A Usability Inspection Approach to Assist in the Software Development Process

Priscila Silva Fernandes, Bruno Araújo Bonifácio
Institute of Exact Sciences and Technology (ICET/UFAM)
Federal University of Amazonas
Itacoatiara, Brazil
{pry.bila, brunobonni}@gmail.com

Priscila Silva Fernandes, Tayana Uchôa Conte
Software Engineering and Usability Group (USES)
Institute of Computing (IComp)
Federal University of Amazonas
Manaus, Brazil
{priscila.fernandes, tayana}@icomp.ufam.edu.br

Abstract— Several approaches have been proposed to ensure the quality of interactive systems. However, interactive systems continue to reach users with malfunctions, such as usability, communicability and interaction errors. Researches show that the lack of usability knowledge in software development organizations is an obstacle for usability evaluation. Our research goal is to popularize usability inspections so that even novice inspectors are able to perform it. Aiming to provide an approach to be used during the development process of web application, we have proposed the WE-QT technique. We are using an experimental methodology to evolve our technique and transfer it from the academy to industry. This paper presents a new comparative study; the results show that WE-QT technique is more efficient than and as effective as the compared technique.

Keywords—usability inspection; novice inspectors; web application; experimental study.

I. INTRODUCTION

Interactive systems have been widely developed and used nowadays. Integrating this growth to web services, users are able to be more connected. Despite the great evolution of web applications, and the existence of various approaches addressing their development [1], users sometimes experience malfunctions when interacting with them. These malfunctions are generally caused by the poor interaction design [2].

Web applications with bad interaction design leads not only to users dissatisfaction and frustration, but also to rework during the development and maintenance phases, costs surpass, and market disadvantage [3]. Usability is a quality factor that can improve interaction design of software products [4][5]. Therefore, improving usability of web application can minimize users’ interaction difficulty and improve the quality of these applications [6][7]. However, researches show that a large fraction of this problem is originated on the development process of these systems, which sometimes does not embody Human Computer Interaction (HCI) principles and methods for the development and evaluation of these applications [8][9][10]. Some authors [8][9][10] state that developers do not use, avoid or incorrectly apply HCI principles. This fact is due to the scarcity of knowledge and experience in concepts and practice of HCI area.

We proposed the Web Evaluation – Question Technique (WE-QT) [11][12] seeking to assist software developers performing usability inspections, and hence to improve the quality of the software products. The WE-QT technique is currently in the third version. We are using an experimental methodology to evaluate and evolve our technique [14]. This paper presents a comparative experimental study between the new version of WE-QT and its base technique, the Web Design Perspectives-based Inspection – Reading Technique (WDP-RT). Results show that our technique is as effective as and more efficient when compared to WDP-RT. Future work includes running studies with a major sample and comparing WE-QT to other methods.

The remaining of this paper is organized as follows. Section II addresses usability concepts and methods, along with the description of the new version of the WE-QT technique. Section III describes the experimental methodology and the results of the study. Section IV presents the conclusions.

II. USABILITY EVALUATION

According to Krug [15], users do not want to spend time trying to discover what the web application is about, figuring out whether an unusual button is actually a link, or how to go back to a previous visited page. Several researches propose approaches to ensure the quality of software product. Some propose tools and techniques specific to evaluate usability [6]. Usability evaluation methods can be divided into two groups: usability inspections and usability tests [16][18]. Usability inspection consists of a detailed interface analysis by an expert, while usability test seeks to uncover problems based on user observation [6]. Usability test is often more expensive because it requires users’ time and specific material or infrastructure, such as usability labs [6]. Usability inspection was proposed as a better cost-effective method [6]. The majority of the researches focus on usability tests [8]. Our work, however, is centered on usability inspections. We also restricted the software products to web applications, and the target public to novice inspectors.

Literature provides similar works [10][21][22][23][24][25][26][27]. Conte et al. [10] proposed the Web Design Perspective-Based Usability Evaluation Technique (WDP). The WDP is a checklist technique that uses the Nielsen’s Heuristic Evaluation [2] illuminated by the three web-design perspective: Presentation, Conceptual and Navigation [10]. The authors...
state that the technique is feasible, however inspectors had difficulties to apply the technique due the lack of knowledge and experience in usability and inspections [10]. Gomes et al. [27] propose the WDP-RT technique. The WDP-RT has the WDP technique as base and it is detailed in the next subsection.

