

# An Empirical Study to Evaluate the Feasibility of a UX and Usability Inspection Technique for Mobile Applications

Ingrid Nascimento, Williamson Silva, Adriana Lopes, Luis Rivero, Bruno Gadelha, Elaine Oliveira, Tayana Conte  
Federal University of Amazonas (UFAM)  
Manaus, Brazil  
{inc, williamson.silva, adriana, luisrivero, bruno, elaine, tayana}@icomp.ufam.edu.br

**Abstract** — Usability and UX (User eXperience) are some of the most important factors for evaluating the quality of mobile applications. They focus on how easy to use an application is and the emotions that such use evokes. However, these aspects are often evaluated separately in industry through different evaluation techniques. Although it is possible to identify more usability and UX problems by employing different UX and usability evaluation methods, this distributed approach may not be cost effective and may not allow to thoroughly explore the identified issues. In order to support the identification of both UX and usability problems in a single evaluation, we have proposed Userbility, an UX and usability inspection technique that allows evaluating these aspects in mobile applications. This paper presents an empirical study over the second version of Userbility to verify its feasibility. In this study, we compared Userbility with the UX and Usability Guidelines Approach (UUGA) that helps the evaluation of usability and UX separately in mobile applications. According to the quantitative results, considering efficiency, UUGA was better than the Userbility technique. However, the qualitative results suggest that Userbility pointed more improvement suggestions, which could be useful for redesigning the evaluated application.

**Keywords**- *User experience; usability; mobile applications; evaluation; inspection; empirical study.*

## I. INTRODUCTION

In the past few years, the development of mobile applications has grown considerably due to growth of the mobile devices market [1]. Therefore, to help the quality of the developed mobile applications, it is necessary to carry out specific evaluations considering their dynamism and the aspects that make it difficult to evaluate them (e.g mobile context, connectivity, size of the screen and others) [2].

Usability is one of the factors for the adoption of mobile applications [2]. According to the ISO 9241-210 norm [3], usability is the extent to which “*a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a context of use*”. Moreover, another factor related to quality in use is User eXperience (UX). UX considers the quality of an application regarding the feelings of the user when interacting with it. This term denotes the overall experience of the user of an application [4], and UX is crucial for the adoption of an application with success [4]. In this sense, while usability focuses on the ease of carrying out tasks and overall satisfaction, UX focuses on aspects related to

the emotions, perceptions and judgements of an application. Therefore, software development teams willing to increase the quality in use of the developed mobile applications need to evaluate both of them.

To evaluate usability and UX together, in our previous work [13], we developed Userbility in order to support inspectors in the evaluation of both UX and usability in mobile applications at the same time, to assess whether Userbility can support inspectors in detecting usability and UX problems. Nascimento et al. [13] conducted a study with five mobile applications. The results showed that it is possible to identify improvements in applications, and allowed us to identify problems during the use of the technique. Based on this, in this paper, we proposed a new version of the technique and an empirical study to evaluate the feasibility of Userbility. We have compared the Userbility to an approach proposed by De Paula et al. [5], which evaluates UX and usability separately.

The remainder of this paper is organized as follows. Section II presents a background on UX and usability evaluation techniques that can be applied to evaluate mobile applications. Then, Section III shows the Userbility technique in its second version. Section IV presents the empirical study where we compared Userbility with another evaluation approach. In Section V, we present the results of the empirical study. Finally, Section VI presents our conclusions and future work.

## II. BACKGROUND

Traditional methods for evaluating usability and UX may not be adequate for evaluating mobile applications since mobile devices offer a new paradigm for Human Computer-Interaction (HCI) that these traditional evaluation methods do not consider [2]. Therefore, specific methods for the evaluation of mobile devices may allow providing more beneficial data for mobile applications than those of traditional evaluation [2]. The following are some methods for such purpose.

Von Wangenheim et al. [6] proposed the Checklist Heuristic Evaluation for Smartphones Applications. This checklist has 12 heuristics for evaluating applications. Each heuristic has usability items that allow measuring the traditional usability of mobile applications.

