 28, 31, 40, 43, 58]). There are 4 papers (e.g., [20, 47, 53, 62]) that evaluated product within business and finance domain, for example digital wallet, cryptocurrency website, crowdsourcing mobile application, and ERP (Enterprise Resource Planning) system. Product in commercial domain is also evaluated with a total of 4 papers (e.g., [46, 48, 52, 57]). In social and communication domain, there are 2 papers in total (e.g., [19, 21]). 6 papers (e.g., [22, 33, 54, 65, 66, 78]) are evaluated as news and book for example online news, e-Book, or e-Reader. There are 2 papers (e.g., [31, 68]) that evaluate product in industrial domain. The least evaluated product is in culinary domain (e.g., [42]) and home management domain (e.g., [60]).

4.3 What Type of Metric That Was Evaluated by the UXEM?

Based on the result, most of the paper that were reviewed is collecting metric with quantitative data (60 papers e.g. [18-44], [46-78]). While, qualitative data is measured in 21 papers (e.g., [19, 21, 32, 33, 35–37, 42, 45, 48, 50, 55, 57–61, 63, 66, 68, 75]). Qualitative data consist of issue-based metric and self-reported metric. Though, most of the self-reported metric is categorized into quantitative data. All of Interaction metric and stress and emotion metric consist of quantitative data.

There are 12 papers (e.g., [21, 23, 32, 36, 37, 42, 47, 57–59, 61, 68]) that measure issue-based metric. This evaluated metric represents problems that user encounter during interaction. There are also paper that measure severity level and the number of usability issue. 9 reviewed papers (e.g., [18, 22, 41, 52, 56, 57, 65, 72, 78]) measure

Table 2: The Distribution of User Experience Evaluation, Procedure, Evaluated Product and Evaluated Metric

Ref	User Experience Evaluation Method (UXEM)							Evaluated Product	Evaluated Metric					Data Type	
	UXEM Categories			Theo.	Method of	Exp. Loc.	Eval.		SR	INT		SE	IB	QN	QL
	SR	OBS	PHY	Framework	Sampling		Per.			B	P				
[18]			GSR				DU	Edu. game				√		√	
			BVP									√		√	
	Quest.						AU		√					√	
[19]	Quest.					Lab	AU	Communication	√					√	
	Interview							mobile app.	√						√
[20]	AttrakDiff					Field	AU	Mobile wallet SCWS wallet	√					√	
[21]	HE						DU	Social web. svc. Travel web. svc. Health web. svc.					√	√	√
[22]	Smiley-o-Meter				Convenience	Lab	AU	e-Book	√					√	
	Interview						AU		√					√	
			ET				DU					√		√	
		PE					DU				√			√	
	Ranking survey						AU		√					√	
[23]	HE					Lab	DU	Travel booking					√	√	
		PE						web.			√			√	
[24]	Quest.					Online	LT	HIS	√					√	
[25]	QUIS				Convenience	Online	LT	HIE	√					√	
[26]	GEQ				Quota	Lab	AU	Mobile game app.	√					√	
[27]	Quest.			UTAUT	Convenience	Online	LT	HIS	√					√	
		INTR								√				√	
[28]	Attrakdiff				Purposive	Lab	BU	Travel Mobile AR	√					√	
	SUXES						AU	app.	√					√	
	Emocard						AU	Virtual Pet AR app.	√					√	
[29]		INTR			Convenience		DU	VR game		√	√			√	
	ITC-SOPI						BU	Desktop game	√					√	
	Presence						AU								
	SSQ								√					√	
[30]	Think aloud					Lab	DU	m-Health app.	√						√
		PE									√			√	
	Quest.						AU		√					√	
[31]	SUXES						BU	Travel public	√					√	
	Quest.			Experience Pyramid			AU	display app. Industrial public display app.	√					√	
[32]	Think aloud						DU	Bioinformatic software					√		√
	SUMI						AU		√					√	
	Interview												√		√
[33]	Flow Short Scale				Convenience Snowball	Field	LT	e-Book	√					√	
	Interview						AU		√						√
[34]	ASQ					Online	AU	Edu. web.	√					√	
	WAMMI								√					√	
		INTR					DU			√	√			√	

[35]	Quest.				Random	Lab	AU	BCI serious game	√				√		
	Quest.								√						√
[36]	Think aloud					Lab	DU	HIS				√			√
	Quest.					Online	LT		√				√		
		VULab log data								√			√		
	VULab quest.								√						√
[37]	SUS					Lab	AU	Exercise game	√				√		
	Interview			UTAUT				software				√			√
	Ranking card								√				√		
[38]	GEQ				Convenience		AU	Edu. games	√				√		
	IMI								√				√		
[39]	Checklist				Random			Edu. web.	√				√		
	Quest.								√				√		
[40]	UEQ				Random		AU	TSP app.	√				√		
[41]			EEG				DU	Edu. VR				√	√		
	SAM						AU		√				√		
[42]	Quest.					Lab	AU	Culinary mobile app. Culinary wearable app.	√				√	√	√
[43]		INTR				Online	DU	Travel online quiz web.		√	√			√	
[44]	Quest.				Purposive		AU	Educational game	√					√	
		INTR					DU	web. app.		√				√	
		FEO										√		√	
			GSR									√		√	
			BVP									√		√	
			RESP									√		√	
	Quest.			PAD model					√					√	
[45]	Interview				Convenience	Field	AU	VR serious game	√						√
[46]		PE				Field	DU	m-Commerce svc.			√			√	
	SUS						AU		√					√	
[47]	HE				Convenience			ERP system					√	√	
[48]	SUS					Lab	AU	m-Commerce app.	√					√	
	Think-aloud						DU		√						√
[49]		INTR				Lab	DU	e-Learning system		√	√			√	
	SUS					Online	AU		√					√	
[50]	SUS							Fall prevention	√					√	
	PACES							system	√					√	
	Quest.			DART					√					√	
	Interview								√						√
[51]	SUS					Lab	AU	m-Health app.	√					√	
		PE					DU				√			√	
[52]	Quest.					Lab	AU	House rent web.	√					√	
			ET				DU					√		√	
			HRV									√		√	
[53]	Quest.			ISO 25010		Lab	AU	Crowdsourcing mobile app.	√					√	
[54]	Quest.					Field	AU	e-Reader mobile app.	√					√	
		PE					DU				√			√	

