

Evaluating the Acceptance and Quality of a Usability and UX Evaluation Technology Created for the Multi-Touch Context

Guilherme Eduardo Konopatzki Filho¹^a, Guilherme Corredato Guerino²^b
and Natasha Malveira Costa Valentim¹^c

¹Department of Computing, Federal University of Paraná, R. Evaristo F. Ferreira da Costa, Curitiba, Brazil

²Department of Computing, University of the State of Paraná,
Av. Minas Gerais, 5021 - Nucleo Hab. Adriano Correia, Apucarana, Brazil

Keywords: Multi-Touch, UX, Usability, Evaluation.

Abstract: The multi-touch context is a poorly explored field when it comes to usability and User eXperience (UX) evaluation. As any kind of system, it must be properly evaluated in order to be truly useful. A Systematic Mapping Study (SMS) showed that there is no technologies being used to evaluate the UX and usability of multi-touch systems that were specifically built for it. The use of generic technologies can leave behind important perceptions about the multi-touch systems specificities. To fill this gap, the User eXperience and Usability Multi-tiuch Evaluation Questionnaire (UXUMEQ) was created. UXUMEQ is a questionnaire that seeks to evaluate the UX and usability of multi-touch systems taking into account the most relevant aspects being used to this end, such as performance, workload, intuition, error tolerance and others. As any new technology, UXUMEQ must be evaluated in order to be improved. In this paper, we carried out a quantitative analysis to verify the public acceptance of UXUMEQ when compared with generic technologies being used to evaluated multi-touch systems. This analysis showed greater public acceptance about UXUMEQ regarding usefulness and ease of use. We also invited Human-Computer Interaction (HCI) experts to inspect UXUMEQ through a qualitative study. Their perceptions were collected and evaluated through the Grounded Theory method, that will contribute to provide a most refined version of UXUMEQ.


1 INTRODUCTION


When a system recognizes two or more touches at the same time, it can be considered a multi-touch system Lampert (1986). The quality of this kind of interaction needs to be studied and explored. The smartphones are the main representative devices that provides the multi-touch interaction. Composing a market that moved more than 1.4 billion units in 2021 (Statista Research Department, 2022) and its predicted to move more than US\$ 490 billion in 2026 (Market Data Forecast, 2022), its importance in our daily lives become evident. Multi-touch systems must have their software quality evaluated to be truly useful for their users. The usability and User eXperience (UX) are well-known criteria for this purpose because they can provide the necessary subsidy for


the software quality to be achieved Madan and Kumar (2012).

Usability is defined by ISO/IEC 25010 as “the degree to which users can use a product or system to achieve specific objectives to achieve specific goals effectively, efficiently, and satisfactorily in a specified context of use”. UX is defined by ISO 9241-210 as “perceptions and person’s responses that result from use and/or prior use of a product, system, or service.”

The technology concept used here is defined by Santos et al. 2012 as being a generalization for metrics, tools, methodologies, and techniques. Seeking to identify the technologies being used to evaluate the UX and usability, Guerino and Valentim found a lack of technologies being used to evaluate the multi-touch context, throughout a Systematic Mapping Study (SMS). This SMS revealed that only 11.76% of these technologies were evaluating the multi-touch systems. Seeking to characterize these technologies, Filho et al. 2022 carried out another SMS, discovering that the technologies being used

^a <https://orcid.org/0000-0002-7957-0684>

^b <https://orcid.org/0000-0002-4979-5831>

^c <https://orcid.org/0000-0002-6027-3452>

to evaluate multi-touch systems are generic (i.e., can be used to evaluate any kind of system). The scenario becomes worrying when realizing that generic tools may not extract the most credible result possible Blake (2011). This raises concerns about the influence of generic assessment technologies regarding interactions that are not so conventional to the user. Based mainly on this gap, the User eXperience and Usability Multi-touch Evaluation Questionnaire (UX-UMEQ) technology was proposed.

By seeking to refine UXUMEQ (the only known technology built for this context) through quantitative and qualitative studies, we can achieve a more significant maturation of the technology and a valuable expansion of evidence-based knowledge. This contribution can and should lead to a greater understanding of user behavior and needs in this context, in addition to expanding the state of the art in the necessary but poorly explored niche of multi-touch. Furthermore, more mature and refined usability and UX evaluation technology can contribute to better software quality reaching the hands of the users.

The remainder of this paper is organized as follows: Section 2 contains studies that served as basis of this work; Section 3 contains a presentation about UXUMEQ; Section 4 presents the methodology used in the two studies; Section 5 contains the quantitative analysis regarding the user acceptance about UXUMEQ; Section 6 presents the experts perceptions regarding the UXUMEQ content validity; Section 7 contains the discussion linking the results and Section 8 presents the final considerations and future work.

2 RELATED WORK

Considering the limitations of the generic usability and UX evaluation technologies found in the SMS of Filho et al. 2022, we looked for usability and UX evaluation technologies built considering the multi-touch context. Three works presented technologies considered built for this context. Ghomi et al. 2013 presented a study about a multi-touch input technique for learning chords and a recognizer and guidelines for building chord vocabularies. The experiment consisted of a reproduction of multi-touch gestures presented on screen by 12 participants. The usability and UX aspects of understandability and comfort were evaluated in a questionnaire with a 5-point Likert scale that was not made available for consultation. In other phase, 24 participants tried to perform the demonstrated chord. The data collection for this second experiment was a system log used to calculate success, help, and recall ratios. Both methods ex-

tracted quantitative data.