A. WDP-RT

WDP-RT is a reading technique specific designed to inspectors with little knowledge on usability and inspections. The WDP-RT technique consists of a three pages document containing instructions to assist the evaluator uncovering usability problems. The reading approach provided by the WDP-RT technique does not simplify the inspections to novice inspectors and it is generally very time consuming, since inspectors have to read the three-page-document to carry out the inspection. The results of empirical studies to evaluate the WDP-RT technique indicated that the inspectors still have difficulty on understanding its instructions and applying it [28]. The WDP-RT technique is available at [33]. Both WDP and WDP-RT techniques require training on usability and on the technique before the inspection.

B. The WE-QT Technique

The WE-QT technique is an approach to guide novice evaluators performing usability inspections, and has the WDP-RT as base [11][12]. This research focused on novice inspectors aiming to reach industry workers that have little experience/knowledge on human-computer interaction concepts and practices, more specifically on usability and inspections. Our technique is composed by questions [12]. The questions lead the inspector through a flow that is adaptable by the elements present on the interface [14]. The mapping, provided by the question flow is illustrated in Figure 1.

The WE-QT technique is currently in its third version. Our technique does not require training on usability or on the technique itself before utilizing it. Since this approach can be applied by development team itself, reducing the need of executing usability tests or hiring an expert inspector, it is considered a cheaper option to improve the quality of web application. Some improvements made for the third version of the WE-QT technique are as follows: (1) Adding descriptions/examples to illustrate each question/affirmative, aiming to increase the information to assist the novice

![Figure 1. Extract of the WE-QT technique.](image-url)
inspectors to better judge possible problems, Table I shows an extract of the descriptions and examples of the questions/affirmatives; (2) Implementing the scope of the inspected elements, we added questions addressing pop-ups and logins; (3) Adding initial instructions on how to conduct an inspection flow; (4) Relocating some questions, aiming to avoid mistakes concerning switching pages when the inspection flow is executed.

TABLE I. ILLUSTRATION OF THE EXAMPLES PROVIDED BY THE WE-QT TECHNIQUE – THIRD VERSION.

<table>
<thead>
<tr>
<th>Question/Affirmative</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>I cannot see the messages easily</td>
<td>If the message source is small, or if the message is in a difficult location to be seen</td>
</tr>
<tr>
<td>Does the page inform you in which part of the application you are at?</td>
<td>Some mechanism to inform you on which page of the system you are at</td>
</tr>
<tr>
<td>Is the page part of a page sequence of a task (e.g., a registration with several steps)?</td>
<td>A registration form with several steps, or a form with several pages</td>
</tr>
<tr>
<td>Are the mandatory fields to be filled in well defined?</td>
<td>Defined by a symbol as &quot;*&quot; or a similar one</td>
</tr>
</tbody>
</table>

III. EMPIRICAL EVALUATION

Following the experimental methodology, we designed a observational comparative study to evaluate the WE-QT technique. As the WE-QT has the WDP-RT as its base technique, we initially used the WDP-RT technique as the compared technique. We used statistical test to analyze the quantitative data and we present the subjective results. The second observational study is detailed below.

We selected 16 students from the third year of Information System course at the Federal University of Amazonas (UFAM), which were attending an Analysis and Design class. All the participants were familiar with Web applications. One participant had high knowledge and experience on usability, three participants had medium, four of them had low and the others had none. We divided them into two groups: Group 1 (participants used the WE-QT technique) and Group 2 (participants used the WDP-RT technique). All the participants carried out the inspection individually. Table II shows the participants characterization regarding usability knowledge and experience. The participants had also participated in another study, concerning usability of models, in which it was provided two hours of usability training.

The evaluation object was the MPS.Br Portal [20]. This application is responsible for providing information concerning the MPS.Br program. The MPS.Br is a nationwide program, equivalent to the CMMI [29], for software process improvement in Brazilian organizations. The MPS.Br aims to establish a feasible pathway for organizations to achieve benefits from implementing software process improvement at reasonable costs, especially small and medium-size enterprises [30]. It was the same application used in the feasibility study, as described in [11]. However, due to the large number of pages to be evaluated in the feasibility study, we only selected two tasks, with two web pages each, to be executed by the inspectors during the evaluation. The following tasks comprised the inspection’s context: Obtain information about the Implementation Guides. These guides describe orientations on how to implement some expected results of the MPS.Br program and access the presentations provided by the MPS.Br Portal, such as presentations about the MPs.Br program, workshops, and projects.

