EVALUATING THE USE OF THE COGNITIVE WALKTHROUGH USABILITY METHOD

Alice Demaria Silva PENHA[1], Flávio Anthero Nunes Vianna dos SANTOS[2]
[1] Universidade do Estado de Santa Catarina <alicedemaria@gmail.com>
[2] Universidade do Estado de Santa Catarina <flavioanvs@hotmail.com>

ABSTRACT

This paper addresses the evaluation of the Cognitive-Walk-Through (CWT) usability method by the application of an experiment with two groups: the control group and the experimental group. While one group carried out a task assessment without the use of any method, the other group used the CWT method. Both groups were intentionally formed only by designers, which aimed at evaluating the capability of what this method proposes: the possibility of the expert to put him or herself in the user's position. Although both groups came up with similar answers, significant differences were found. The evaluation of the group that used the method was deeper and more sensitive to the condition of the user. Furthermore, the method proved itself as an important guide in the search for answers focused on the task at hand.

KEYWORDS

Usability. Method evaluation. Interface.

1. INTRODUCTION

This study aims at validating a usability inspection method called the Cognitive Walkthrough Method (CWT). It was presented in the *Handbook of Human Factors and Ergonomics Methods* (SMITH-JACKSON, 2005) as one to be applied by experts and not by final users, based upon the belief that the evaluator can put him or herself in the user's position in order to identify problems related to design.

The CWT is a method that can be applied either by an expert or a designer, as long as the system is observed through the perspective of the user. Designers are used to developing interfaces that provide the best user experience (ROYO, 2008). Therefore, this research aims at discovering whether or not the use of a method really affects the search for problems. The research proposal is to discover if there are significant differences in the analysis of the same task undertaken by two groups of designers. In order to achieve this purpose, the control group will analyze the task without using the method, only relying upon the application of their previous design knowledge, while the other group – the experimental one – will use the method to evaluate the same task.

To carry out this research, the following contributions of both groups will be analyzed: ability to perform the task, content and interface analysis, runtime of the proposed task, as well as the verbal remarks of participants.

1.1. The method

The CWT is an analytical method, very similar to task analysis. It is used to identify usability problems related to products or to the apprenticeship of a system.

This method proposal suggests that a group of usability experts and/or designers can go through many scenarios of product use and identify the required cognitive stages so that the user can complete the task. For this reason, in order to accomplish the task, experts will have to suppose what users will know, think and learn, as well as how they will react or comprehend (SMITH-JACKSON, 2005).

The method is composed by the six stages below (SMITH-JACKSON, 2005):

1. Developing a complete understanding of the knowledge of users;

2. Identifying the tasks which represent what users will achieve in real life;

3. Creating detailed task-based scenarios;

4. Exploring the correct sequence of the actions required to complete the selected task;

5. Identifying and discussing the cognitive task processes that will be performed by the user to complete the scenarios successfully;

6. Identifying the apprenticeship or the most common reactions in product exploration.

One of the greatest advantages of this method is the low implementation cost, since it does not require the application of usability tests with users. However, one of its drawbacks is the low consistency between problems found and those reported by users in usability tests. (SMITH-JACKSON, 2005).

Usability tests were originated in experimental psychology. In the field of ergonomics, usability tests are used to test and evaluate the usability of systems and products by means of the observation of users at the exact time of use. However, usability tests tend to raise the price of interface evaluation projects (ROSA; MORAES, 2010). Thus, the use of the CWT method is proposed in this research in order to simulate the user's thoughts .

1.2. The task

The task chosen for analysis is part of the semester enrollment process of undergraduate students in a distance learning course of a university in Santa Catarina. In 2010, this university made available, at a unique website, an exclusive Academic Bulletin which contained all the information required for the enrollment process. Before that, the Academic Bulletin was a printed guidebook sent to students by mail and made available on digital file at the institution's website.

In order to enroll, the student needs to look at the curriculum of his/her course, choose the subjects of interest to attend the next semester and take note of the class number linked to the chosen subjects. There is one complicating factor in this task: the subjects must be chosen according to a schedule of subjects offered related to the specific dates of in-person assessments. Since it is a distance learning modality, the student cannot enroll in more than three subjects with the same in-person assessment date. According to the methodology of the institution, the student will not have time to take more than three in-person assessments on the same day.