Another approach is the Expressing Emotions and Experiences (3E), which is a method that aims to capture the experience and feelings of the user [7]. In 3E, the user is

provided with a simple pictorial template for expressing emotions and experiences.

Wetzinger et al. [8] proposed a process for evaluating UX and usability comparing the use in tablets and laptops. This process was composed of the following approaches: the SUS (Software Usability Scale) questionnaire, to measure usability through scales; the AttrakDiff to measure UX; and the SEQ (Single Ease Question), to measure ease of use.

De Paula et al. [5] suggested an approach that uses guidelines to evaluate UX and usability of mobile application by employing Design Thinking, called UX and Usability Guidelines Approach (UUGA). This approach is based on: (a) Usability guidelines by Gong and Tarasewich [9], guidelines by Shneiderman [10], the ISO 9241-210 [3] and Guidelines by Nielsen [11]; and (b) UX guidelines, based on the questionnaire by Chen and Zhu [12], where scores are given to each criterion that influences the experience of the user.

Regarding the methods, Von Wangenheim et al. [6] focuses only in usability evaluation, while 3E method focuses in UX evaluation. Also, the process proposed by Wetzinger et al. [8] did not carry out the inspection of an application; it only uses SUS for measuring usability through a five-point scale, ranging from totally agree to totally disagree for about 10 questions. De Paula et al. [5] suggested the use of an approach to evaluate UX and usability separately in mobile applications. In this approach, the identified UX and usability problems are not directly associated, which could generate repetition of information for the analysis and could not associate the problems to the possible improvements captured with UX evaluation.

### III. THE USERBILITY TECHNIQUE

Userbility v 1.0 (Integrating **User** eXperience and **Usability**) was proposed to allow the evaluation of the UX and usability of mobile applications by unexperienced inspectors in the area of HCI [13]. We improved the Userbility v 1.0 based on the methods Checklist Heuristic Evaluation for Smartphones Applications [6] and 3E [7]. We chose these methods as they provide effective means to evaluate specific usability attributes of mobile applications [6] and allow gathering rich emotional responses from users, thus complementing others methods. The integration of these methods aims to make it easier and faster to evaluate UX for practitioners that are not experts in carrying out UX and usability evaluation. Figure 1 shows an example of Userbility in its second version.

Item 1 (Usability Heuristic) shows one of the twelve Usability Heuristics for mobile applications that were included to support the inspection process. We followed the suggestions by Von Wangenheim et al. [6], including usability items for the heuristics so that they could guide the inspectors in finding usability problems. Item 2 (Usability items and an Examples) presents the Usability items and an example for each item. Items 3 (UX Q1) and 4 (UX Q2) present two UX questions with an example for each of them. These questions were included considering the 3E method, in order to support the evaluator when reporting his/her experience for each aspect.

The two UX questions are: (Q1) “How did you react towards the application regarding this aspect? Why did you react that way?”, and (Q2) “What do you think or would improve about this aspect? Point where it should be improved in the application”. Item 5 (Satisfaction Item) shows an item related to the degree of satisfaction of the evaluator. This item was adapted from the Self-Assessment Manikin [14] and is composed of a 5-point scale (very unsatisfied, unsatisfied, neither satisfied nor unsatisfied, satisfied and very satisfied), where a human face depicting an emotion represents each item. With this item, we intend to obtain more information about how much the evaluator was satisfied regarding the evaluated usability heuristic. Figure 1 also shows an example of application and a part of inspection, in Problems (P1 and P2) and Improvements box.

