Proc. of the Seventh Intl. Conf.on Advances in Computing, Communication and Information Technology - CCIT 2018

ISBN: 978-1-63248-162-7 DOI: 10.15224/ 978-1-63248-162-7-05

THINKING ALOUD METHOD TO EVALUATE THE USABILITY OF A NOWCASTING APP

Regina de Moraes França, Leandro Guarino de Vasconcelos,
Luiz Augusto Toledo Machado, Luiz Eduardo Guarino de Vasconcelos

***Abstract*— Brazilians have great interest in weather forecasting, be it a trip plan, physical activity, the clothes that will go in the day or just for an elevator talk. There are several institutions and applications in Brazil and in the world that carry out the weather forecast, one of them being the Center for Weather Forecasting and Climate Research (CPTEC/INPE). One of these applications is the SOS CHUVA, developed by CPTEC, which shows images of radar and satellite in order to assist society in decision-making. The main goal of this work is to show the usability assessment performed in the SOS CHUVA app, in order to understand what aspects and user interfaces need improvement, in order to provide an even better product to society. For this, a usability assessment was made using the Thinking Aloud methods and the heuristic evaluation with the 10 Nielsen heuristics. Results are considered satisfactory showing the need for in-app adjustments.**

Keywords*— *evaluation of usability, thinking aloud, heuristic evaluation, weather forecasting, application

I. Introduction

Brazilians have great interest in weather forecasting, be it a trip plan, physical activity, the clothes that will go in the day or just for an elevator talk. In addition to citizens interested in weather forecasting, there is interest in the various business sectors, such as agriculture, industry, transportation, energy generation and transmission, commerce, tourism, education, among others. There are several institutions in Brazil and in the world that carry out the weather forecast, one of them being the Center for Weather Forecasting and Climate Research (CPTEC) of the National Institute of Space Research (INPE). Having the competitive differential is fundamental to fully fulfill the mission of the CPTEC, which aims at the benefit of society in general. A new science related to the weather forecast, is the nowcasting, i.e. immediate weather forecast, for the next few hours.

Regina de Moraes França, Leandro Guarino de Vasconcelos, Luiz Augusto Toledo Machado, Luiz Eduardo Guarino de Vasconcelos
Instituto Nacional de Pesquisas Espaciais (INPE/CPTEC)
Brazil

Regina de Moraes França, Leandro Guarino de Vasconcelos, Luiz Eduardo Guarino de Vasconcelos
FATEC Guaratinguetá
Brazil

According to Machado and Martins [1], nowcasting are widely used by several segments of society. Flood control in urban areas, monitoring of risk areas during storms, information on weather conditions on highways, monitoring of precipitation in certain regions, information on fishing activities, promotion of sporting events on weather conditions, information to companies' managers of electricity and telephone networks, among others, are examples of the use of nowcasting forecasts.

There are a number of solutions (i.e. applications or app) available in the market that provide the weather forecast. One of these applications is the SOS CHUVA, developed by CPTEC, which shows images of radar and satellite in order to assist society in decision-making.

This app, available on Android and iOS, has already had more than 106,000 downloads and more than 600 comments in the official stores (i.e. Play Store and Apple Store). For any app, the big challenge is to design screens with features that people can understand and use without difficulty. This is related to UX (User eXperience) and usability.

The main goal of this work is to show the usability assessment performed in the SOS CHUVA app, in order to understand what aspects and user interfaces need improvement, in order to provide an even better product to society. For this, a usability assessment was made using the Thinking Aloud [2, 3] methods and the heuristic evaluation with the 10 Nielsen heuristics [2]. Results are considered satisfactory showing the need for in-app adjustments.

This paper is organized as follows. Section 1 is Introduction. Section 2 presents the main concepts discussed in this paper. Section 3 presents the methodology used in conducting the usability assessment with the participants. Section 4 shows the main results obtained. Finally, the conclusions, the suggestions for future work and the acknowledgments are made.

II. Background

Some surveys [4] indicate that 25% of installed mobile applications are opened and used only once because of the user's poor experience of using the app. People want solutions that are easy to use, have quality and accuracy.