[55]		PE				Lab	DU	m-Health app.		√		√		
	Think aloud								√				√	
	Quest.						AU		√			√		
	Emoji card								√			√		
[56]		FEO			Purposive	Lab	DU	Mobile game app.			√	√		
[57]	SAM						AU	Travel mobile app.	√			√		
		FEO					DU				√		√	
	Quest.						AU					√	√	
[58]	HE						DU	Museum guide mobile app.				√	√	√
[59]	SUS				Convenience	Lab	AU	e-Learning system	√			√		
	AttrakDiff								√			√		
	Interview								√				√	
	Think aloud						DU					√		√
		PE								√		√		
[60]	MDMQ					Lab	BU AU	Spatial AR app.	√			√		
	SAM								√			√		
	Quest.						AU		√			√		
	Quest.								√			√		
	Quest.								√				√	
[61]	UEQ				Random		AU	SIsKA web.	√			√		
	HE				Purposive		DU				√	√	√	
[62]	SEQ					Lab	AU	Cryptocurrency web.	√			√		
	SUS								√			√		
	QUIS								√			√		
	USE								√			√		
	Product reaction cards								√			√		
		PE					DU			√		√		
[63]	UEQ				Purposive	Lab	AU	LMS	√			√		
	Interview								√				√	
[64]	NuHISS					Online	LT	HIS	√			√		
[65]			EEG			Lab	DU	Online news article			√	√		
[66]	UEQ					Online		e-Learning system	√			√		
	HE quest.								√				√	
[67]	SUS			ISO 9241-11			AU	m-Health app.	√			√		
		PE					DU			√		√		
[68]		PE				Field	DU	Industrial AR HMD app.		√		√		
	Interview						AU		√		√		√	
[69]	GUESS					Lab	AU	VR game Desktop game	√			√		
	Presence Scale								√			√		
[70]	UEQ						AU	Entertainment mobile AR app.	√			√		
[71]	Quest.					Lab	AU	Edu. VR system	√			√		
[72]	Quest.					Lab	BU	Entertainment mobile AR app.	√			√		
		PE					DU			√		√		
			EEG								√	√		
	UEQ						AU		√			√		
	SUDs								√			√		

[73]	SUS				Stratified	Online	AU	Student	√				√		
	EWPL							information system	√				√		
	Quest.								√				√		
		PE				Lab	DU						√		
[74]	Quest.					Online	AU	Edu. game	√				√		
[75]	uMARS				Convenience	Online	AU	HIS	√				√		
	Quest.								√						√
[76]	PEDEACLWQ				Convenience	Online	AU	Edu. web.	√				√		
[77]	Quest.					Online	LT	HIS	√				√		
[78]	UES					Lab	AU	News web.	√				√		
	CAS								√				√		
	SUS								√				√		
		INTR					DU			√	√		√		
			HRV									√	√		
			EDA									√	√		
			EMG									√	√		

User Experience Evaluation Method (UXEM). Theo. Framework: Theoretical Framework; Exp. Loc.: Experiment Location; Eval. Per.: Evaluation Period.
UXEM Categories. SR: Self-Reported; OBS: Observational; PHY: Physiological.
Eval. Per. AU: After Usage; DU: During Usage; BU: Before Usage; LT: Long Term
Evaluated Metric. SR: Self Report; INT: Interaction; B: Behavior; P: Performance; SE: Stress and Emotion; IB: Issue-Based
Data Type. QN: Quantitative; QL: Qualitative
SR. QUIS: Questionnaire for User Interface Satisfaction; GEQ: Game Experience Questionnaire; SSQ: Simulator Sickness Questionnaire; SUMI: Software Usability Measurement Inventory; ASQ: After-Scenario Questionnaire; WAMMI: Website Analysis and Measurement Inventory; SUS: System Usability Scale; IMI: Intrinsic Motivation Inventory; SAM: Self-Assessment Manikin; PACES: Physical Activity Enjoyment Scale; MDMQ: Multidimensional Mood State Questionnaire; SEQ: Single Ease Question; UEQ: User Experience Questionnaires; NuHISS: National Usability-focused HIS Scale; GUESS: Game User Experience Satisfaction Scale; SUDs: Subjective units of distress scale; EWPL: Emotion Words Prompt List; uMARS: Mobile App Rating Scale; PEDEACLWQ: Perceived Design Effectiveness of Nigerian Academic Library Websites Questionnaire; USE: Usefulness, Satisfaction, and Ease of use Questioner; CAS: Cognitive Absorption Scale; UES: User Engagement Scale; Quest.: Questionnaire
OBS. PE: Performance evaluation; INTR: Interaction record; FEO: Facial expression observation
PHY. GSR: Galvanic Skin Response; BVP: Blood Volume Pulse; ET: Eye Tracking; EEG: Electroencephalography; RESP: Respiration; HRV: Heart Rate Variability; EDA: Electrodermal Activity; EMG: Electromyography
Theo. Framework. UTAUT: Unified Theory of Acceptance and Use; DART: The Dynamic Acceptance Model for the Re-evaluation of Technologies; PAD: Pleasure Arousal Dominance
Evaluated Product. App.: Application; HIS: Health Information System; HIE: Health Information Exchange; AR: Augmented Reality; VR: Virtual reality; ERP: Enterprise Resource Planning; SIsKA: Academic Progress Information System; LMS: Learning Management System; TSP: Travelling Salesman Problem; Web.: Website; Edu.: Educational; Svc.: Service; BCI: Brain–Computer Interface