The figure shows a form for evaluating a usability heuristic. It is organized into five main sections, each with a corresponding label on the right side:

- 1. Usability Heuristic:** A1. Visibility of System Status. It contains three items (I1, I2, I3) with Yes/No checkboxes. I1: "Does the application show the system state, feedbacks and instant alert messages?" (Yes/No). I2: "Do the interactive components such as buttons, menus and selection boxes that were selected are highlighted when compared to the others?" (Yes/No). I3: "Do the slower operations offer feedback?" (Yes/No).
- 2. Usability Items and Examples:** This section provides context for the items, such as "Ex.: after a task such as delete user, there is a confirmation."
- 3. UX Q1:** "How did you react towards the application regarding this aspect? Why did you react that way?" with an example: "I felt lost after registering because I had no feedback if I managed to finish it."
- 4. UX Q2:** "What do you think or would improve about this aspect? Point where it should be improved in the app?" with an example: "I believe that this aspect is not adequate. I would include feedback messages for the registration functionality."
- 5. Satisfaction Item:** A 5-point scale from "Very Unsatisfied" to "Very Satisfied" using human face icons.

Below the satisfaction scale is an **Application Summary** for "Water Your Body" with identified **Problems** (P1: "The app does not return feedback to start.", P2: "The goal to drink water is not clear.") and **Improvements** ("An interactive tutorial that teaches to use the app.").

Figure 1. Example using one of the heuristics of Userbility.

### IV. THE EMPIRICAL STUDY

We conducted the empirical study in order to compare Userbility v 2.0 with the UX and Usability Guidelines Approach (UUGA) [5], to verify to what extent it could improve the performance of the evaluators in terms of identified problems, time and improvements suggestions to the problems. We chose UUGA due to the following reasons: the evaluator does not need any experience to use the method and it also allows identifying and describing usability problems.

#### A. Planning

During the planning stage, we defined the hypotheses of the experiment, its context, the selection of subjects and their training in the techniques, and the tasks to be carried out.

1) **Hypotheses:** the experiment was conducted to test the following hypotheses:  $H_{01}$  – There is no difference in terms of efficiency regarding Userbility and UUGA;  $H_{A1}$  – There is a difference between the efficiency regarding Userbility and UUGA;  $H_{02}$  – There is no significant difference in terms of effectiveness regarding Userbility and UUGA;  $H_{A2}$  – There is a significant difference between the effectiveness regarding Userbility and UUGA. In this sense, effectiveness is defined as “the percentage of identified defects divided by total (known) defects”, and efficiency is defined as “the total number of defects found by an inspector divided by the total time spent to find them” [15].

2) **Context and Subjects:** We conducted the study at Federal University of Amazonas (Brazil) using both approaches (Userbility and UUGA) to evaluate a mobile application called “In the Tip of the Tongue” (Na Ponta da Língua – in Portuguese). “In the Tip of the Tongue” is an application that aims to help students when they are learning Portuguese, by describing the origin of the words and helping users fix their meaning through a simple game. At all, 49 students from two courses on Human Computer-Interaction (HCI) and Collaborative Systems agreed to participate in the study as evaluators. All subjects signed a consent form to participate in the study and filled out a characterization questionnaire to measure their knowledge in HCI (KH) and their knowledge in software Analysis and Design (KA). The characterization form was employed to categorize the subjects as having: (a) No Experience – does not have knowledge in HCI or possesses some notions on usability acquired through lectures/ readings, but without practical experience; (b) Low – participated in at least one usability evaluation/ project in class; (c) Medium – participated in 1 to 4 usability evaluation/ projects in industry; (d) High – participated in 5 usability project/ evaluations in industry. The expertise on KA was classified following the same standards. From this characterization, we divided the subjects into two groups to carry out the inspection: the group from the Userbility was formed of 25 subjects and the UUGA by 24 subjects. The group of UUGA had one more subject with high knowledge degree in software analysis and design when compared to the Userbility group (see Table I – third and fourth columns).

3) **Training:** We trained the subjects introducing the concepts and examples of Usability and UX. In addition, they had exercises in which they had to give their impressions on the usability and UX of one application of their choice. After balancing the subjects, we randomly distributed them into two groups where each group received equivalent training on each of the techniques that they would apply.

4) **Tasks definition:** Each of the evaluators had to carry out six main tasks during his/her evaluation of the application. The tasks were: (a) To train, answering at least three training questions; (b) To start a game, choose the correct word; (c) To view the final score in the game; (d) To search for a word in the dictionary; (e) To view the graph of the languages; and (f) To find information about questions of Portuguese language.