Faced with a more complex scenario that includes much more demanding and well-informed customers, in addition to increasingly better competitors, the only certainty is that if we do not change and adapt to the demands of our customers (ie society), the more difficult it will be we remain competitive and viable operationally and economically in the long run.

Proc. of the Seventh Intl. Conf.on Advances in Computing, Communication and Information Technology - CCIT 2018

ISBN: 978-1-63248-162-7 DOI: 10.15224/ 978-1-63248-162-7-05

In addition, an application (i.e. software or app) without any usability process will bring many problems. To ensure that a product can function properly without user confusion, it is necessary to perform a usability assessment.

Usability assessment is defined as a search technique used to evaluate a product or service. The tests are performed with users representative of the target audience. Each participant tries to perform typical tasks while the analyst observes, hears and notes [5].

A task is a sequence of steps that must be done in the application. The comparison between the sequence of events performed by the user in the execution of a task and the sequence of events defined by the evaluator is able to indicate any usability problems. Among the works that use this approach can be highlighted [6, 7].

The term usability was first used in the early 1980s with the primary goal of providing guidance for product developers to develop user-friendly solutions. Currently, we use ISO 9241-11 [8] that defines usability as "the ability of a product to be used by specific users to achieve specific goals with effectiveness, efficiency and satisfaction in a specific context of use." In addition, we use this standard to quantify the efficiency and effectiveness of an application.

According to [8], effectiveness is the precision and completeness with which users achieve specific goals, accessing the correct information or generating the expected results; efficiency is the precision and completeness with which users achieve their goals, in relation to the amount of resources spent; and satisfaction is the comfort and acceptability of the product, measured by subjective and / or objective methods.

The efficiency [9] is obtained by equation 1:

$$Overall\ relative\ efficiency = \frac{\sum_{j=1}^{R}\sum_{i=1}^{N} n_{ij}t_{ij}}{\sum_{j=1}^{R}\sum_{i=1}^{N} t_{ij}} * 100 \quad (1)$$

Where:

• N is the total of tasks;

• R is the number of participants;

• n_{ij} is the result of task i by user j; when the user successfully completes the task, then n_{ij} is equal to 1, otherwise nij equals 0.

• t_{ij} is j user time spent to complete task i. When the task does not complete successfully, the time is measured until the user exits the task.

The effectiveness [9] is obtained by dividing the total number of tasks successfully completed by the total number of tasks performed, multiplying by 100 (Equation 2).

$$Effectiveness = \frac{total\ number\ of\ tasks\ completed\ successfully}{total\ number\ of\ task\ undertaken} * 100 \quad (2)$$

Satisfaction can be obtained by applying the SUS (System Usability Scale) questionnaire [10]. The SUS is a simple and reliable tool used to quickly measure how people perceive the usability of applications. A SUS score above 68 would be considered above average and anything below 68 would be below average, but the best way to interpret its results involves the "normalization" of scores to produce a percentile ranking [11, 12].

The heuristic evaluation method was proposed by Nielsen in 1994 [13], which is a heuristic-guided inspection, which indicates the general principles of good interface design, aimed at maximizing the usability of the artifact. Some experts in the field of human factors usually conduct a heuristic assessment. Together, experts discuss their findings; establish the most familiar and serious problems; and make suggestions for problem solving. The evaluations are based on a set of usability heuristics defined by Nielsen [14, 15, and 16] for user interface design.

Heuristics are general rules that one or more experts apply to evaluate the usability of a software or application.

The Thinking Aloud [2, 3] method requires 5 to 8 participants [17] who should carry out the test verbalizing their thoughts on how they decide to interact in the application screens.

III. Methodology

Initially, 16 tasks were defined so that they could be performed by the participants. For each task were defined the appropriate steps that each participant should carry out until the completion of the task.

To participate in the usability assessment, we selected people who already had knowledge in technology; that were not of the area of meteorology; who had experience using mobile devices and were already users of the SOS Chuva app. Users of the relationship network of the authors of this article were selected. An identification questionnaire was developed to define the user profile (i.e. name, age, gender, schooling); to identify user experience in the use of applications in general and with other applications that interact with maps; and what features you usually use of the SOS Chuva.