emotion and stress metric. The metric is evaluated from physiological based emotion recognition to measure bioinformatics (e.g., visual attention, heart beats interval, skin conductivity) and video-based emotion recognition to observe emotion from observing facial expression. Interaction metric is divided into 2 sub metric, performance metric and behavior metric. There are 18 papers (e.g., [22, 23, 29, 30, 34, 43, 46, 49, 51, 54, 55, 59, 62, 67, 68, 72, 73, 78]) that measure performance metric, while there are 8 papers (e.g., [27, 29, 34, 36, 43, 44, 49, 78]) that measure behavior metric. 54 papers (e.g., [18-20], [22], [24-42], [44-55], [57], [59-64], [66-78]) measure self-reported metric for example pragmatic quality, hedonic quality aesthetic, usability, enjoyment, emotional response, perceived usefulness, learnability, ease of use, etc.

5 DISCUSSION

Self-reported measurement method could be used for measuring the problem a user may encounter during interacting with product. The value is categorized into issue-based metric as mentioned by Albert and Tullis [17]. Although most of the issue-based metric is categorized as qualitative data, there are some research that categorized issue-based metric in quantitative data for measuring the severity rating of the usability issue (e.g., [42, 47, 58, 61]). The method used for evaluating issue-based metric is heuristic evaluation (e.g., [21, 23, 47, 58, 61]), interview (e.g., [32, 37, 68]), think aloud (e.g., [32, 36, 59]) and questionnaire (e.g., [42, 57]). Beside issue-based metric, Self-reported measurement evaluating self-report metric in form of qualitative data and quantitative data. Method used for evaluating such metric is think aloud (e.g., [30, 49, 55]),

interview (e.g., [19, 22, 33, 45, 50, 53, 63, 68]), non-standard questionnaire (e.g., [18, 19, 22, 24, 27, 30, 31, 35–37, 39, 42, 44, 50, 52–55, 57, 60, 71, 72, 74, 75, 77]) and standard questionnaire (e.g., [20, 22, 25, 26, 28, 29], [31-34], [37, 38, 40, 41, 46, 48–50, 55, 57, 59], [60-64], [66, 67, 69, 70], [72-76], [78]) . The most used method for evaluating self-report metric is standard questionnaire. The most common method used for standard questionnaire is System Usability Scale (e.g., [37, 46, 49–51, 53, 62, 67, 73, 78]) and User Experience Questionnaire (e.g., [40, 61, 63, 70, 72]). Some of the non-standard questionnaire is developed based on theory, for example Unified Theory of Acceptance and Use model (e.g., [27, 37]), Experience Pyramid theoretical (e.g., [31]), Pleasure-Arousal-Dominance emotion representation model (e.g., [44]), The Dynamic Acceptance Model for the Re-evaluation of Technologies (e.g., [50]), ISO 25010 (e.g., [53]), ISO 9241-11 (e.g., [67]).

Observational method is mostly used to measure interaction method. As previously explained Hussain et al. [7], interaction metric divided again into performance metric and behavior metric. Observational method that mostly used to measure performance metric is interaction record and performance evaluation. Interaction record may assist performance metric measurement that was conducted over the internet (e.g., [34, 43, 49]) or in controlled environment (e.g., [49, 78]). Performance evaluation mostly implemented in controlled environment (e.g., [22, 23, 30, 34, 51, 53, 55, 62, 78]) or in non-controlled environment (e.g., [46, 54, 68, 72, 73]). So, it can be concluded that for performance metric, the most common method is performance evaluation implemented in controlled environment. For behavior metric, observational method that is used is only interaction record. The method is implemented over the internet (e.g., [27, 34, 36, 43, 49]) or in controlled environment (e.g., [49, 78]). Other than interaction metric, observational method measure stress and emotional metric such as facial expression observation (e.g., [56, 57])

Physiological method is mostly used for measuring stress and emotion metric. Physiological method could be used to measure eye movement (Eye tracking e.g. [22, 52]) or bioinformatic such as GSR (galvanic skin response) to measure skin conductivity (e.g., [18, 44]) , BVP (Blood Volume Pulse) to measure blood pulse (e.g., [18, 44]), HRV (Heart Rate Variability) to measure heartbeat interval (e.g., [52, 78]), electrodermal activity (EDA) to measure produced sweat (e.g., [78]), measure respiration (e.g., [44]), electromyography (EMG) to measure muscle activity (e.g., [78]), and EEG to measure brain activity (e.g., [41, 65, 72]). By using physiological information, the researcher can conclude the emotion that the user feels during interaction with the product for example joy, fear, anger, frustration, etc. Few papers implement physiological metric even though this specific method is appropriate for evaluating user's emotion in objective perspective. This measurement method could also only be implemented in a controlled environment due to the need for certain equipment.