We selected the tasks based on their importance within the application to achieve user goals.

### B. Execution

Following their characterization, the subjects then received a tablet or cell phone to evaluate the selected application using one of the two techniques. When starting the evaluation, each subject received an introduction to the evaluated application and the list of tasks they would employ to evaluate its usability and UX individually. Each group carried out the evaluation in separate rooms with a researcher who supervised all subjects. After the evaluation, we grouped the discrepancies (possible usability problems indicated by an inspector) that were reported by each of the techniques in a single list. In this list, we removed the duplicated discrepancies from the same inspector before the discrimination of the discrepancies. After that, the list of discrepancies was reviewed by another researcher related to this study to avoid misclassification.

### C. Analysis

We carried out a discrimination meeting to classify the discrepancies into real problems and false positives (discrepancies which were not problems affecting the usability and the UX). The client (owner of the application), a UX and usability expert and three other researchers participated in this meeting, which lasted for 1 hour and 34 minutes.

## V. RESULTS OF THE EMPIRICAL STUDY

We obtained the quantitative data from measuring the time, number of real problems and false positives for each subject, and the qualitative data from analyzing the description of the problems found in the Userbility technique inspection questionnaire.

### A. Quantitative Analysis

The results from the discrimination meeting have been used to calculate the effectiveness and efficiency of the subjects. Table I shows the results per subject and per technique in terms of discrepancies, identified real problems and time spent to find them. In Table I, we can see that subjects who used Userbility managed to find between 1 and 11 defects spending about 0.47 and 2 hours. On the other hand, the subjects that used UUGA employed between 0.22 and 1.05 hours, however they also found between 1 and 11 defects. The subjects took more time using Userbility according to the applied statistical tests. Table II shows the results per technique in terms of identified problems, unique defects and suggested improvements. The analysis of these results shows that UUGA allowed finding more real defects (NRD). However, among the identified real problems, Userbility allowed identifying more unique defects (total of 37 unique defects - NUD) than UUGA (NUD is 31). Also, we noticed that Userbility allowed pointing out less false positives. This is an important feature, as accurate techniques (those in which the chances of a discrepancy become a real problem is high) are important for identifying more problems.

To test the hypotheses defined in Subsection IV-A, we carried out a statistical analysis using the SPSS tool ( $\alpha = 0.05$ ). We used the Shapiro-Wilk test to check the normality. This test is indicated for sample with size less than 50 [16]. We verified

that efficiency ( $p = 0.06$  for the Userbility and  $p = 0.08$  for the UUGA) and effectiveness ( $p = 0.72$  for the Userbility and  $p = 0.36$  for the UUGA) were normally distributed. Based on the results, we applied the *t-test* [17]. Figure 2 shows the boxplots graphs to facilitate the visualization of the results.

TABLE I. SUMMARY OF INSPECTION RESULT PER SUBJECT.