Martin-SanJose et al. 2017 presented a questionnaire built to the multi-touch context, using a 5-point Likert scale, extracting quantitative data. They used it to evaluate the UX aspect of motivation regarding students using a tabletop application designed to visualize and manipulate the European banknote monetary system. The questionnaire was available and contained questions such as "I find it enjoyable to study the monetary system at the table" and "For me, it was easy to learn the different euro notes and coins".

In the study of Hachet et al. 2011, 16 participants were invited to manipulate 3D objects in a virtual reality environment. After that, they answered a 5-point Likert scale questionnaire that contained questions such as "I felt sick or tired", "I understood the depth well", and "I needed to move my head". This questionnaire evaluated the general feeling, interaction, and manipulation with visual elements, approaching both UX and usability criteria quantitatively.

The questionnaires proposed by Ghomi et al. 2013, Martin-SanJose et al. 2017 and Hachet et al. 2011 are authorial, i.e., they were created to evaluate UX and usability of the specific multi-touch systems of their studies. We noted that these three works collected only quantitative data. We also verified a lack of specific UX and usability technologies for the multi-touch context. Moreover, there is a lack of joint extraction of quantitative and qualitative data, that corroborates with the findings of Guerino and Valentim 2020. These findings demonstrate that only 29.69% of the technologies used to evaluate usability and UX of NUIs do this evaluation jointly. In addition, the above-cited evaluation technologies were not empirically evaluated before use. This lack of validation in the evaluation technologies, mainly in the authorial ones, leads to less reliable results Shull et al. (2001). To improve the reliability of our results, we present in this paper two studies to validate UXUMEQ content.

3 UXUMEQ

To fill the gaps presented in Section 2, UXUMEQ was created and its development process is already published Konopatzki et al. (2023). It is a questionnaire for usability and UX evaluation of multi-touch systems. UXUMEQ was built considering the main aspects used in the context of multi-touch evaluation and contains phrases that can better direct the evaluators' gaze through more specific problems of multi-touch systems. UXUMEQ contains 28 questions, 13 about UX and 15 about usability. There are 21 aspects

Table 1: UXUMEQ aspects.

Usability aspects	UX aspects
Performance	Fun
Ease of use	General Feeling
Efficiency	Comfort
Effectiveness	Innovation
Workload	Intuition
Ease of Learning	Tension
Ease of Remembering	Control
Response Time	Immersion
Satisfaction	Concentration
Usefulness	Distraction
Error Tolerance	

covered in UXUMEQ, being presented in the Table 1.

UXUMEQ presents two ways to provide feedback: an input field to describe the problems encountered, thus gathering qualitative data, and a 5-point semantic differential scale for each question to collect quantitative data. It is suggested to read UXUMEQ once before interacting with the multi-touch system that the user seeks to evaluate. This is due to a question that asks to count how many attempts to finish a task the user took, so it's better to know in advance that there is a necessity to count it. The interaction following this first reading will enable the user to use UXUMEQ to evaluate it. It can be used individually, but it is interesting to use it with groups of users testing an application, as comparisons and ratios can be made with quantitative results and more qualitative responses can lead to a more complete understanding of multi-touch system weaknesses. Some examples of UXUMEQ questions can be found in Figures 1 and 2.

USABILITY	
Performance	
1. Do you consider your performance on the multi-touch system task to be good or bad?	
Too bad	I didn't like it at all
	1 2 3 4 5
	I really liked
Describe the performance issues you noticed	
Ease of use	
2. Was the multi-touch system easy to use? That is, did you reach your goal without major difficulties?	
Very difficult	I had no fun
	1 2 3 4 5
	I had a lot of fun
Describe the difficulties you noticed when using the multi-touch interface	

Figure 1: Two of the 14 questions of UXUMEQ regarding usability.

4 METHODOLOGY

In search of building an evidence-based path, the methodology that most represents this research proposal is presented by Shull et al. (2001), which presents feasibility, observation and case studies to carry out the evaluation of a technology from its

UX	
General Feeling	
15. Did you enjoy using the multi-touch interface?	
I didn't like it at all	I really liked
	1 2 3 4 5
Describe why you didn't like using the interface	
Fun	
16. Did you have fun using the multi-touch interface?	
I had no fun	I had a lot of fun
	1 2 3 4 5
Describe the issues that got in the way of your enjoyment	

Figure 2: Two of the 13 questions of UXUMEQ regarding UX.

proposition to its transfer to be applied in industry. To improve and validate UXUMEQ, two studies were carried out, a feasibility study with students and a qualitative study with experts in Human-Computer Interaction (HCI).

4.1 Feasibility Study with Students

Motivation. In order to choose a technology to compare to UXUMEQ, we analyzed all questionnaires found in the SMS presented by Filho et al. 2022, and priority was given to those that address usability and UX together. However, none of the questionnaires that met these requirements were selected because they were not considered suitable.

