Web Usability Inspection Technique
Based on Design Perspectives

Tayana Conte[1], Jobson Massolar[1], Emilia Mendes[2], Guilherme Horta Travassos[1]

[1]PESC – COPPE/UFRJ, Cx Postal 68.511, CEP 21945-970, Rio de Janeiro, RJ, Brasil
[2]Computer Science Department, The University of Auckland, Private Bag 92019, Auckland, New Zealand

`{tayana, jobson, ght}@cos.ufrj.br, emilia@cs.auckland.ac.nz`

Resumo. *Este artigo apresenta uma técnica de inspeção de usabilidade (WDP - Web Design Perspectives-Based Usability Evaluation) e a metodologia experimental utilizada como base para seu desenvolvimento e aprimoramento. Até o presente momento foram executados três estudos experimentais (um secundário e dois primários) que levaram à versão atual da técnica WDP. Resultados indicam que a técnica é viável e que é possivelmente mais efetiva e tão eficiente quanto a Avaliação Heurística proposta por Nielsen.*

Abstract. *This work describes a usability inspection technique (WDP - Web Design Perspectives-Based Usability Evaluation) as well as the experimentation methodology followed to support its development and improvement. At this point in time, three experimental studies (one secondary and two primaries) were carried out and led to the WDP's current version. Results indicate the feasibility of the WDP technique and its possibility to be more effective than and as efficient as Nielsen's Heuristics Evaluation.*

1. Introduction

In recent years the development demand of Web applications has grown and, today, this category of software already represents a considerable development effort in the general context of software applications. This growth has been observed not only from the point of view of the amount of features of Web applications, but also from the point of view of criticality and risks associated to each one of them. These risks, if not mitigated, may present potential damages to the organizations' image regarding reliability, security, transparency, and so on.

The prominent position of Web applications plus the unique features of this kind of software motivated the emerging of a new research field: Web Engineering, aimed at applying the software engineering principles to develop quality Web applications [Pressman 2000]. According to Pastor (2004), the main goal of Web Engineering is to develop correct Web applications where structure, functionality, navigation and user interaction have to be properly represented with quality.

Regarding quality, Abrahão et al. (2003) stated that Web applications' quality has often been assessed in an *ad-hoc* way, based on common sense, intuition and developers' expertise. Therefore, there is a need for new quality assurance techniques tailored to Web applications' features.

According to Offutt (2002), one of the three quality criteria on the dominant Web development drivers is usability. ISO 9241 standard defines usability as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". There are several usability evaluation methods proposed in the literature and they can be applied to different kinds of software systems [Nielsen 1994, Zhang et al. 1999, Dix et al. 2004].

Usability is one relevant factor of a Web application's quality due to its own features, as: "Web applications are interactive, user-centered, hypermedia-based applications, where the user interface plays a central role" [Olsina et al. 2006]. According to Matera el al. (2006), the acceptability of Web applications by users relies strictly on the applications' usability. A Web application with poor usability will be quickly replaced by another one more usable, as soon as its existence becomes known to the target audience [Mendes et al. 2006]. Web Usability has two main objectives: (1) to drive the design of Web Applications and (2) to evaluate the relevant usability criteria of Web Applications. Defining methods and techniques for ensuring usability is therefore one of the current goals of Web engineering researches [Matera et al. 2006].

This scenario has motivated one of our research's goals: to define a set of inspection techniques tailored to support quality assurance concerned with specific features of Web Applications. These inspection techniques should focus on quality criteria relevant for Web Applications development and should be evaluated using a scientific process in order to support knowledge building. According to Pfleeger (1999), no science can advance without experimentation and measurement. The use of the experimentation is necessary as a source of reliable knowledge to prove the results of the techniques proposed. In addition, experimental studies can also be used as a safe approach for transfer software technologies from academy to industry [Shull et al. 2001, Mafra et al. 2006].