6 CONCLUSION

This research paper presents a systematic literature review of the UX evaluation research. The review has found that the most frequently used method in UXEM is self-reported metric and the least frequently used method is physiological measurement method. Self-reported measurement and observational method could be implemented over the internet (online), in controlled environment (lab), and non-controlled environment (field) while physiological method is only implemented in controlled environment and only in a certain time period of usage (during usage). The result of this study may be used as a reference by researchers or students for analysis regarding UX evaluation, the procedure for implementation, and evaluated metric.

ACKNOWLEDGMENTS

This research was supported by Hibah Publikasi Terindeks Internasional (PUTI) Q2 2020 at Universitas Indonesia (Number: NKB-846/UN2.RST/HKP.05.00/ 2020).

REFERENCES

[1] Mohammad Zarour and Mubarak Alharbi. 2017. User Experience Aspects and Dimensions: Systematic Literature Review. Cogent Engineering 4, 1 (December 2017), 52-59. https://doi.org/10.1080/23311916.2017.1421006

[2] Marc Hassenzahl and Noam Tractinsky. 2006. User experience – a research agenda.Behaviour & Information Technology 25, 2 (March 2006), 91-97. https://doi.org/10.1080/01449290500330331

[3] Effie Lai-Chong Law, Virpi Roto, Marc Hassenzahl, Arnold P.O.S. Vermeeren, and Joke Kort. 2009. Understanding, scoping and defining user experience: a survey approach. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09). Association for Computing Machinery, New York, NY, USA, 719–728. https://doi.org/10.1145/1518701.1518813

[4] ISO 9241-110. 2020. https://www.iso.org/obp/ui/#iso:std:iso:9241:-110:ed-2:v1:en

[5] Lauralee Alben. 1996. Quality of experience: defining the criteria for effective interaction design. interactions 3, 3 (May/June 1996), 11–15. https://doi.org/10.1145/235008.235010

[6] Manon Holman and Funmi Adebesin. 2019. Taking the Subjectivity out of UX Evaluation with Emotiv EPOC+. In Proceedings of the South African Institute of Computer Scientists and Information Technologists 2019 (SAICSIT '19). Association for Computing Machinery, New York, NY, USA, Article 30, 1–10. https://doi.org/10.1145/3351108.3351139

[7] Jamil Hussain, Wajahat Ali Khan, Taeho Hur, Hafiz Syed Muhammad Bilal, Jaehun Bang, Anees Ul Hassan, Muhammad Afzal, and Sungyoung Lee. 2018. A Multimodal Deep Log-Based User Experience (UX) Platform for UX Evaluation. Sensors 18, 5(18 May 2018), 1622. https://doi.org/10.3390/s18051622

[8] Fu Guo, Yi Ding, Weilin Liu, Chang Liu, and Xuefeng Zhang. 2016. Can eye-tracking data be measured to assess product design? Visual attention mechanism should be considered. International Journal of Industrial Ergonomics 53 (May 2016), 229-235. https://doi.org/10.1016/j.ergon.2015.12.001

[9] Camila Loiola Brito Maia and Elizabeth S. Furtado. 2016. A study about psychophysiological measures in user experience monitoring and evaluation. In Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems (IHC '16). Association for Computing Machinery, New York, NY, USA, Article 7, 1–9. https://doi.org/10.1145/3033701.3033708

[10] Yulyani Arifin, Thomas Galih Sastria, and Edo Barlian.2018. User Experience Metric for Augmented Reality Application: A Review, Procedia Computer Science 135 (January 2018), 648-656. https://doi.org/10.1016/j.procs.2018.08.221"

[11] Luis Rivero and Tayana Conte. 2017. A Systematic Mapping Study on Research Contributions on UX Evaluation Technologies. In Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems (IHC 2017). Association for Computing Machinery, New York, NY, USA, Article 5, 1–10. https://doi.org/10.1145/3160504.3160512

[12] Arnold P. O. S. Vermeeren, Effie Lai-Chong Law, Virpi Roto, Marianna Obrist, Jettie Hoonhout, and Kaisa Väänänen-Vainio-Mattila. 2010. User experience evaluation methods: current state and development needs. In Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordiCHI '10). Association for Computing Machinery, New York, NY, USA, 521–530. https://doi.org/10.1145/1868914.1868973

[13] Asbjørn Følstad and Rolf Kenneth Rolfsen. 2006. Measuring the Effect of User Experience Design Changes in e-Commerce Web Sites: a Case on Customer Guidance. In Proceedings of the 2nd International COST294-MAUSE Open Workshop. COST, Oslo, Norway, 10-15.

[14] Hala Magdy Hassan and Galal Hassan Galal-Edeen. 2017. From usability to user experience. In Proceedings of the International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS). IEEE, Okinawa, Japan, 216-222. https://doi.org/10.1109/ICIIBMS.2017.8279761.

[15] Marc Hassenzahl. 2003. The Thing and I: Understanding the Relationship Between User and Product. In: Blythe M.A., Overbeeke K., Monk A.F., Wright P.C. (eds) Funology. Human-Computer Interaction Series, vol 3. Springer, Dordrecht. https://doi.org/10.1007/1-4020-2967-5_4

[16] Sascha Mahlke. 2008. User Experience of Interaction with Technical Systems. PhD Thesis.