Each participant was invited, through direct (verbal) contact, by the authors of the work to participate in the usability evaluation of the application in order to improve the SOS Chuva app. All participants were excited to contribute to the research. With each, a schedule and a date were scheduled according to the participant's availability.

In the day and time combined, it was better explained how the evaluation would be performed, making it clear that what was evaluated was the application and not the participant. Data collection took place in a room where only the experts and the participant stayed. Each task to be performed by the participant was recorded and this was explained previously for each participant. The recordings occurred with the permission of the participants, who understood their need for data collection. The recording was made with the participant sitting with his or her smartphone in the hands performing the tasks that were passed while the analyst was recording, focusing only on the participant's smartphone.

Proc. of the Seventh Intl. Conf.on Advances in Computing, Communication and Information Technology - CCIT 2018

ISBN: 978-1-63248-162-7 DOI: 10.15224/ 978-1-63248-162-7-05

At the time of data collection, each participant underwent the following procedure:

1. The importance of this evaluation, the Thinking Aloud method [3] and what they should do throughout the test was explained;

2. They were asked to complete the identification questionnaire;

3. The tasks in Table 1 were read, one by one, so that the participant could execute them; if the participant could not complete the task, the next task was initiated;

4. They were asked to complete the SUS post-test questionnaire.

5. We thank the participant and the importance of their contribution.

6. Explanation by the expert of tasks that were not successfully completed.

After completing the evaluation with the participant, the data was tabulated. We counted the steps and the time spent in each of them to complete each task.

It is worth mentioning that after the identification questionnaire, the tasks and the post-test questionnaire were elaborated, the procedure was performed with two people who already knew the application, in order to validate the questions and the tasks. Some adjustments to the task statement were needed. These two people were not counted in the results of the usability evaluation, presented in the next section.

IV. Results

In this study, 2 experts did the analysis and identified 43 usability issues in the application.

Next, the Thinking Aloud method was used, which was applied with 8 people between July 25 and August 1, 2018. To fill out the identification questionnaire, the time taken was 5 to 10 minutes. Sixteen tasks were proposed, with the average execution time for all tasks being 462 seconds. The tasks were recorded and the recordings lasted between 6 and 19 minutes. Finally, the participant should respond to a SUS questionnaire [7] in order to measure overall satisfaction with the use of the SOS Chuva app. For this, the time it took was 5 to 10 minutes.

Participants identified 34 usability issues in the application, representing 79% of all usability issues encountered by participants and experts. Through this method it is possible to discover approximately 80% of the usability problems of an application.

The effectiveness obtained at the conclusion of the tasks was of 64.8% and the efficiency was of 61.07%. The mean SUS score [10] was 80 points.

The results obtained after the participants perform the task are complete task (or not) and the execution time. These results are used to calculate effectiveness and efficiency. The data for calculating efficacy can be seen in table I. P1 to P8 are the participants. In case the task has been completed, the cell is numbered 1. Otherwise, the cell is 0.

The total time taken to complete each task considering the attempts until they gave up completing the task was 3696 seconds. The following table II is a table of execution time of completeness task. Cells with 0 show the tasks that were not completed.

TABLE I. EFFECTIVENESS IN THE CONCLUSION OF TASKS

Tasks	Task-completeness (1/0)								QP	Effectiveness
	P1	P2	P3	P4	P5	P6	P7	P8		
1	1	1	1	1	1	1	0	1	7	87.5
2	1	1	1	1	0	0	0	1	5	62.5
3	1	1	0	0	0	0	0	1	3	37.5
4	1	1	1	1	1	0	1	1	7	87.5
5	1	1	1	0	0	1	1	0	5	62.5
6	1	1	1	1	1	1	0	1	7	87.5
7	1	1	1	1	1	1	1	1	8	100.0
8	1	0	0	1	0	0	0	0	2	25.0
9	1	0	1	1	0	0	0	1	4	50.0
10	0	1	1	0	1	1	1	1	6	75.0
11	1	1	1	1	0	1	1	1	7	87.5
12	0	1	1	1	1	1	1	1	7	87.5
13	0	0	1	1	1	0	1	1	5	62.5
14	1	0	1	0	0	1	1	0	4	50.0
15	0	0	0	0	0	1	0	1	2	25.0
16	1	1	0	1	0	0	0	1	4	50.0
TCP	12	11	12	11	7	9	8	13	5.19	64.8

ICP - correct tasks per person; QP - number of people who completed the task.