Thus, priority was given to finding two questionnaires to compose the comparative set, one that only evaluates usability and the other UX. Among the questionnaires that specifically evaluate usability, SUS Brooke (1996) was chosen because it is the most used in SMS Filho et al. (2022) and because it is widely used in several other studies in the literature. The questionnaire considered most appropriate to represent the UX criteria was the INTUI Ullrich and Diefenbach (2010), for the number of questions (UXUMEQ = 28 questions vs. SUS + INTUI = 27 sentences/questions) and the UX aspects addressed.

Goal. The goal of this study, following the Goal-Question-Metric (GQM) paradigm Basili and Rombach (1988), is to analyse UXUMEQ, to evaluate it, concerning its acceptance, from the point of view of HCI and Software Quality students, in the context of usability and UX evaluation in multi-touch systems.

Hypothesis. The study was planned and conducted in order to test the following hypotheses (null and alternative, respectively):

- H01 - There is no difference between the ease of use of UXUMEQ and SUS+INTUI;
- HA1 - There is difference between the ease of use of UXUMEQ and SUS+INTUI;
- H02 - There is no difference between the perceived usefulness of UXUMEQ and SUS+INTUI;

USABILITY/SUS			
1. I think that I would like to use this system frequently.			
	1 2 3 4 5		
2. I found the system unnecessarily complex.			
	1 2 3 4 5		
3. I found the system was easy to use.			
	1 2 3 4 5		
UX/INTUI			
While using the product...			
11	I acted deliberately	1 2 3 4 5 6 7	I acted on impulse
12	I tried very hard to reach my goal	1 2 3 4 5 6 7	I reached my goal effortlessly
13	I performed unconsciously, without reflecting on individual steps	1 2 3 4 5 6 7	I consciously made a step after another

Figure 3: Examples of SUS and INTUI questions.

- HA2 - There is difference between the perceived usefulness of UXUMEQ and SUS+INTUI;
- H03 - There is no difference between the future use intention of UXUMEQ and SUS+INTUI;
- HA3 - There is difference between the future use intention of UXUMEQ and SUS+INTUI;

Context. We carried out the comparative study with undergraduate students at UXUMEQ. They were attending the classes of "Software Quality" and "Human-Computer Interaction". This study was accepted by the Research Ethics Committee of UXUMEQ.

Variable Selection. The dependent variables selected were the TAM indicators Davis (1989), ease of use, perceived usefulness and future use intention. The independent variable was the usability and UX evaluation technology type (UXUMEQ and SUS+INTUI).

Selection of Participants. Forty-seven participants signed a consent form and filled out a characterization form measuring their expertise in usability, UX evaluation, and multi-touch systems. Among them, we have 44 men and 3 women.

Experimental Design. Participants were divided into two groups (group UXUMEQ and group SUS+INTUI), to evaluate the same application, Google Earth¹, with the same instructions for use. We chose Google Earth because it was one of the apps that presented a great variety of multi-touch gestures. We characterized the two groups from their experience with UX and Usability evaluation and from their familiarity with multi-touch systems. This study follows a design between groups.

Instrumentation. Several artifacts were defined to support the study: characterization and consent forms, the UXUMEQ and SUS+INTUI themselves, instructions for the evaluation and the post-evaluation

¹<https://earth.google.com>

questionnaire TAM Davis (1989). These instruments can be found in a Figshare repository ².

Preparation. The participants received two-hour training on usability and UX evaluation. For each group, we made a 15-min presentation about the evaluation technology that the group would apply.

Execution. At the study's beginning, a researcher was responsible for passing the information from the evaluation to the participants. They were then divided into two groups for each technique. First, each participant received the artifacts described previously. After the evaluation, they delivered the post-evaluation questionnaire filled out. We had 24 participants used the UXUMEQ technology, and 23 used the SUS+INTUI technology.

4.2 Qualitative Study with Experts

Goal. The goal of this study, following the Goal-Question-Metric (GQM) Basili and Rombach (1988) paradigm, is to evaluate UXUMEQ, to improve it, concerning its content validity, from the point of view of SE and HCI experts, in the context of usability and UX evaluation of multi-touch systems.

Context. The minimum requirement to be considered an expert was to have a complete master's degree in SE and/or HCI areas. This study was accepted by the Research Ethics Committee of [anonymous institution]. Invites were sent to experts through e-mails containing a short research presentation and a contextualization of the study. Eleven SE and HCI experts from Brazil accepted the invite, six men and five women.

Preparation. The experts passed through a first online meeting where the study proposal was fully presented. At this initial meeting, the expected role to carefully and critically evaluate UXUMEQ was presented. The evaluation was asynchronous, within a period of two weeks. They received the UXUMEQ in a digital .docx file and consent and characterization forms.

Characterization. The participants filled and signed the characterization and consent forms. The characterization can be seen in Table 2.

Execution. After two weeks, a second meeting was conducted to interview the experts and collect perceptions about the UXUMEQ validity. Then, we conducted a semi-structured interview with nine scripted questions. The speeches of the interview were recorded, transcribed and imported into the Atlas.ti.v9 program, where the two first phases of the Grounded Theory method Corbin and Strauss (2014) were carried out. These data were then analyzed

²<https://encurtador.com.br/sDKN7>

Table 2: Experts Characterization.