[17] William Albert and Thomas Tullis. 2013. Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics. Morgan Kaufmann. https://doi.org/10.1016/C2011-0-00016-9

[18] Helmut Prendinger, Junichiro Mori, and Mitsuru Ishizuka. 2005. Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. Int. J. Hum. Comput. Stud. 62, 2 (February 2005), 231–245. https://doi.org/10.1016/j.ijhcs.2004.11.009

[19] Natasa Milic-Frayling, Martin Hicks, Rachel Jones, and Jamie Costello. 2007. On the design and evaluation of web augmented mobile applications. In Proceedings of the 9th international conference on Human computer interaction with mobile devices and services (MobileHCI '07). Association for Computing Machinery, New York, NY, USA, 226–233. https://doi.org/10.1145/1377999.1378011

[20] Gerald Madlmayr, Dominik Brandlberger, Josef Langer, and Josef Scharinger. 2008. Evaluation of SmartCard webserver as an application platform from a user's perspective. In Proceedings of the 6th International Conference on Advances in Mobile Computing and Multimedia (MoMM '08). Association for Computing Machinery, New York, NY, USA, 360–363. https://doi.org/10.1145/1497185.1497263

[21] Kaisa Väänänen-Vainio-Mattila and Minna Wäljas. 2009. Developing an expert evaluation method for user eXperience of cross-platform web services. In Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era (MindTrek '09). Association for Computing Machinery, New York, NY, USA, 162–169. https://doi.org/10.1145/1621841.1621871

[22] Areej Al-Wabil, Ebtisam Alabdulqader, Latifa Al-Abdulkarim, and Nora Al-Twairesh. 2010. Measuring the user experience of digital books with children: An eyetracking study of interaction with digital libraries. In Proceedings of the 2010 International Conference for Internet Technology and Secured Transactions. IEEE, London, 1-7.

[23] H Idyawati, Murni Mahmud, and Abu Osman Md Tap. 2010. User Experience: Assessing the effectiveness of Internet booking service. In Proceeding of the 2010 International Conference on User Science and Engineering (i-USEr). IEEE, Shah Alam, Malaysia, 1-6. https://doi.org/10.1109/IUSER.2010.5716712

[24] Johanna Viitanen, Hannele Hyppönen, Tinja Lääveri, Jukka Vänskä, Jarmo Reponen, and Ilkka Winblad, 2011. National questionnaire study on clinical ICT systems proofs: Physicians suffer from poor usability. International Journal of Medical Informatics, 80, 10, 708-725. https://doi.org/10.1016/j.ijmedinf.2011.06.010.

[25] Cynthia S Gadd, Yun-Xian Ho, Cather Marie Cala, Dana Blakemore, Qingxia Chen, Mark E Frisse, and Kevin B Johnson. 2011. User perspectives on the usability of a regional health information exchange. Journal of the American Medical Informatics Association, 18, 5 (27 May 2011), 711-6. https://doi.org/10.1136/amiajnl-2011-000281

[26] Stephan Engl, and Lennart E. Nacke. 2013. Contextual influences on mobile player experience – A game user experience model. Entertainment Computing, 4, 1, 83-91. https://doi.org/10.1016/j.entcom.2012.06.001."

[27] Annemie Heselmans, Bert Aertgeerts, Peter Donceel, Siegfried Geens, Stijn Van de Velde, and Dirk Ramaekers. 2012. Family Physicians' Perceptions and Use of Electronic Clinical Decision Support During the First Year of Implementation. J Med Syst 36, 3677-3684. https://doi.org/10.1007/s10916-012-9841-3

[28] A. Dhir and Mohammed Al-kahtani. 2013. A case study on user experience (UX) evaluation of mobile augmented reality prototypes. J. Univers. Comput. Sci. 19, 1175-1196. https://doi.org/10.3217/jucs-019-08-1175

[29] Jean-Luc Lugrin, Marc Cavazza, Fred Charles, Marc Le Renard, Jonathan Freeman, and Jane Lessiter. 2013. Immersive FPS games: user experience and performance. In Proceedings of the 2013 ACM international workshop on Immersive media experiences (ImmersiveMe '13). Association for Computing Machinery, New York, NY, USA, 7–12. https://doi.org/10.1145/2512142.2512146

[30] Jia Tan, Kari Rönkkö, and Cigdem Gencel. 2013. A Framework for Software Usability and User Experience Measurement in Mobile Industry. In Proceedings of the 2013 Joint Conference of the 23nd International Workshop on Software Measurement (IWSM) and the 8th International Conference on Software Process and Product Measurement (IWSM-MENSURA '13). IEEE Computer Society, USA, 156–164. https://doi.org/10.1109/IWSM-Mensura.2013.31

[31] Tuuli Keskinen, Jaakko Hakulinen, Tomi Heimonen, Markku Turunen, Sumita Sharma, Toni Miettinen, and Matti Luhtala. 2013. Evaluating the experiential user experience of public display applications in the wild. In Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia (<i>MUM '13). Association for Computing Machinery, New York, NY, USA, Article 7, 1–10. https://doi.org/10.1145/2541831.2541840

[32] Casper Shyr, Andre Kushniruk, and Wyeth W Wasserman. 2014. Usability study of clinical exome analysis software: Top lessons learned and recommendations. J Biomed Inform 51, (October 2014), 129-136. https://doi.org/10.1016/j.jbi.2014.05.004

[33] Luca Colombo and Monica Landoni. 2014. A diary study of children's user experience with EBooks using flow theory as framework. In Proceedings of the 2014 conference on Interaction design and children (IDC '14). Association for Computing Machinery, New York, NY, USA, 135–144. https://doi.org/10.1145/2593968.2593978

[34] Sharmistha Roy, Prasant Kumar Pattnaik, and Rajib Mall. 2014. A quantitative approach to evaluate usability of academic websites based on human perception. Egyptian Informatics Journal 15, 3, 159-167. https://doi.org/10.1016/j.eij.2014.08.002.