TABLE II. EXECUTION TIME OF COMPLETENESS TASK

Tasks	Execution Time of Completeness Task							
	P1	P2	P3	P4	P5	P6	P7	P8
1	33	95	78	24	35	10	0	4
2	32	25	43	67	0	0	0	15
3	36	18	0	0	0	0	0	27
4	16	12	10	11	16	0	19	23
5	41	19	15	0	0	62	48	0
6	8	14	6	4	7	11	0	18
7	6	9	11	7	32	14	6	20
8	64	0	0	31	0	0	0	0
9	28	0	19	19	0	0	0	23
10	0	21	24	0	63	32	27	25
11	27	14	30	19	0	16	14	18
12	0	3	4	3	8	8	3	4
13	0	0	240	38	83	0	14	100
14	3	0	3	0	0	1	6	0
15	0	0	0	0	0	8	0	5

Proc. of the Seventh Intl. Conf.on Advances in Computing, Communication and Information Technology - CCIT 2018

ISBN: 978-1-63248-162-7 DOI: 10.15224/ 978-1-63248-162-7-05

16	47	61	0	53	0	0	0	41
TTP	341	291	483	276	244	162	137	323

TTP - total time per person. Sum of the time of the tasks.

The efficiency was 61.07%, obtained from the division of the sum of the TTP divided by 3696 seconds.

The response of the SUS questionnaire can be seen in table III. P1 to P8 are the participants. Q1 to Q10 are the SUS issues.

TABLE III. SUS CALCULATION

SUS Calculation											
Participant	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**	**Q7**	**Q8**	**Q9**	**Q 10**	**SUS Score**
P1	4	2	2	1	4	1	2	2	4	1	**72.5**
P2	5	1	2	1	5	1	5	1	5	1	**92.5**
P3	4	1	4	2	4	2	4	2	4	2	**77.5**
P4	5	2	4	1	4	2	1	2	4	1	**75**
P5	4	2	2	1	4	2	2	2	4	1	**70**
P6	5	1	4	2	4	1	5	1	4	4	**82.5**
P7	5	2	5	1	5	2	5	1	5	1	**95**
P8	4	2	4	2	4	2	4	2	4	2	**75**
											80

In Figure 1, the total of usability problems identified by heuristics can be visualized.

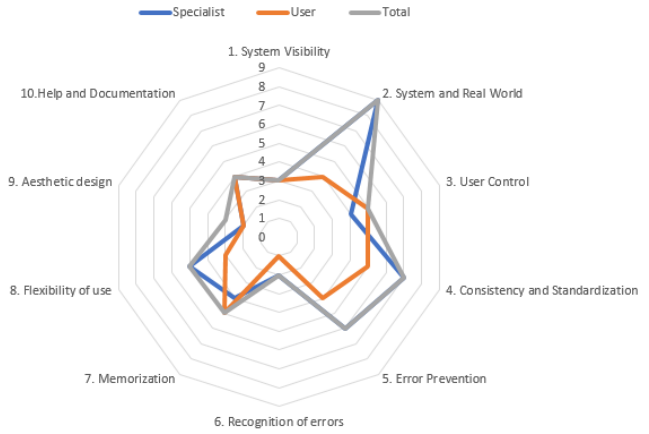

Figure 1. Number of usability problems per heuristic

Given this, it is noticed that the application has reasonable usability, needing improvements in several screens.

v. Conclusions

The main goal of this work was to show the usability evaluation performed in the SOS Chuva app, in order to understand what aspects and user interfaces need improvement, in order to provide an even better product to society. The main results obtained were (i) the effectiveness in the accomplishment of the tasks was of 64.8%, being considered reasonable, showing the need of adjustments in the application; (ii) the efficiency, related to the execution time of the tasks, was of 61.07%, being considered reasonable, showing the need for adjustments in the application; (iii) the satisfaction level, as measured by SUS, was 80 points, showing that people can learn as they use the application.