Expert	UXUAP	UXUIP	MTEX
P01	2	0	High
P02	1	0	High
P03	5	2	High
P04	20	10	High
P05	0	10	High
P06	4	1	High
P07	1	0	High
P08	5	20	Medium
P09	3	0	High
P10	15	8	High
P11	10	1	High
UXUAP - UX and Usability Academic Projects quantity; UXUIP - UX and Usability Industrial Projects quantity; MTEX - Multi-Touch EXperience;			

and brought insights to make improvements on UXU-MEQ.

5 ANALYSIS OF USER ACCEPTANCE

The post-evaluation questionnaire using Technology Acceptance Model (TAM) Davis (1989) filled by the participants was analyzed. This questionnaire was built to evaluate the general acceptance of UXU-MEQ and SUS+INTUI. TAM defines three indicators: (i) perceived ease of use, defined as the degree to which a person believes that using a specific technology would be effortless, (ii) perceived usefulness, as the degree to which a person believes that the technology could improve his/her performance at work; and (iii) future use intention, which assesses users' intention to use the technology again in the future. The possible answers are: totally agree (bold green), partially agree (light green), I do not agree nor disagree (yellow), partially disagree (light red), and totally disagree (bold red). Their answers can be seen in Figures 4 and 5. In that questionnaire, the participants answered their degree of agreement with the TAM statements regarding the ease of use, usefulness, and future use intention of UXU-MEQ.

The mean answers to the TAM affirmations were compared, and the results are presented in Table 3. As can be seen, all the means in the UXU-MEQ group were higher than the means in the SUS+INTUI group. To verify the statistical relevance of these results, the Shapiro-Wilk normality test was applied, with $\alpha = 0.05$, which revealed that none of the samples presented a normal data distribution. The Mann-Whitney

Table 3: Comparison between the UXU-MEQ and SUS+INTUI groups.

TAM Question	Group	Mean	Mann-Whitney U test
E1	SUS+INTUI	4,09	0.055
	UXU-MEQ	4,54	
E2	SUS+INTUI	3,96	0.279
	UXU-MEQ	4,25	
E3	SUS+INTUI	4,39	0.046
	UXU-MEQ	4,71	
E4	SUS+INTUI	3,48	0.007
	UXU-MEQ	4,08	
U1	SUS+INTUI	3,13	0.002
	UXU-MEQ	4,08	
U2	SUS+INTUI	3,35	0.144
	UXU-MEQ	3,79	
U3	SUS+INTUI	3,39	0.040
	UXU-MEQ	3,96	
U4	SUS+INTUI	3,74	0.020
	UXU-MEQ	4,33	
I1	SUS+INTUI	2,91	0.052
	UXU-MEQ	3,54	
I2	SUS+INTUI	3,09	0.203
	UXU-MEQ	3,5	

statistical significance test was then carried out, also with $\alpha = 0.05$, where it was found that questions E3, E4, U1, U3 and U4 had significant statistical relevance. This lead us to conclude that the HA1, about ease of use, was partially accepted (E3 and E4 showed statistical significance). The second alternative hypothesis, HA2, regarding perceived ease of use, was almost fully accepted (U1, U3 and U4 showed statistical significance). And the H03 hypothesis was confirmed, since there is no statistical significance in the I1 and I2.

6 QUALITATIVE RESULTS

Findings on Specific Issues. The experts provided several insights that will help to guide the decisions of necessary changes. Some of these perceptions involved doubt about the meaning of performance (see quote from P10 below) or even the perception of rep-

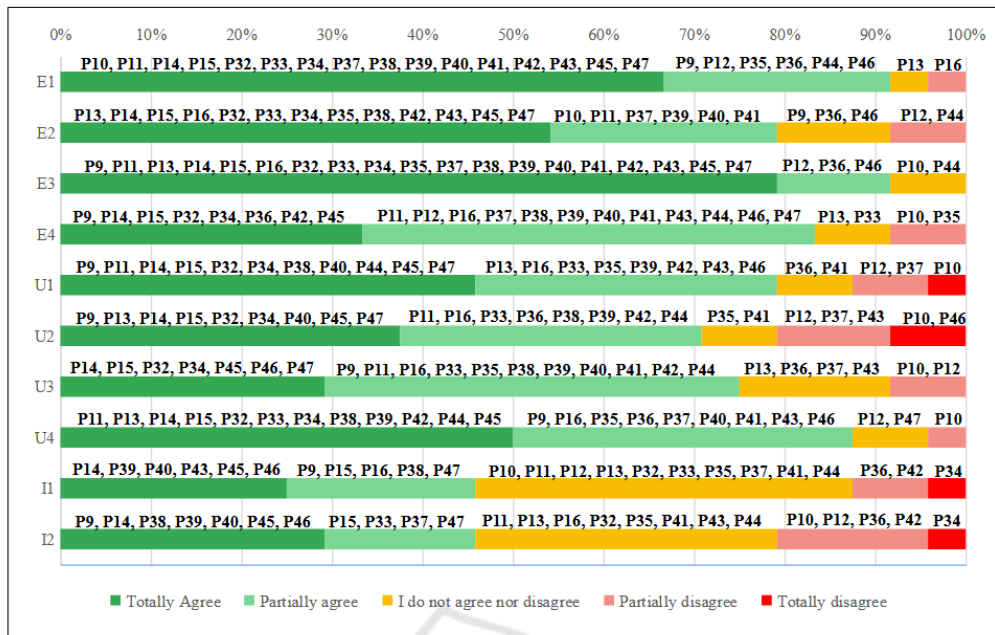


Figure 4: Results of users acceptance from UXUMEQ group using TAM questionnaire.

etition between the aspects of performance, effectiveness and efficiency (see quote from P06 below). These points will lead us to add examples and explanations of what these terms mean.