[35] Fotis Liarokapis, Kurt Debattista, Athanasios Vourvopoulos, Panagiotis Petridis, and Alina Ene. Comparing interaction techniques for serious games through brain–computer interfaces: A user perception evaluation study. Entertainment Computing 5, 4, 391-399. https://doi.org/10.1016/j.entcom.2014.10.004.

[36] A Kushniruk, J Kaipio, M Nieminen, H Hyppönen, T Lääveri, C Nohr, A M Kanstrup, M Berg Christiansen, M-H Kuo, and E Borycki. 2014. Human Factors in the Large: Experiences from Denmark, Finland and Canada in Moving Towards Regional and National Evaluations of Health Information System Usability. Contribution of the IMIA Human Factors Working Group. Yearb Med Inform 9, 1 (15 August 2014), 67-81. https://doi.org/10.15265/IY-2014-0023

[37] Ather Nawaz, Nina Skjæret, Kristine Ystmark, Jorunn L. Helbostad, Beatrix Vereijken, and Dag Svanæs. 2014. Assessing seniors' user experience (UX) of exergames for balance training. In Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (NordiCHI '14). Association for Computing Machinery, New York, NY, USA, 578–587. https://doi.org/10.1145/2639189.2639235

[38] Luis Gustavo Rotoly de Lima, André de Lima Salgado, and André Pimenta Freire. 2015. Evaluation of the user experience and intrinsic motivation with educational and mainstream digital games. In Proceedings of the Latin American Conference on Human Computer Interaction (CLIHC '15). Association for Computing Machinery, New York, NY, USA, Article 11, 1–7. https://doi.org/10.1145/2824893.2824904

[39] Ankur Pant. 2015. Usability evaluation of an academic library website: Experience with the Central Science Library, University of Delhi. The Electronic Library 33, 5, 896-915. https://doi.org/10.1108/EL-04-2014-0067

[40] Nova Eka Diana and Ocky Aditia Saputra. 2015. Measuring user experience of a potential shipment tracking application. In Proceedings of the International HCI and UX Conference in Indonesia (CHIuXiD '15). Association for Computing Machinery, New York, NY, USA, 47–51. https://doi.org/10.1145/2742032.2742039

[41] Carofiglio, Valeria and Ricci, Giuseppe and Abbattista, Fabio. 2015. User brain-driven evaluation of an educational 3D virtual environment. In Proceedings of 2015 10th Iberian Conference on Information Systems and Technologies (CISTI 2015). IEEE, Aveiro, Portugal, 1-7. https://doi.org/10.1109/CISTI.2015.7170553

[42] Marcin Wichrowski, Danijel Koržinek, and Krzysztof Szklanny. 2015. Google Glass Development in Practice: UX Design Sprint Workshops. In Proceedings of the Mulitimedia, Interaction, Design and Innnovation (MIDI '15). Association for Computing Machinery, New York, NY, USA, Article 11, 1–12. https://doi.org/10.1145/2814464.2814475

[43] Nouzha Harrati, Imed Bouchrika, Abdelkame Tari and Ammar Ladjailia. 2015. Automating the evaluation of usability remotely for web applications via a model-based approach. In Proceedings of the First International Conference on New Technologies of Information and Communication (NTIC). IEEE, Mila, Algeria, 1-6. https://doi.org/10.1109/NTIC.2015.7368757

[44] Jakub Miler, and Agnieszka Landowska. Designing effective educational games - a case study of a project management game. 2016. In Proceedings of the 2016 Federated Conference on Computer Science and Information Systems (FedCSIS). IEEE, Gdańsk, Poland, 1657-1661. https://doi.org/10.15439/2016F434"

[45] Bruno Fanini and Alfonsina Pagano. 2015. Interface Design for Serious Game Visual Strategies-The Case Study of "Imago Bononiae". In Proceedings of the 2015 Digital Heritage. IEEE, Granada, Spain, 623-626. https://doi.org/10.1109/DigitalHeritage.2015.7419583

[46] Pride Bongiwe Nyambi and Mamello Thinyane Telkom. 2014. Persona mapping for usability of ICTD services. In Proceedings of the 2014 IEEE 6th International Conference on Adaptive Science & Technology (ICAST). IEEE, Ota, Nigeria, 1-7. https://doi.org/10.1109/ICASTECH.2014.7068128

[47] Joelma Choma, Luciana A.M. Zaina, and Tiago Silva da Silva. 2016. Working beyond technical aspects: an approach for driving the usability inspection adding the perspective of user experience. In Proceedings of the 34th ACM International Conference on the Design of Communication (SIGDOC '16). Association for Computing Machinery, New York, NY, USA, Article 11, 1–10. https://doi.org/10.1145/2987592.2987607

[48] Yucheng Jin, Joris Klerkx, and Katrien Verbert. 2016. CircleBuy: a visual search based second screen application of buying products in videos. In Proceedings of the 8th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '16). Association for Computing Machinery, New York, NY, USA, 287–292. https://doi.org/10.1145/2933242.2935871