T	#S	KH	KA	ND	NFP	NRD	Ti(h)	Effic.	Effec.(%)
Userbility	S01	N	N	3	0	3	1.42	2.12	4.41
	S02	N	N	2	0	2	0.83	2.40	2.94
	S03	N	N	5	0	5	2.00	2.50	7.35
	S04	L	N	5	1	4	0.60	6.67	5.88
	S05	L	N	14	3	11	1.80	6.11	16.18
	S06	L	N	4	2	2	0.83	2.40	2.94
	S07	L	L	10	1	9	1.50	6.00	13.24
	S08	L	L	3	2	1	0.75	1.33	1.47
	S09	L	L	4	0	4	0.92	4.36	5.88
	S10	L	L	5	1	4	0.67	6.00	5.88
	S11	L	L	9	4	5	1.28	3.90	7.35
	S12	L	L	7	1	6	0.92	6.55	8.82
	S13	L	L	7	0	7	0.80	8.75	10.29
	S14	L	L	6	3	3	0.83	3.60	4.41
	S15	L	L	6	1	5	1.15	4.35	7.35
	S16	M	L	3	0	3	0.78	3.83	4.41
	S17	M	L	2	0	2	0.50	4.00	2.94
	S18	M	M	4	1	3	-	-	4.41
	S19	M	M	4	0	4	1.00	4.00	5.88
	S20	L	L	6	1	5	0.53	9.38	7.35
	S21	L	N	4	0	4	1.65	2.42	5.88
	S22	L	L	4	0	4	0.50	8.00	5.88
	S23	N	L	7	0	7	1.00	7.00	10.29
	S24	L	M	6	1	5	-	-	7.35
	S25	H	H	8	0	8	0.47	17.14	11.76
S26	N	N	3	0	3	0.55	5.45	4.41	
S27	N	N	7	3	4	0.50	8.00	5.88	
S28	L	N	3	0	3	0.77	3.91	4.41	
S29	L	N	5	1	4	0.52	7.74	5.88	
S30	L	N	5	2	3	0.67	4.50	4.41	
S31	L	N	4	1	3	0.65	4.62	4.41	
S32	L	L	8	3	5	0.78	6.38	7.35	
S33	L	L	11	2	9	1.05	8.57	13.24	
S34	L	L	8	2	6	0.60	10.00	8.82	
S35	L	L	5	1	4	0.53	7.50	5.88	
S36	L	L	7	1	6	0.43	13.85	8.82	
S37	L	L	4	1	3	0.75	4.00	4.41	
S38	L	L	9	1	8	0.97	8.28	11.76	
S39	L	L	4	0	4	0.52	7.74	5.88	
S40	L	L	3	1	2	0.77	2.61	2.94	
S41	M	L	3	2	1	0.58	1.71	1.47	
S42	L	M	7	1	6	0.45	13.33	8.82	
S43	M	M	10	0	10	0.90	11.11	14.71	
S44	M	H	5	2	3	0.22	13.85	4.41	
S45	L	L	3	1	2	0.43	4.62	2.94	
S46	N	N	2	0	2	0.85	2.35	2.94	
S47	N	L	5	0	5	0.23	21.43	7.35	
S48	L	L	12	1	11	0.47	23.57	16.18	
S49	H	H	10	3	7	0.77	9.13	10.29	

**Footnote - T:** Technique; **#S:** Subject; **KH:** Knowledge in HCI; **KA:** Knowledge in software Analysis and Design; **H:** High; **M:** Medium; **L:** Low; **N:** None; **ND:** Number of Discrepancies; **NFP:** Number of False Positives; **NRD:** Number of Defects; **Ti(h):** Time; **Effic.:** Defects per Hour; **Effec.(%):** Effectiveness.

Figure 2 (Efficiency) presents the boxplot graphs comparing the efficiency indicator. When analyzing the median of both groups, we can see that the group applying UUGA had higher efficiency than the group that applied the Userbility. The p-value ( $0.009$ )  $< 0.05$  rejects  $H_{01}$  and supports  $H_{A1}$ , which indicates that evaluators using UUGA are more efficient than those applying Userbility. A possible cause for this result may

be the number of items that evaluators had to fill in, when applying Userbility.

TABLE II. ANALYSIS OF THE DISCREPANCIES PER TECHNIQUE.

Technique	ND	NFP	NRD	NUD	SI
Userbility	138	22	116	37	190
UUGA	143	29	114	31	11
<b>Total</b>	<b>246</b>	<b>45</b>	<b>201</b>	<b>68</b>	<b>201</b>

**Footnote - UUGA:** UX and Usability Guidelines Approach; **ND:** Number of Discrepancies; **NFP:** Number of False Positives; **NRD:** Number of Defects; **NUD:** Number of Unique Defects; **SI:** Improvements Suggested.