"If the focus is on end users, will they know what performance means?". (P10)

"For example, the question of performance, effectiveness and efficiency, what is the difference? This can become confusing (...)". (P06)

The question involving efficiency received praise in relation to its clarity (see quote from P01 below), some notes about its lack of breadth (see quote from P03a below) and subjectivity (see quote from P07) and also notes about the aspect of efficiency overlaps with issues of workload (see quote from P03b below). The overlap was not considered a problem, since its a way to better evaluate the aspects, and the subjectivity will be addressed with examples and explanations.

"For example, there are others here where efficiency is clearer this way". (P01)

"Efficiency is not just time, right? Efficiency is time, effort, is resources that are used to evaluate". (P03a)

"I think that "time you consider viable" is something very personal". (P07)

"I think that efficiency has other dimensions than just time. I think that down there you evaluate the workload. But then there is an overlap between efficiency and workload, right?". (P03b)

Questions 25 and 26, about immersion, included doubts about what immersion would be and how ex-

amples of involvement through the senses can confuse rather than help the user (see quote from P03 below). An expert noticed that the text that indicates what should be described in the qualitative field was repeated in both questions (see quote from P02 below). The point most cited by experts was the strangeness in the expression "feeling part of the experiment", as it does not clearly express the opposite of being an observer, as it does not lead the user to reflect on their experience and because it has a confusing and difficult to understand (see quote from P08 below). It was considered that these difficulties could lead the user to not answer the question, in addition to being considered a question not applicable in many situations (see quotes from P04 below). As the other questions, these will be reformulated as well to correct the pointed issues.

"I don't know what immersion is. Is it a very involved hearing, a very involved vision? I think this question doesn't assess immersion and I don't even know what to answer with it". (P03)

"Here, within immersion, the descriptions are the same, right? If you noticed, and then it's the same thing in the second one, I think that even if they want to, they won't answer both". (P02)

"I find the term "part of the experiment" strange. The sentence has to lead the user to reflect on their experience of use." (P08)

"(...) possibly not applicable to many situations". (P04)

Q27, which deals with concentration, had some

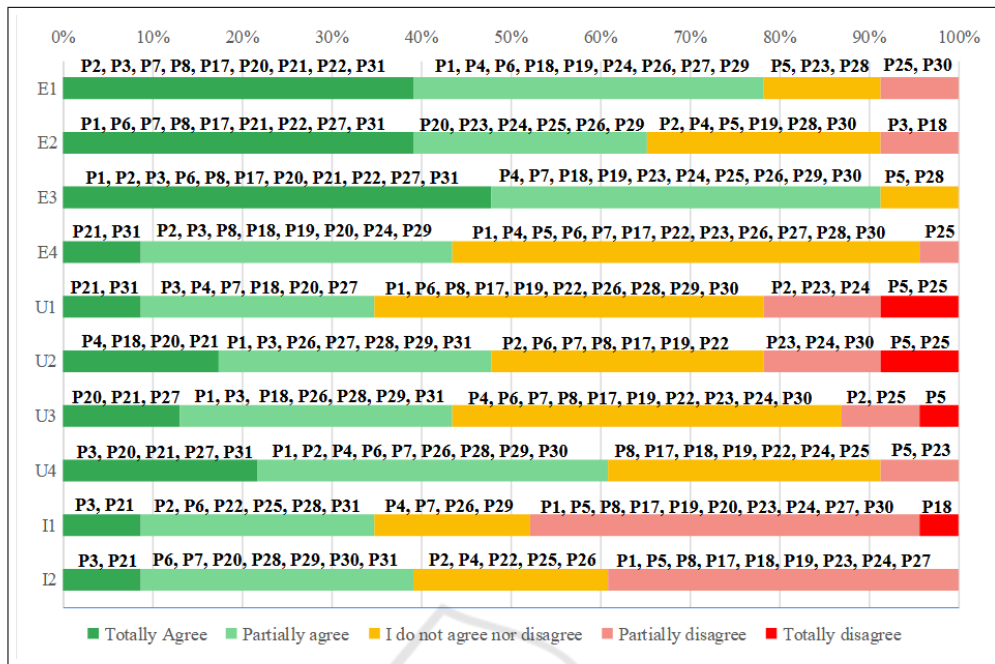


Figure 5: Results of users acceptance from SUS+INTUI group using TAM questionnaire.

notes regarding its need when compared to Q28 which deals with distraction (see quote from P01 below). The difference between focusing on the necessary activities and the mechanisms to carry out such activities was not clear (see quote from P08 below). We considered important to keep both questions of concentration and distraction, since they allow a wider gathering of answers, and the question will be reformulated to make the difference between the activities and the mechanisms to carry them out.