To support the development and validation of our usability inspection technique, we have adopted the experimental methodology presented in [Mafra et al. 2006], summarized on Figure 1. This methodology, that represents an extension of [Shull et al. 2001], comprises six stages: 1) the execution of secondary studies to identify, evaluate and interpret all available research relevant to a particular research question or topic area [Kitchenham 2004]; 2) the creation of the initial version of the technique based on the result of secondary studies; 3) the execution of feasibility studies to determine the usage possibility of the technology; 4) the execution of observational studies to improve the understanding and the cost-effectiveness of the technology; 5) the execution of case studies in real lifecycle to characterize the technology application during a real lifecycle, and; 6) the execution of case studies in industry to identify if technology application fits into the industrial setting.

This work presents a Web usability inspection technique, as well as the experimental methodology adopted to support its development and improvements. In two previous studies [Conte et al 2005, Conte et al 2007], we discussed specific results of this research: the evidence gained through a secondary study regarding Web development processes and the results of a formal experiment to assess the second version of the technique. In this article, we present an overview of the development of this technique including the description of all the studies that helped define and improve

our technique. The elaboration cycle of the technique is illustrated and discussions on how the methodology guided each phase of its development are presented, including the studies that were performed and how the methodology impacted the development of the Web usability evaluation technique.

The remainder of this paper is structured as follows: Section 2 presents the secondary study performed to obtain evidence related to Web applications development. Section 3 describes the initial proposal of the technique and the first study aimed at the evaluation of the technique feasibility. Section 4 describes the redesign of the technique and the second study: a formal experiment performed to evaluate the second version of the technique. Conclusions and comments on future works are given in Section 5.
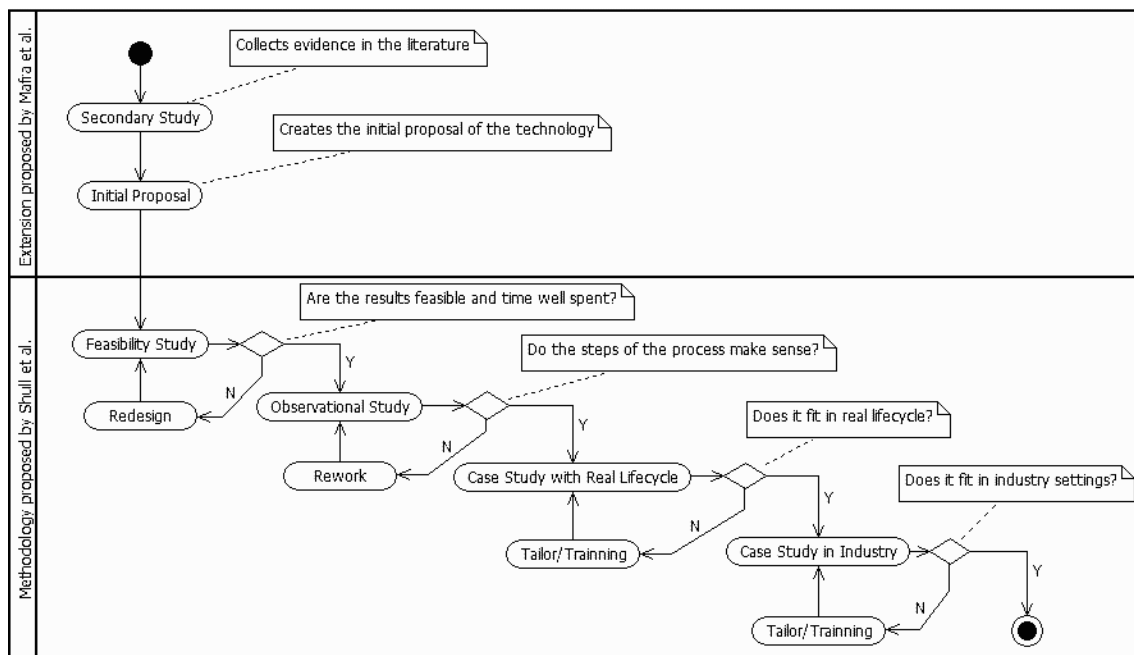