In order to apply the method, the following task was developed: to choose subjects for enrollment of a student of the Information Technology Management course who joined the university in the

first semester of 2010. He has already attended some subjects of the first and second semester and, now, must enroll in the third semester. The difficulty included in the task is that the student did not attend a subject considered as a mandatory requirement for attendance of another subject of the third semester. Therefore, the dates of the in-person assessments must be carefully examined along with the limited amount admitted by the university.

2. APPLICATION

2.1. Participants

The participants chosen for the research are graduated designers with at least two years of market experience, collaborators of the university design team. The controlled group (C Team) was composed of 4 people (2 women and 2 men, between 25 and 35 years old) and the experimental group (E Team), of 6 people (3 women and 3 men, between 25 and 35 years old).

2.2. Gathering instruments

Both sessions were recorded directly on the computer used for the task. The audio and video recordings captured the reactions of the participants. All participants agreed to the use of the recordings and their transcriptions for the research purposes.

2.3. Procedure

Both teams had 30 minutes to perform the analysis. The C Team was asked to evaluate the task of choosing the subjects for the enrollment process. There was no explicit method: only a task without a detailed scenario. The E Team learned about the method and stages to achieve and was presented to a detailed scenario for the accomplishment of the task. The most significant observations were transcribed in Table 1.

Time	C Team	Time	E Team
02:35	Enroll yourself! Choose the enrollment. Before that, you have to check the information that it'll offer.	06:41	If I were a user, I would go to "Enroll yourself", then, I'd check if I have to go back.
05:21	Oh, here are the courses…	09:48	Yeah... this is the course.
05:25	Where's the code? Where did it say that there was a code?	09:50	What's the difference between TI 3 and TI 4?
05:30	Class number.	09:52	From the first semester of 2007... and the other one? First of 2012…
05:43	It wants us to take note of what?	10:10	We're gonna have to take note.
09:17	The student gets confused, he thinks "oh, I didn't take this here [the subject]", but remembers he has taken it, so he thinks he passed, or that he didn't pass.	10:56	Oh, you should go by the date of the in-person assessment, and choose three [subjects] on the spot.
10:07	The student has to follow his academic area. First thing... the	11:50	He can't take Computer Networks II.

	student doesn't follow it.		
12:00	Because that's what draws attention. I go straight to what I'm interested in because I don't want to waste any time.	13:47	We have to get the number of the class. Did you get it? We didn't…
12:05	I think each user interacts in a different way.	14:01	Go back there, which one is the subject?
12:07	The user has to struggle a bit as well. Not just start clicking there…	14:14	In our case, we would choose the subjects and go to [academic system] and, oh, no… go back there again.
12:22	So, when you get to the ending, something goes wrong and you have to go back to the beginning.	21:45	It's an absurd excess of information... something that we never really stop to look at.
12: 28	Here it shows you a text. It gives you an introduction, so then, later, it tells you how the content was organized.	21:50	And also the language used, why does it need to begin a sentence with "In possession of the list of subjects in which you will enroll and their respective class numbers for enrollment"?
12:37	So, here, it could be more succinct.	23:02	Not nice at all. There is no reason for formal treatment of something that will have to reach all kinds of people, ages and so on.
18:22	Yeah, it's kinda confusing, I'd call, like a good student, I'd call and say: can you explain to me what's going on!	30:25	I think it's a big mess, tiring, stressful, let's call [student services]. It's discouraging.
32:21	It's because this is curriculum 2, freshmen from the first semester of 2012... it looks like there was a change in the curriculum.	31:11	This much information in there makes it even harder… the enrollment process.
		31:31	Whoever thinks about the Bulletin has never used it. And doesn't even think about using it. It's all about "I suppose".

Table 1: Excerpts from transcriptions of C and E Teams

3. ANALYSIS

In the beginning of the task, both teams behaved identically: they chose the "*Matricule-se*" ("Enroll Yourself") item from the menu. The C Team [02:35] and the E Team [06:41] assertively chose that item

from the menu, but right after that, they realized that the enrollment item did not correspond to their expectations.

Still in terms of language and textual component of the interface, C Team noticed the enormous amount of textual information contained in big blocks [12:28] and E Team noticed, besides that, an excessive formality of the text [21:50; 23:02].