"Then I had the same doubt, does it make sense to have the two and what would be the difference, you know?". (P01)

"What's the difference between focusing on "the necessary activities" and "the mechanisms used to carry out these activities"?". (P08)

As for distraction, some of the main points raised were the difficulty of interpreting and answering the question (see quote from P08 below) and the confusion about attention capture having a good or bad nature, in addition to the doubt to what this attention capture refers to (see quotes from P03a below). The inadequacy of the term "automatic" was also noted (see quote from P11 below), in addition to the confusion about how attention capture relates to the aspect of distraction (see quote from P03b below).

"I found it a difficult question to interpret and answer". (P08)

"Capturing attention in relation to distractions or in relation to the interface?". (P03a)

"That term doesn't seem appropriate here". (P11)

"I couldn't understand what question 28 is for. Isn't distraction being measured? Why are you asking about attention capture?". (P03b)

Corrections to Consider. An expert pointed that the acronym for UX was not defined in the document, as well as the acronym for the UXUMEQ, besides declaring that specific jargons can make the understanding difficult or impossible (see quotes from P03a, P03b and P03c below). To fix this, the acronyms will be make explicit, and the jargons will be changed by simpler words. A simplification of many aspects will be applied, following the statement of P10 (see P10 quote below).

"The acronym UX is not defined (User eXperience)". (P03a)

"It was not clear how the acronym UXUMEQ was created (from which letters of which words it was created)". (P03b)

"Eh and other jargons that are used can make it impossible to directly understand the content of the dimension". (P03c)

"This technique has to be much simpler for the users, okay?". (P10)

Observations. Some observations were made, that can help to understand better the peculiarities of UX-UMEQ. The possibility to customize the technique was noted (see quote from P03a below). The neutral point in scales is a common discussion regarding psychometrics, and an expert told that he doesn't see

a problem with the neutral point here (see P09 quote below). The perception that questionnaires are not focused on qualitative evaluation was perceived

"And that is a very good result. It's cool that using a very generic technique, I can customize it for different types of context and bring specific characteristics that are specific to that context". (P03a)

"but I think there's no problem with this issue of having the neutral point". (P09)

"I don't see the questionnaire as a technique which is focused on qualitative". (P03b)

"So I think the questionnaire is useful for providing an indication of UX in a quantitative way". (P03c)

Positive Findings. Some positive points were noted, as the not tiring aspect of UXUMEQ (see P06 quote below), the user's guidance to find problems (see P11a quote below), the usefulness that a well-based tool brings (see P01 quote below), the possibility of reducing costs when using it (see P11b quote below) and the well-covered aspects (see P07 quote below). The ease to tabulate and process data was pointed (see P11c quote below), as the simplicity and objectivity of the questions (see P05 quote below) and the adequate number of questions (see P09 quote below).

"It doesn't seem to be very tiring to use". (P06)

"(...) guides you through what problems could occur when using the multi-touch system". (P11a)

"Ah, these categories of performance, ease of use, efficiency. So you have a foundation, you know? Huh So I think he's useful for that. He knows? There's a basis there.". (P01)

"By having a checklist, you can greatly reduce the cost of a project.". (P11b)

"It covers well several aspects". (P07)

"It's an even more practical way for you to tabulate this data and process this data later". (P11c)

"the questions be simple, right? They are objective questions.". (P05)

"I think there just aren't that many questions, right? I didn't find a lot of questions, I don't think twenty-eight is that big, right?". (P09)

Scope of UXUMEQ. The UXUMEQ scope was considered good about the usability aspects covered (see P06 quote below), in this way being considered very complete (see P07 quote below). This coverage was considered enough to the P09 to work with it (see P09 quote below).

"(...) it raises a lot of questions related to usability". (P06)

"(...) I think it's very complete". (P07)

"Yes [I would work with UXUMEQ], because he is covering many aspects here.". (P09)

Ease of Use. The ease of use regarding the division

into categories, the agility to fill, the practicality and the interpret and read part were raised (see quotes from P02, P09, P08 and P05 below).

"(...) the categories are fundamental, right? So much for being able to separate what is well evaluated and what is not. As much as to facilitate the logical flow, right?". (P02)

"Using this scale is for me, even though it's twenty-eight, it helps you, it's agility when filling it out too". (P09)

"You have the possibility of using it in print, right? Very practical". (P08)

"There's not much to say, they're easy to interpret, easy to read". (P05)

Notes on Semantic Differential Scale. A semantic differential scale was used to gather the quantitative data, and the perception that it was aligned with the questions was brought, as well as its importance to reduce the confusion (see P09 and P06 quotes below). By the other way, the presence of a neutral point was considered a potential source of noise in the data, as well as the bias generated when asking just about negative points in the questions (see quotes from P04 and P10 below). We did not consider the neutral point as being a problem worth of modification, since not having a neutral point is also a problem, but the questions will be reformulated to ask about positive points also.

"For me they are in line with the question". (P09)

"But the Likert scale can be confusing sometimes, right? So I think it's important to have the statements". (P06)

"Not that it's a problem with the structure of the questionnaire, but I think this could be a source of noise". (P04)

"So this is a huge bias, okay? Just have questions that are all negative on one side and positive on the other". (P10)

Explicit Division of Questions into Categories and Aspects. The questions and criteria of UXUMEQ are divided by labels, and were thought to help the user. The experts were divided in those who did not see a need to have it (see quotes from P08 and P05 below) and those that considered it helpful (see quotes from P01, P08 and P03 below). Since no problems were highlighted, but rather a non-necessity, we do not intend to make changes in relation to this aspect.