We used Morae (version 3.3) usability testing software to capture the inspection section of each inspector and to assist the collection of the perceptions of each inspector during the evaluation. Subjective data was gathered at the completion of the inspection phase using post-inspection questionnaires. We provided the subjects with the Inspection Guide and a Consent Form (all the subjects signed the consent form before starting the inspection).

To support the mapping process of the WE-QT technique, we developed an automated tool called WE-QT Assistant. The support tool was designed to minimize the effort of the inspectors during the problem detection phase. Therefore, the tool was developed to be located at the left side of the screen, allowing the inspection to be performed without the need to switch windows. The left side of Figure 2 illustrates the WE-QT Assistant. The Assistant provides text boxes in order to allow the inspectors describing the identified usability problems instead of needing an extra document to report the problems.

TABLE II. SUMMARY OF INSPECTION RESULTS PER SUBJECT

<table>
<thead>
<tr>
<th>No</th>
<th>Usabil. Exp.</th>
<th>Discr. ep.</th>
<th>Nº Real Prob.</th>
<th>False-positives</th>
<th>Time (min)</th>
<th>Effectiv. (%)</th>
<th>Effic. (Prob./hour)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>None</td>
<td>9</td>
<td>7</td>
<td>2</td>
<td>37</td>
<td>7.22%</td>
<td>11.35</td>
</tr>
<tr>
<td>2</td>
<td>Low</td>
<td>17</td>
<td>9</td>
<td>8</td>
<td>37</td>
<td>9.28%</td>
<td>14.59</td>
</tr>
<tr>
<td>3</td>
<td>Low</td>
<td>26</td>
<td>16</td>
<td>10</td>
<td>35</td>
<td>16.49%</td>
<td>27.43</td>
</tr>
<tr>
<td>4</td>
<td>Medium</td>
<td>28</td>
<td>14</td>
<td>14</td>
<td>22</td>
<td>14.43%</td>
<td>38.18</td>
</tr>
<tr>
<td>5</td>
<td>None</td>
<td>37</td>
<td>31</td>
<td>6</td>
<td>72</td>
<td>31.96%</td>
<td>25.83</td>
</tr>
<tr>
<td>6</td>
<td>None</td>
<td>65</td>
<td>55</td>
<td>10</td>
<td>75</td>
<td>56.70%</td>
<td>44.00</td>
</tr>
<tr>
<td>7</td>
<td>Medium</td>
<td>31</td>
<td>25</td>
<td>6</td>
<td>39</td>
<td>25.77%</td>
<td>38.46</td>
</tr>
<tr>
<td>8</td>
<td>None</td>
<td>30</td>
<td>22</td>
<td>8</td>
<td>53</td>
<td>22.68%</td>
<td>24.91</td>
</tr>
<tr>
<td>9</td>
<td>None</td>
<td>13</td>
<td>7</td>
<td>6</td>
<td>80</td>
<td>8.64%</td>
<td>5.25</td>
</tr>
<tr>
<td>10</td>
<td>Low</td>
<td>21</td>
<td>9</td>
<td>12</td>
<td>86</td>
<td>11.11%</td>
<td>6.28</td>
</tr>
<tr>
<td>11</td>
<td>Medium</td>
<td>43</td>
<td>24</td>
<td>19</td>
<td>109</td>
<td>29.63%</td>
<td>13.21</td>
</tr>
<tr>
<td>12</td>
<td>High</td>
<td>19</td>
<td>9</td>
<td>10</td>
<td>65</td>
<td>11.11%</td>
<td>8.31</td>
</tr>
<tr>
<td>13</td>
<td>Low</td>
<td>65</td>
<td>44</td>
<td>21</td>
<td>163</td>
<td>54.32%</td>
<td>16.20</td>
</tr>
<tr>
<td>14</td>
<td>None</td>
<td>24</td>
<td>15</td>
<td>9</td>
<td>104</td>
<td>18.52%</td>
<td>8.65</td>
</tr>
<tr>
<td>15</td>
<td>None</td>
<td>31</td>
<td>21</td>
<td>10</td>
<td>107</td>
<td>25.93%</td>
<td>11.78</td>
</tr>
<tr>
<td>16</td>
<td>None</td>
<td>13</td>
<td>6</td>
<td>7</td>
<td>77</td>
<td>7.41%</td>
<td>4.68</td>
</tr>
</tbody>
</table>