Both teams took detours along the task. The C Team talked about other enrollment systems and both teams logged into the enrollment system.

After the detour, both teams reported that they would ask for help. The C Team verified that the process is confusing [18:22]. The E Team verified that in a more intense way, especially at the end of the session [30:35; 31:11]. Moreover, both teams suggested improvements for the Academic Bulletin.

In terms of the processes required for the task, the C Team partially completed the task in a shorter period of time: in about 5 minutes the task had been completed, although the only item analyzed was the relation between the curricula numbering and the students' year of entry [32:21]. The E Team finished the entire task in a detailed and successful way within a period of time a little longer than 15 minutes [15:09].

Each team had a different perspective during the test: the C Team guessed the behavior of the user [09:17; 10:07; 12:05; 12:07; 12:22] and the E Team felt the difficulties encountered by the user while performing the task [21:45; 31:31].

It was interesting to realize that both teams began the task in the exact same way and made a mistake at the same time and place (menu). Although not using the same method as a basis for the evaluation, the C Team proved that designers do endeavor to think like users.

In terms of language, their considerations were alike even though the E Team performed a more profound analysis. Besides talking about the amount of text, the E Team also talked about its content and how the user would receive the message.

Both teams realized that, at some point, the user would simply give up trying to perform the task by him or herself and would seek for help.

Detours were not convenient for the research, since they were not part of the task and made the analysis too long and out of focus. However, the detour was important for some members to understand the need for the task in the user's life.

The E Team took more time to accomplish the task, but did it with a more profound understanding of the process. In this sense, the method helped to specify the task. The E Team knew exactly what to do and the restrictions to be faced. Without using the method, the C Team accomplished the task in a shorter period of time, but did not understand the whole process that the user would have to go through.

As designers and collaborators of the institution that produces the Academic Bulletin, it was not a surprise to notice that both teams felt free to suggest improvements for the Bulletin, even though that was not part of the method.

It was even more interesting to realize the difference in the feelings between both groups of evaluators. While participants of the C Team just tended to observe and judge the behavior of users in the presence of problems found, members of the E Team placed themselves as users and described feelings of frustration before difficulties presented by the task.

4. CONCLUSION

After testing both groups, it was possible to understand the importance of using a method to inspect usability in systems. The method chosen for evaluation served as a guide, specifying what should be performed and keeping the focus on what should be discussed.

The method evaluation result was satisfactory: the remarks from the experimental group were richer when related to "placing themselves as the user" (basis of the method) and both teams were able to achieve the task taking into account the interface and language of the evaluated material. In a way, this observation could be related to the fact that all participants were designers, meaning that they tend to think as users when developing products. The evaluation of the group that did not use the method was more superficial and unfocused, whereas participants of the group that used the method made more consistent remarks, liable of understanding and internal reflection.

The application of this research with two groups (control and experimental) was very interesting due to the observation of the participants' behavior. It was possible to notice that the control group was disperse from its objective and was not able to finish the test, although having completed the task in 5 minutes. On the other hand, the experimental group followed the steps established and was able to be more objective.

During test application, we realized that the literature recommendation to offer a reward for participants works and, also, that an evaluation generates anxiety in participants.

Amongst the directions of this research are found: a) the evaluation of other usability inspection methods and b) investigation about the contributions of those methods in the human-machine systems project in other situations, such as redesign or continuous improvement of projects.

5. ACKNOWLEDGEMENT

We would like to thank our fellow designers for participating in this research and providing their time and resources for the accomplishment of this study. We would also like to thank the Confesol board for allowing the development of the present research.

6. REFERENCES

[Moraes10]Anamaria de Moraes, José Guilherme Santa Rosa. *Avaliação e Projeto no Design de Interfaces*. Rio de Janeiro: 2AB Editora, 2010.

[Rojo08] Javier Rojo. *Design Digital*. São Paulo: Rosari, 2008. (Fundamentos do design)

[Smith-Jackson05] Tonya Smith-Jackson. Cognitive Walk-Through Method. In: STANTON, Neville et al. (Ed.). *The handbook of human factors and ergonomics methods*. Boca Raton, Flórida: CRC Press LLC, 2005. 82-1 – 82-7.