[49] Nouzha Harrati, Imed Bouchrika, Abdelkamel Tari, and Ammar Ladjailia. 2016. Exploring user satisfaction for e-learning systems via usage-based metrics and system usability scale analysis. Comput. Hum. Behav. 61, C (August 2016), 463–471. https://doi.org/10.1016/j.chb.2016.03.051

[50] Daryoush D. Vaziri,Konstantin Aal, Corinna Ogonowski, Thomas Von Rekowski, Michael Kroll, Hannah R. Marston, Rakel Poveda, Yves J. Gschwind, Kim Delbaere, Rainer Wieching, and Volker Wulf. 2016.Exploring user experience and technology acceptance for a fall prevention system: results from a randomized clinical trial and a living lab. Eur Rev Aging Phys Act. 13, 6 (10 Juny 2016). https://doi.org/10.1186/s11556-016-0165-z

[51] Noraini Ibrahim, Masitah Ghazali, Tan Wee Quan, Rooster Tumeng, and Shahliza Abd Halim. 2017. An evaluation study on Dengue-Entomological Surveillance system using alpha acceptance test. International Journal on Advanced Science, Engineering and Information Technology, 7, 4-2. 10.18517/ijaseit.7.4-2.3410

[52] Qing-Xing Qu, Fu Guo, and Vincent G. Duffy. 2017. Effective use of human physiological metrics to evaluate website usability: An empirical investigation from China. Aslib Journal of Information Management, 69, 4, 370-388. https://doi.org/10.1108/AJIM-09-2016-0155

[53] Amorim, Ana and Boechat, Glaucya and Novais, Renato and Vieira, Vaninha and Villela, Karina. 2017. Quality Attributes Analysis in a Crowdsourcing based emergency. In Proceedings of the 19th International Conference on Enterprise Information Systems - Volume 2: ICEIS. SciTePress, Porto, Portugal, 501-509. https://doi.org/10.5220/0006360405010509

[54] Azham Hussain, Emmanuel O. C. Mkpojiogu, Ja'afaru Musa, and Salah Mortada. 2017. A user experience evaluation of Amazon Kindle mobile application. In Proceedings of the AIP Conference Proceedings. AIP, 1 (October 2017). https://doi.org/10.1063/1.5005393

[55] Azham Hussain, Emmanuel O. C. Mkpojiogu, Najdawati Mohd Fadzil, and Norhasizasuriati Mohd Hassan. 2017. The UX of amila pregnancy on mobile device. In Proceedings of the AIP Conference Proceedings. AIP, 1 (October 2017). https://doi.org/10.1063/1.5005394

[56] Kazi Md. Munim, Iyolita Islam, Mahmuda Khatun, Md Mahboob Karim, and Muhammad Nazrul Islam. 2017. Towards developing a tool for UX evaluation using facial expression. In Proceedings of The 2017 3rd International Conference on Electrical Information and Communication Technology (EICT). IEEE, Khulna, Bangladesh, 1-6. https://doi.org/10.1109/EICT.2017.8275227

[57] Giulia de Andrade Cardieri and Luciana Martinez Zaina. 2018. Analyzing User Experience in Mobile Web, Native and Progressive Web Applications: A User and HCI Specialist Perspectives. In Proceedings of the 17th Brazilian Symposium on Human Factors in Computing Systems (IHC 2018). Association for Computing Machinery, New York, NY, USA, Article 9, 1–11. https://doi.org/10.1145/3274192.3274201

[58] Mohd Kamal Othman, Muhd Nur Shaful Sulaiman, Shaziti Aman, and Vesna Popovic. 2018. Heuristic Evaluation: Comparing Generic and Specific Usability Heuristics for Identification of Usability Problems in a Living Museum Mobile Guide App. Adv. in Hum.-Comp. Int. 2018 (2018). https://doi.org/10.1155/2018/1518682

[59] Joachim Majors, Anette Bengs, Sören Granlund, Anu Ylitalo, and Maria Byholm. 2018. Moodle Moods? A User Experience Study of a Small Private Online Course for Higher Teacher Education. In Proceedings of the 22nd International Academic Mindtrek Conference (Mindtrek '18). Association for Computing Machinery, New York, NY, USA, 228–235. https://doi.org/10.1145/3275116.3275146

[60] Lisa Schuhmacher, Anne Pagenkopf, Ragavendra Lingamaneni, and Jürgen Scheible. 2018. Emotion Enhancement through Ubiquitous Media Technology in a Smart Kitchen Environment. In Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia (MUM 2018). Association for Computing Machinery, New York, NY, USA, 317–325. https://doi.org/10.1145/3282894.3282904

[61] A.A. Istri Ita Paramitha, Gede Rasben Dantes, and Gede Indrawan. 2018. The evaluation of web based academic progress information system using heuristic evaluation and user experience questionnaire (UEQ). In Proceedings of The 2018 Third International Conference on Informatics and Computing (ICIC). IEEE, Palembang, Indonesia, 1-6. https://doi.org/10.1109/IAC.2018.8780430

[62] Bagus Anugrah Ramadhan and Billy Muhamad Iqbal. 2018. User Experience Evaluation on the Cryptocurrency Website by Trust Aspect. In Proceedings of 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS). IEEE, Bangkok, Thailand, 274-279. https://doi.org/10.1109/ICIIBMS.2018.8550019

[63] Wahyu Nur Hidayat, Setiadi Cahyono Putro, Mukhamad Angga Gumilang, and Indriana Hidayah. 2018. What is Informatics Education Students Impression of Using Metacognitive Training System at The First Time?. In Proceedings of The 2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS). IEEE, Yogyakarta, Indonesia, 213-218. https://doi.org/10.1109/ICACSIS.2018.8618171

[64] Johanna Kaipio, Anne Kuusisto, Hannele Hyppönen, Tarja Heponiemi, and Tinja Lääveri. 2019. Physicians' and nurses' experiences on EHR usability: Comparison between the professional groups by employment sector and system brand. International Journal of Medical Informatics 134. https://doi.org/10.1016/j.ijmedinf.2019.104018.