Figure 1 Experimental Methodology's Overview

2. Collecting Evidence in Literature regarding Quality Assurance Techniques for Web Applications

According to the experimental methodology, our first step was to obtain evidence about Web development processes and their quality assurance techniques. Therefore, a systematic review was conducted following the method proposed in [Kitchenham 2004] and using the protocol given by [Mendes and Kitchenham 2004] as a model. The know-how in running systematic reviews and the lessons learned with planning and execution such reviews are discussed in [Mian et al 2005].

2.1. Secondary Study: Systematic Review of Web Development Processes

The systematic review aimed to characterize Web Development Processes and, in addition, their quality assurance activities, more specifically inspection related activities.

The complete systematic review is described in [Conte et al. 2005]. The two research questions investigated were as follows:

- First Research Question: "What development processes have been used to develop Web applications?"
- Secondary Research Question: "What processes have been used to inspect Web applications for quality control?"

The population of the study was that of Web development projects and the outcome was Web development processes. The strategy used to search for primary studies was to search electronic databases including digital libraries selected from CAPES Periodicals Portal. Due to the limitations of some digital libraries, the review had to be restricted to just two digital libraries (IEEExplore and Elsevier digital libraries).

After applying the selection criteria, 13 articles in IEEE digital library and 9 articles in Elsevier digital library were collected for the First Research Question. However, some of those articles described the same process (or development method). Therefore in total 19 different processes were found. In relation to the Secondary Research Question, none of the selected papers described or suggested the use of inspections during the Web development process.

The proposed criteria to analyze the 19 Web development processes were:
- Process Amplitude (All Cycle/ Restricted to Phases): if the process covered the entire Web development cycle (analysis, design, implementation and delivery) or only a few phases. 13 articles described processes applied to the entire development life cycle, and six articles described processes restricted to phases.
- Specific Web Design Artifacts Production (Yes/ No): if the process generates specific artifacts, such as a navigational map or a presentation diagram. 16 processes include specific artifacts' production and three do not.
- Computational Support (It requires specific support/ it suggests specific support/ None): if the process requires or depends on a specific case tool. Seven processes require specific computational support and 12 do not require computational support.

2.2. Secondary Study's Contributions

As pointed out in Section 2.1, none of the collected papers describes the use of quality assurance techniques such as reviews or inspections. This reinforces Abrahão et al. (2003) statement (see Section 1) and highlights the need to develop specific quality assurance techniques for Web development.

After conducting the systematic review, we initiated a detailed study about Web process and design artifacts. The review criteria for processes' categorization allowed us to select a subset of relevant processes:

- Process Amplitude: we chose only processes that cover the whole cycle.
- Specific Web Design Artifacts Production: we selected processes that suggest the design of specific Web artifacts.
- Computational Support: we selected processes that did not require specific computational support.

Due to search engines' limitation, we had to restrict the search space to only two digital libraries. Therefore, we added three well-known Web processes to the original list of the systematic review results. The final list contained eight selected processes:

- Object-Oriented Hypermedia Design Model (OHDM) (Schwabe et al., 1996)
- Web Modeling Language (WebML) (Ceri et al., 2000).
- W2000 (Baresi et al., 2000).
- Object-oriented Hypermedia method (OO-H) (Gómez et al. 2001).
- Web Application Extension (WAE) (Conallen, 2002).
- UML-based Web Engineering (UWE) (Koch and Kraus 2002).
- Object Oriented Web Solution (OOWS) (Fons et al., 2003).
- Ariadne Development Method (ADM) (Diaz et al., 2004).