In order to minimize the threats to validity, we developed a tool to support the WDP-RT technique as well. The tool is similar to the WE-QT Assistant, and also has text boxes to problem description and it is located on the left side of the screen.
In this study, we used the effectiveness and efficiency indicators. These indicators have been employed in other studies concerning usability inspection methods as well [10][11][19]. Effectiveness is defined as the ratio between the number of detected problems and the total of existing problems; and Efficiency is defined as the ratio between the number of detected problems and the time spent in the inspection.

We also collected participants’ subjective impressions of the techniques through a post-inspection questionnaire, which was based on the Technology Acceptance Model (TAM) [31]. This model aims to examine a new technology usage and verify the user perceptions concerning usefulness and ease-of-use, the key determinants of individual technology adoption.

The experiment was used to test the following hypotheses concerning to the effectiveness and efficiency (null and corresponding alternative hypotheses are given):

- H01: There is no difference between the effectiveness of techniques WE-QT and WDP-RT.
- H1A: The effectiveness of the WE-QT technique is greater than the effectiveness of the WDP-RT technique.
- H02: There is no difference between the efficiency of techniques WE-QT and WDP-RT.
- H2A: The efficiency of the WE-QT technique is greater than the efficiency of the WDP-RT technique.

The inspection phase was carried out by each subject individually. They were provided with the instruments to accomplish the inspection and received instructions about the inspection flow plus a five minutes exemplification of usability problems detection using the WDP-RT technique and its instructions. It is worth to mention that the subjects from Group 1 did not receive training on the WE-QT technique. Once the inspector understood the procedures, the inspection process began. Figure 2 shows a subject from Group 1 evaluating the MPS.Br Portal with the WE-QT technique (WE-QT technique, problem detection and description); it also illustrates in the interface of the application an example of an uncovered problem. One researcher acted as the facilitator, being responsible for conducting the detection phase and passing the initial information to the subjects. After the detection phase, the subjects received the post-inspection questionnaire by e-mail and they could answer them at home.

At the end of the inspection phase, the researches elaborated a list containing all usability problems detected by the inspectors, without duplicates. Then, a meeting attended by the researchers and a control group formed by usability specialists took place. The list of problems was discriminated to classify these problems into real problems or false positives. To eliminate any possible influence during the discrimination meeting, the problem list did not contain any information about which technique uncovered which problems. The authors of the technique did not influence the discrimination, they were not allowed to comment or give
their opinion about whether the discriminated problems were real problems or false-positives.

A. Results

After the discrimination meeting, we were able to analyze the gathered data. We computed the number of detected problems, false-positives, time spent during the inspection phase, efficiency and efficacy for each inspector of each group.

1) Quantitative results: As a result of the inspection, the inspectors identified a total of 135 real problems, including both techniques. The WE-QT group detected a total of 97 problems, while the WDP-RT group uncovered 81. Table III shows the averages for the time, and effectiveness and efficiency indicators. Regarding the efficiency indicator, inspectors detected an average of 31.91 defects per hour using the WE-QT technique. Table II presents the overall result of the usability evaluation for each inspector, including their experience level.

TABLE III. AVERAGE EFFECTIVENESS AND EFFICIENCY PER GROUP.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Average Effectiveness (%)</th>
<th>Average Efficiency (Per hour)</th>
<th>Average Time (min)</th>
<th>Total Known Defects</th>
</tr>
</thead>
<tbody>
<tr>
<td>WE-QT</td>
<td>23.07</td>
<td>28.09</td>
<td>46.25</td>
<td>97</td>
</tr>
<tr>
<td>WDP-RT</td>
<td>20.83</td>
<td>9.29</td>
<td>98.88</td>
<td>81</td>
</tr>
</tbody>
</table>

We carried out a statistical analysis using the statistical tool SPSS (SPSS Statistics version 17.0), and \( \alpha = 0.05 \). This choice of statistical significance was motivated by the small sample size used in this experiment.