"for the independent user, okay? I don't think there would be a need". (P08)

"Hey, I think it would be simpler to not have that extra information that I have to somehow interpret". (P05)

"For me it's interesting to be divided, right, so for me it makes more sense to be that way". (P01)

"For me, for those of us who work with usability and UX, it makes it easier, right?". (P08)

"the categories are fundamental". (P03)

Focus on Multi-Touch. Several comments were made regarding the multi-touch aspect. An expert reported that he could not identify the specific metrics used in the UXUMEQ, and some experts found the questions very generic (see quotes from P10a and P03a below). The generic questions comes from the generic technologies from which the UXUMEQ was based. In order to improve UXUMEQ, questions regarding multi-touch gestures will be added.

Some experts pointed several issues of understanding, such as the meaning of performance, the meaning of multi-touch systems and concluded that maybe a familiarity with the scientific terms can be necessary to truly understand the UXUMEQ (see quotes from P03b, P03c and P07 below). To mitigate these points, we propose to add the meaning of performance in the question, as well to present the meaning of "multi-touch interface" and other specific jargon.

Another expert pointed that there are occasions where multi-user systems use multi-touch, with the specificity to verify if the gesture is being linked with the right user (see P08 quote below). As UXUMEQ is a modular questionnaire, we consider that would be a good advance to add a question regarding this issue.

"I can't identify what the indicators actually are, what the specific metrics were for these types of interface, okay". (P10a)

"I keep thinking that if I were to use it, removing the word multi-touch for any other type of application, would it change? Maybe not. Maybe not". (P03a)

"I don't know what performance means in a multi-touch system. Is it able to play? Is it having a 1x1 relationship for touch and action performed? I feel that this dimension is very comprehensive and abstract, without being able to really address what will be evaluated". (P03b)

"I feel like, to use the questionnaire, familiarity with the scientific literature on multi-touch interfaces/systems is necessary". (P03c)

"Isn't it worth putting some sentence on what multi-touch systems are?". (P07)

"I also think, for example, of those larger systems that allow user collaboration and the system has to identify who is making the gesture when you have more than one hand there, right? Two-handed using, for example, one, right? Using it, is the system recognizing it properly, right? When there are multi-users". (P08)

Intention to Use UXUMEQ by Participants. Some participants stated that they would use UXUMEQ (see P03 quote below), mainly in the design stage

when developing a software (see P11 quote below). Some participants stated that they would use UXUMEQ with conditions, such as access to the documentation, if it were shorter and if it had some more adjustments (see P04, P05 and P02 quotes below).

"Yes. As you bring this in a systematic way, right? Already with questions and easy to apply. It's great to be able to apply this now". (P03)

"[I would use it], probably. It is a very objective tool and this makes it much easier to adopt a legal project. Mainly in the previous stage, in the design stage". (P11)

"if I had access to the manual for this questionnaire, it would indicate the validity of the evidence collected, the reliability estimated, yes, I would use it". (P04)

"If it were shorter, and if it were more focused on having more ergonomic issues". (P05)

"I would use it, but then it would have to make some more adjustments". (P02)

Incongruence Between Question Format and Scale. Some experts perceived a clash between the questions format and the scale being used. The main issue was some questions that could be answered with yes or no, contrasting with the semantic scales goal, that is differentiate the scales extremes through opposing words (see quotes from P02 and P10 below). These points provide a subsidy to further modifications in order to adjust the type of answers.

"The questions shouldn't be: determine the degree of ease of use of the multi-touch system?". (P02)

"Often the type of response does not make it possible to understand the Likert thing". (P10)

Projections About Public Use. A general perception that UXUMEQ would be easy to use by people with experience in UX evaluation was found (see P03 quote below). In the same way, an expert pointed that users with no experience could have some difficulty using it (see quote from P01 below).

"A person who has already evaluated the UX in some way will have no difficulty with the questionnaire". (P03)

"I don't know if it would be so clear for a user with no experience". (P01)

7 DISCUSSION

The evaluation of the first study allowed us to understand, through TAM's answers, that there is greater public acceptance for UXUMEQ than for the generic SUS+INTUI technologies. This acceptance focuses on the terms of perceived usefulness and ease of use. All means of the Likert scales present in the TAM sen-

tences were higher for the UXUMEQ group than for the SUS+INTUI group. In five of these sentences, statistical significance was proven. It is interesting to note that none of the statistically significant averages pertain to UXUMEQ's intended future use. We assume this is due to the very narrow niche of evaluating multi-touch interfaces. The greater acceptance regarding ease of use and perceived usefulness of UXUMEQ reinforces some perceptions obtained through the literature, which served as motivation for its creation. These perceptions involve the idea that generic technologies fail to consider the specificities of certain contexts. It also involves the perception that there are few technologies extracting quantitative and qualitative data. In this way, greater public acceptance of UXUMEQ reinforces the idea that a technology built specifically for a context can perform a better assessment.