Figure 2 (Effectiveness) presents the boxplot graph comparing the effectiveness indicator. When analyzing the medians, we can see that the effectiveness of the Userbility group was almost the same as the one of the group that applied UUGA. The p-value ( $0.886$ )  $< 0.05$  supports  $H_{02}$ , which indicates that is not possible to make statements about effectiveness of Userbility and UUGA. It is important to note that Userbility allows finding more unique (not repeated) defects and can support identifying improvements based on the perceptions of the subjects, which is an important feedback for the correction of problems and redesigning the evaluated application. The subjects S25 and S48, that had high and low skill level, were more efficient for the approaches. And the subjects S05 and S48, that had low and none skill level, were more effective for the approaches. This indicates that the evaluator's skills levels may not influence in the effectiveness.

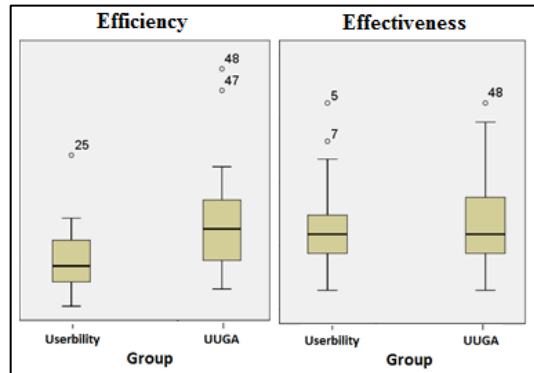


Figure 2. Boxplot graphs for the efficiency and effectiveness indicators.

### B. Qualitative Analysis

The analysis of the qualitative data began with the examination of the answers within the Userbility technique questionnaire. The number of improvements suggested to the Userbility (190) was much higher than the number of improvements suggested to the UUGA (11), showed in the Table II. The improvements suggested by the subjects were related to the problems identified in the application. The improvements suggestions to some problems were:

The application stopped without warning – “The app should have a message about what happened...” (S04);

Difficulties in understanding the symbols – “They could add a help button about the features of the app” (S07);

Many steps to perform the task to find information about questions (task f in the Section IV) – “The app could have a review option about the performance of some tasks...” (S07);

The graph didn't provide any relevant information for the application – *When we touch the screen and we choose Greek language, for instance, there could be some fun facts about such language.*” (S13);

More feedback to the user in the game – *“There could be some way to alert that the bubbles would appear and how to play, since I failed to notice that I should choose the correct word.”* (S18).

These improvements suggestions were useful for redesigning the application and improving the quality of the application. And they also could be useful to improve the UX, in order to generate the successful adoption of the application.

## VI. CONCLUDING REMARKS AND FUTURE WORK

This paper described and evaluated the second version of Userbility, a technique to integrate and facilitate UX and usability inspections in mobile applications. In our empirical study, we compared the Userbility with the UX and Usability Guidelines Approach (UUGA). Among the results, the statistical analysis suggested that Userbility was less efficient than UUGA. Nevertheless, when analyzing the results from Table II, we noticed that Userbility allowed inspectors to point out less false positives. This is an important feature as the degree of accuracy of a technique can avoid inspectors to waste their time in indicating problems not affecting the quality of the evaluated application. In addition, Userbility allowed inspectors to point out more unique problems and improvements suggestions, which can be useful for redesigning the evaluated application. Furthermore, this empirical evaluation showed evidence of its feasibility.

We considered also the threats to validity in this empirical study. The main threats to validity were: (1) training effects: we controlled this risk by providing equivalent training for both approaches; (2) students are probably not good substitutes for professional inspectors: since we were looking for novice inspectors with no knowledge on the use of the applied techniques; (3) academic environments do not represent day to day experience in the industry: we carried out the evaluation over a real mobile application which can help resemble a real industry environment; (4) homogeneity of the sample: there is a limitation regarding our study, even though the study was carried out in classes from different courses; (5) the subjects did not use the same device: not using the same device for inspecting the application could cause bias to our results, but we needed to identify problems in both devices to verify the correctness of the application in different platforms.

As future work, we intend to carry out improvements in the Userbility technique to enhance its effectiveness and efficiency when compared to other inspection approaches. Also, we intend to replicate this study to draw further conclusions on the feasibility and applicability of Userbility in industrial environments.