[65] Manon Holman and Funmi Adebesin. 2019. Taking the Subjectivity out of UX Evaluation with Emotiv EPOC+. In Proceedings of the South African Institute of Computer Scientists and Information Technologists 2019 (SAICSIT '19). Association for Computing Machinery, New York, NY, USA, Article 30, 1–10. https://doi.org/10.1145/3351108.3351139

[66] Siti Nur Aisyiah Syahrir and Sfenrianto. 2019. User experience questioner and heuristics evaluation in online learning environment. Journal of Theoretical and Applied Information Technology 97, 4 (February 2019), 1071-1081.

[67] Abdul Samad Dahri and Ahmaed Al-Athwari and A. Hussain. 2019. Usability evaluation of mobile health application from AI perspective in rural areas of Pakistan. Int. J. Interact. Mob. Technol. 13, 11 (2019), 213-225. https://doi.org/10.3991/ijim.v13i11.11513

[68] Andrew Pringle, Stefanie Hutka, Jesse Mom, Robin van Esch, Niall Heffernan, and Paul Chen. 2019. Ethnographic study of a commercially available augmented reality HMD app for industry work instruction. In Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '19). Association for Computing Machinery, New York, NY, USA, 389–397. https://doi.org/10.1145/3316782.3322752

[69] Michael Carroll, Ethan Osborne, and Caglar Yildirim. 2019. Effects of VR gaming and game genre on player experience. In Proceedings of the 2019 IEEE Games, Entertainment, Media Conference (GEM). IEEE, New Haven, USA, 1-6. https://doi.org/10.1109/GEM.2019.8811554

[70] Davidavičienė, V., Raudeliūnienė, J. and Viršilaitė, R. 2019. User experience evaluation and creativity stimulation with augmented reality mobile applications. Creativity Studies. 12, 1 (Mar. 2019), 34-48. https://doi.org/10.3846/cs.2019.3576.

[71] Chen-Wei Chang, Shih-Ching Yeh, Mengtong Li, and Eason Yao. 2019. The Introduction of a Novel Virtual Reality Training System for Gynecology Learning and Its User Experience Research. IEEE Access, 7, 1 (2019), 43637 - 43653. https://doi.org/10.1109/ACCESS.2019.2905143

[72] Fahad Ahmed Satti, Jamil Hussain, H. S. Bilal, W. Khan, A. Khattak, Ju Eun Yeon, and S. Lee. 2019. Holistic User eXperience in Mobile Augmented Reality Using User eXperience Measurement Index. In Proceedings of The 2019 Conference on Next Generation Computing Applications (NextComp). IEEE, Mauritius, Mauritius, 1-6. https://doi.org/10.1109/NEXTCOMP.2019.8883528

[73] Denizhan Demirkol and Cagla Seneler. 2019. Evaluation of Student Information System (SIS) In Terms of User Emotion, Performance and Perceived Usability: A Turkish University Case (An Empirical Study). Procedia Computer Science 158 (2019), 1033-1051. https://doi.org/10.1016/j.procs.2019.09.145.

[74] Jonathan Moizer, Jonathan Lean, Elena Dell'Aquila, Paul Walsh, Alphonsus (Alfie) Keary, Deirdre O'Byrne, Andrea Di Ferdinando, Orazio Miglino, Ralf Friedrich, Roberta Asperges, and Luigia Simona Sica. 2019. An approach to evaluating the user experience of serious games. Computers & Education 136 (2019), 141-151. https://doi.org/10.1016/j.compedu.2019.04.006

[75] Najeebullah Soomro, Meraj Chhaya, Mariam Soomro, Naukhez Asif, Emily Saurman, David Lyle, and Ross Sanders. 2019. Design, Development, and Evaluation of an Injury Surveillance App for Cricket: Protocol and Qualitative Study. JMIR Mhealth Uhealth 7, 1 (22 January 2019). https://doi.org/10.2196/10978

[76] Fasola P. Abifarin, Shaka Apodoghe Imavah, and Ayodele S. Olobashola. 2019. Design effectiveness of academic library web sites A comparison of university, polytechnic, and college sites in Nigeria. The Electronic Library 37, 3 (Juny 2019), 577-591. https://doi.org/10.1108/EL-08-2018-0159

[77] Ryu Hyeongju and Kim Jeongeun. 2019. Evaluation of User Experience of New Defense Medical Information System. Healthcare Informatics Research 25, 2(April 2019), 73-81. https://doi.org/10.4258/hir.2019.25.2.73

[78] H. O'Brien and M. Lebow. 2013. Mixed-methods approach to measuring user experience in online news interactions. J. Assoc. Inf. Sci. Technol. 64, 1543-1556. https://doi.org/10.1002/asi.22871