With the aim of understanding directly from experts what the points for improvement would be, the second study was carried out and brought enlightening results. The way in which various characteristics of UXUMEQ are presented have been questioned. From inconsistencies between the format of the scale and the question, to the organization of questions by categories and aspects. The most relevant group of notes was certainly those related to multi-touch, which allowed the emergence of insights considered important and unique for the improvement of UXUMEQ. This importance is due to the scarcity of literature on multi-touch. In this way, the perception of experts about this field becomes a source that provides greater scope for improvements and concepts to be worked on.

8 FINAL CONSIDERATIONS AND FUTURE WORK

This paper presented two studies carried out to better understand the weaknesses and strengths of UXUMEQ, in an attempt to improve it and make it a truly useful technology for evaluating usability and UX in the multi-touch context. Through a feasibility study, the superiority of UXUMEQ was demonstrated in terms of ease of use and usefulness perceived by the public, when compared to generic technologies. A qualitative study was also carried out with experts, which provided the necessary basis to understand the points that should be improved.

Below are presented some possible future perspectives for this research:

- Feasibility study: a new feasibility study with the aim of verifying the user acceptance, effectiveness

and efficiency after the generation of the 3rd version of UXUMEQ;

- Creation of an analysis tool: development of a tool that can support the analysis of data collected by UXUMEQ, with the aim of further assisting more the work of researchers and developers;
- Expansion of criteria: carrying out research to include aspects such as accessibility and communication at UXUMEQ;

ACKNOWLEDGMENTS

We thank the funding and support of the Coordination for the Improvement of Higher Education Personnel (CAPES) - Program of Academic Excellence (PROEX).

REFERENCES

- Basili, V. R. and Rombach, H. D. (1988). Towards a comprehensive framework for reuse: A reuse-enabling software evolution environment. In *NASA, Goddard Space Flight Center, Proceedings of the Thirteenth Annual Software Engineering Workshop*, number UMIACS-TR-88-92.
- Blake, J. (2011). *Natural user interfaces in. NET: WPF 4, Surface 2, and Kinect*. Manning.
- Brooke, J. (1996). Sus: A "quick and dirty" usability scale. In JORDAN, P. W. and *et al.*, editors, *Usability evaluation in industry*, pages 1–7, London. Taylor&Francis.
- Corbin, J. and Strauss, A. (2014). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage publications.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pages 319–340.
- Department, S. R. (2022). Global smartphone sales to end users since 2007. <https://www.statista.com/statistics/263437/global-smartphone-sales-to-end-users-since-2007/>.
- Filho, G. E. K., Guerino, G. C., and Valentim, N. M. (2022). A systematic mapping study on usability and user experience evaluation of multi-touch systems. In *Proceedings of the 21st Brazilian Symposium on Human Factors in Computing Systems*, pages 1–12.
- Forecast, M. D. (2022). Smartphone market size and growth. <https://www.marketdataforecast.com/market-reports/smartphone-market>.
- Ghomi, E., Huot, S., Bau, O., Beaudouin-Lafon, M., and Mackay, W. E. (2013). Arpège: learning multitouch chord gestures vocabularies. In *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces*, pages 209–218.

- Guerino, G. C. and Valentim, N. M. C. (2020). Usability and user experience evaluation of natural user interfaces: a systematic mapping study. *IET Software*, 14(5):451–467.
- Hachet, M., Bossavit, B., Cohé, A., and de la Rivière, J.-B. (2011). Toucheo: multitouch and stereo combined in a seamless workspace. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 587–592.
- ISO 9241-210 (2019). *Ergonomics of Human System Interaction - Part 210: Human-Centered Design for Interactive Systems*. International Organization for Standardization.
- ISO/IEC 25010 (2011). *Systems and Software Engineering - SquaRE - Software product Quality Requirements and Evaluation: System and Software Quality Models*. International Organization for Standardization.
- Konopatzki, G. E., Guerino, G., and Valentim, N. (2023). Proposal and preliminary evaluation of a usability and ux multi-touch evaluation technology. In *Proceedings of the XIX Brazilian Symposium on Information Systems*, pages 317–324.
- Lampert, L. A. (1986). The gnats and gnus document preparation system. *G-Animal's Journal*, 41(7):73+.
- Madan, A. and Kumar, S. (2012). Usability evaluation methods: a literature review. *International Journal of Engineering Science and Technology*, 4.
- Martin-SanJose, J.-F., Juan, M.-C., Mollá, R., and Vivó, R. (2017). Advanced displays and natural user interfaces to support learning. *Interactive Learning Environments*, 25(1):17–34.
- Santos, G., Rocha, A. R., Conte, T., Barcellos, M. P., and Prikladnicki, R. (2012). Strategic alignment between academy and industry: a virtuous cycle to promote innovation in technology. In *2012 26th Brazilian Symposium on Software Engineering*, pages 196–200. IEEE.
- Shull, F., Carver, J., and Travassos, G. H. (2001). An empirical methodology for introducing software processes. *ACM SIGSOFT Software Engineering Notes*, 26(5):288–296.
- Ullrich, D. and Diefenbach, S. (2010). From magical experience to effortlessness: an exploration of the components of intuitive interaction. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, pages 801–804.