As suggestions for future work, we highlight the usability analysis with other profiles, especially with users who have never used the app, as long as they are interested in forecasting time; suggestions for refactoring app screens; performing A / B tests to determine the best icon for each feature; making new usability assessments, after the modifications in the app.

Acknowledgment

We thank FAPESP, process 15/14497-0, for the support for the development of this work, also highlighting the student of Scientific Initiation, process 18/04754-3.

References

[1] Machado, L.A.T.; Martins, R.C.G. Caracterização de Tempestades na Amazônia Durante os Experimentos RACCI e WET-AMC. https://daac.ornl.gov/LBA/lbaconferencia/2005_lba_student_conf/abstracts/225_ab.html

[2] Nielsen, J, Usability Engineering. 1993. ISBN-10: 0125184069.

[3] Nielsen, J. Thinking Aloud: The #1 Usability Tool. January 16, 2012. https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/

[4] eMarketer "App Marketing 2015: Fighting for Downloads and Attention in a Crowded Market," July 2015.

[5] Christian, R.. When to Use Which User-Experience Research Methods. October 12, 2014. https://www.nngroup.com/articles/which-ux-research-methods/

[6] Paganelli, L. and F. Paternò (2002). Intelligent analysis of user interactions with web applications. In Proceedings of the 7th international conference on Intelligent user interfaces, IUI '02, pp. 111–118. ACM.

[7] Tiedtke, T., C. Martin, and N. Gerth (2002). AWUSA: A tool for automated website usability analysis. In Proceedings of the 9th International Workshop on the Design, Specification and Verification of Interactive Systems, DSVIS '02, pp. 251–266. Springer.

[8] ISO 9241-11:2018. https://www.iso.org/standard/63500.html

[9] Adhy, S., Noranita, B., Kusumaningrum, R., Wirawan, P.W., Prasetya, D.D., Zaki, F. Usability Testing of Weather Monitoring on a Web Application. ICICoS. 2017.

[10] A. Bangor, P.T. Kortum, and J.T. Miller. Grade rankings of SUS scores from "Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale", 2009, Journal of Usability Studies, 4(3), 114-123.

[11] Sauro, Jeff. Measuring Usability with the System Usability Scale (SUS). February, 2, 2011. https://measuringu.com/sus/

[12] Brooke, John. SUS: A Retrospective. Journal of Usability Studies. Vol. 8, Issue 2, February 2013 pp. 29-40. .http://uxpajournal.org/wp-content/uploads/pdf/JUS_Brooke_February_2013.pdf

[13] Nielsen, J. (1994). Heuristic evaluation. In Nielsen, J., and Mack, R.L. (Eds.), Usability Inspection Methods. John Wiley & Sons, New York, NY.

[14] Nielsen, J. 10 Usability Heuristics for User Interface Design. Retrieved August 25, 2018 from: http://www.nngroup.com/articles/tenusability-heuristics/ (1995).

[15] Nielsen, J. How to conduct a heuristic evaluation. Retrieved August 25, 2018 from: http://www.gerrystahl.net/hci/he2.htm (2001).

ISBN: 978-1-63248-162-7 DOI: 10.15224/ 978-1-63248-162-7-05

[16] Nielsen, J. Usability 101: Introduction to Usability. Retrieved August 25, 2018 from: http://www.nngroup.com/articles/usability-101-introduction-to-usability/ (2012).

[17] Nielsen, J. (1994). Estimating the number of subjects needed for a thinking aloud Test. International Journal of Human-Computer Studies, 41(3), 385–397. https://doi.org/10.1006/ijhc.1994.1065.