Concerning H1: Effectiveness of Techniques WE-QT and WDP-RT, we compared the two samples, Group 1 (WE-QT) and Group 2 (WDP-RT), using the Mann-Whitney test, a non-parametric test, we found no significant differences between the two groups (\( p = 0.753 \)). These results show that both techniques provided similar effectiveness when used to inspect the MPS.Br Portal. Figure 3 shows the boxplot with the distribution of effectiveness per subject, per technique.

Regarding H2: Efficiency of Techniques WE-QT and WDP-RT, the boxplots with the distribution of efficiency per subject, per technique (see Figure 4) show that Group 1 (WE-QT) was considerably more efficient than Group 2 (WDP-RT) to inspect the usability of the MPS.Br Portal: Group 3’s median is significantly higher than Group 2’s. When we compared the two samples using the Mann-Whitney test, it confirmed significant statistical differences between the two groups (\( p = 0.021 \), which supports the alternative hypothesis HA2, and therefore rejects the null hypothesis H02. These results suggest that the WE-QT technique efficiency was significantly higher than the WDP-RT’s. Results show that effectiveness of both techniques is similar; however the WE-QT technique was nearly three times more efficient then the WDP-RT technique.

2) Subjective results: When we proposed the WE-QT technique, one of our aims was to improve users’ satisfaction when using the technique, therefore we are also evaluate this aspect in this study. We collected the subjects’ opinions with respect to key determinants of individual technology adoption, perceived ease of use and usefulness; collected with the post-inspection questionnaire, based on the TAM model. The questionnaire had closed and opened questions. The closed questions were based on a 6 value scale — 0% (Totally Disagree), 1%-30% (Strongly Disagree), 31%-50% (Partially disagree), 51%-69% (Partially Agree), 70%-99% (Strongly Agree) and 100% (Totally Agree); note that it did not have a neutral option, forcing the subjects to stand a position on whether they agree or disagree. Figure 5 shows the average subject ratings, together with standard deviations. The ease of use
perception comprises factors such as learnability, customization of use, ability gain and understanding of the technique; while usefulness perception covers factors such as usefulness, performance improvement, productivity, and efficiency when using the technique. We also add questions addressing the language of the techniques, to identify any possible improvements suggestions. The questionnaire contained discursive questions as well. According to Figure 5, the WE-QT technique was perceived slightly more easy to use and useful than the WDP-RT technique. The both techniques were ranked similarly to the language aspects. The participants’ subjective answers could be affected by the tool to support the techniques automatons, which we will evaluate in further studies. The subjective data were important to improve the WE-QT technique.

IV. CONCLUSION AND FUTURE WORK

In this paper, we evaluated the WE-QT technique, a question-based usability evaluation technique for web applications, specifically tailored for software developers with little knowledge HCI principles and concepts, more specifically on inspections and usability. We used a formal statistic experiment to compare the efficiency and effectiveness of both techniques: WE-QT and WDP-RT. The results showed that our technique in significantly more efficient than and as effective as the WDP-RT technique. We also evaluate the techniques concerning the perceived ease of use, usefulness and language of the techniques. Subjective results showed that our technique was perceived slightly more easy to use and useful then the WDP-RT technique. These results are very promising. However, we will continue to research our technique. Limitations of this research include: focusing on novice inspectors, additional research is required to specific address this topic; the small sample of participants; comparing WE-QT only to one technique. Future work includes: (1) improvement of the technique based on a detailed analysis of the detected usability problems, false-positives and time spend; (2) investigation of a new arrangements for each usability questions, for instance, if efficiency, effectiveness and user satisfaction can be improved if the questions regarding the web application as a whole came first then the questions regarding each individual page; (3) further studies comparing the WE-QT technique with other usability inspection techniques specific for evaluate web applications, with a greater number of subjects; and (4) the replication of the experiment in an industrial environment. With this research we also aim to encourage professionals involved on the development process of interactive systems, such as developers, analysts, testers and stakeholders, to use HCI principles and methods in the development cycle.

ACKNOWLEDGMENT

The authors thank the support granted by CAPES process 00.889.834/0001-08; the researchers Luis Rivero and Natasha Valentin for participating on the discrimination meeting; Martha Fernandes and also all participants of the conducted study.
REFERENCES