## ACKNOWLEDGMENT

We thank the financial support granted by CAPES process 175956/2013 and FAPEAM through processes numbers: 062.00600/2014; 062.00578/2014. Furthermore, we would like to thank all the volunteers who participated in the study. We also thank Natasha Valentim for her remarks on this research.

## REFERENCES

- [1] IDC, Smartphone OS Market Share, Q1 2015. Available at: <http://www.idc.com/prodserv/smartphone-os-market-share.jsp>, 2015.
- [2] D. Zhang and B. Adipat, “Challenges, methodologies, and issues in the usability testing of mobile applications,” in *International Journal of Human-Computer Interaction*, vol. 18, n. 3, pp. 293-308, 2005.
- [3] International Organization for Standardization, ISO FDIS 9241-210: Ergonomics of human system interaction – Part 210: Human-centered design for interactive systems (formerly known as 13407). Switzerland, 2009.
- [4] A. Charland and B. Leroux, “Mobile application development: web vs. native,” in *Communications of the ACM*, vol.54, n.5, pp. 49-53, 2011.
- [5] D. F. De Paula, B. H. Menezes and C. C. Araújo, “Building a Quality Mobile Application: A User-Centered Study Focusing on Design Thinking, User Experience and Usability,” in *Part of HCI International 2014. Design, User Experience, and Usability. User Experience Design for Diverse Interaction Platforms and Environments*, vol. 8518, pp. 313-322, 2014.
- [6] C. G. Von Wangenheim, T. A. Witt, A. F. Borgatto, J. V. Nunes, T. C. Lacerda, C. Krone and L. de Oliveira Souza, “A Usability Score for Mobile Phone Applications Based on Heuristics,” *International Journal of Mobile Human Computer Interaction*, vol. 8, n. 1, pp. 23-58, 2016.
- [7] M. Tähti and M. Niemelä, “3e-expressing emotions and experiences,” in *Workshop on Innovative Approaches for Evaluating Affective Systems. HUMAINE*, 2006, pp. 15-19.
- [8] W. Wetzlinger, A. Auinger and M. Dörfinger, “Comparing effectiveness, efficiency, ease of use, usability and user experience when using tablets and laptops” in *Part of HCI International 2014. Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience*, vol. 8517, pp. 402-412, 2014.
- [9] J. Gong and P. Tarasewich, “Guidelines for handheld mobile device interface design,” in *Annual Meeting, Northeastern University, Boston*, 2004, pp. 3751-3756.
- [10] B. Shneiderman, “Designing the User Interface - Strategies for Effective Human-Computer Interaction,” Addison-Wesley, 1998.
- [11] J. Nielsen, “Usability inspection methods,” in *Conference Companion on Human factors in computing systems*, 1994, pp. 413-414.
- [12] Z. Chen and S. Zhu, “The Research of Mobile Application User Experience and Assessment Model,” in *International Conference on Computer Science and Network Technology*, vol. 4, pp. 2832-2835, 2011.
- [13] I. Nascimento, W. Silva, B. Gadelha, T. Conte, “Userbility: A Technique for the Evaluation of User Experience and Usability on Mobile Applications,” in *18th International Conference on Human-Computer Interaction*, in press.
- [14] M. M. Bradley and P. J. Lang, “Measuring emotion: the self-assessment manikin and the semantic differential,” *Journal of behavior therapy and experimental psychiatry*, vol. 25, n. 1, pp. 49-59, 1994.
- [15] A. Fernandez, S. Abrahão and E. Insfran, “Towards the validation of a usability evaluation method for model-driven web development,” in *International Symposium on Empirical Software Engineering and Measurement*, 2010, pp. 54-57.
- [16] S. Shapiro and M. Wilk, “An Analysis of Variance Test for Normality (Complete Samples),” *Biometrika*, vol. 52, pp. 591-611, 1965.
- [17] N. Juristo and A. M. Moreno, *Basics of Software Engineering Experimentation*. Boston, MA: Kluwer Academic Publishers, 2001.