The study of these processes enabled us to identify design perspectives commonly used in Web development. In addition, our own experience with industrial Web projects suggests that, for some Web applications, the following types of design perspectives should be used: conceptual, presentation, navigational and structural. They are defined as follows:

- **Conceptual**: represents the conceptual elements that make up the application domain.
- **Presentation**: represents the characteristics related to application layout and arrangement of interface elements.
- **Navigation**: represents the navigational space, defining the information access elements and their associations.
- **Structural:** represents the structural and architectural characteristics of the application, that is, how the application is structured in terms of components and their associations.

The motivation to use these four types of design perspectives is based on the following criteria:

- The need to map the application's domain to the presentation perspective, that is, to define views of the conceptual model taking into account the presentation requirements of different user groups;
- The need to map the application's domain to the navigation perspective, that is, to define responsibilities according to the requirements and rules defined in the navigation model; and that;
- There are user requirements that cannot be mapped to or handled by the conceptual, presentation, navigational design perspectives, as they are more related to business rules than to domain concepts, having impact on the final architecture of the Web application.

After carrying out a detailed study about Web development processes and Web design perspectives, we focused on a specific Web quality criterion: usability. Web Usability can be considered as the ability of Web applications to support their main tasks with effectiveness, efficiency, and satisfaction [Matera et al. 2006]. Therefore, we decided to investigate and propose an inspection technique focusing on the evaluation of Web usability.

3. Initial Proposal of a Technique: Web Design Perspectives-based Usability Evaluation

General usability evaluation methods can be divided in two categories: (1) Usability Inspections - evaluation methods based on Experts' Analysis; and (2) Evaluation Methods involving User Participation, such as: Laboratory studies, Think Aloud and Cooperative Evaluation [Dix et al. 2004]. Recently, Web usage analysis has also emerged as a method for studying user behavior, using the recorded users' access to the application's Web pages [Matera et al. 2006].

When using Evaluation Methods involving User Participation, usability problems are discovered by means of observation and interaction with users, while they perform tasks or provide suggestions about the interface design and its usability. In Usability Inspection, the focus of our research, problems are discovered by experts and their inspection techniques. Different usability inspection techniques have been developed and used, such as: Heuristic Evaluation [Nielsen 1994], Cognitive Walkthrough [Polson et al. 1992], and Guidelines and Checklists [Dix et al. 2004].

Usability inspections are naturally less expensive than evaluation methods that involve user participation, since they do not need, besides the inspectors, any special equipment or laboratory.

According to Zhang et al. (1999), it is difficult for an inspector to detect all kind of problems at the same time. Due to that, they proposed a usability inspection technique based on perspectives called "Usability Based Reading" (UBR). In UBR technique each inspection session focus on a subset of usability issue. It defines three perspectives for usability: novice use, expert use and error handling. The assumption behind perspective-based inspections techniques is that, thanks to the focus, each inspection session can detect a greater percentage of defects in comparison to other techniques that do not use perspectives. In addition, the combination of different perspectives can detect more defects than the same number of inspection sessions using a general inspection technique [Zhang et al. 1999].

Motivated by the Zhang's et al. (1999) results and assuming that Web design perspectives could be explored while building Web applications (see Section 3.2), we argue that the adoption of Web design perspectives could be also used to improve the efficiency of Heuristic Evaluation (HEV) [Nielsen 1994]. Basically, Web design perspectives are used as a guide to interpret Nielsen's heuristics. This way, each heuristic evaluation session evaluates a Web application's usability regarding domain concepts, presentation, navigation and application structure. This derived technique we call Web Design Perspectives-based Usability Evaluation (WDP).

In light of Web design perspectives, we define usability as follows:

- **Usability related to Domain Concepts**: relates to the clarity and the concision of problem domain's elements. Under this perspective, the usability is satisfactory if different users easily understand domain terms, what prevents mistakes caused by ambiguous, inconsistent or unknown terms.
- **Usability related to Presentation**: relates to how consistent the information is presented to the user. Under this perspective, the usability is satisfactory if the

arrangement of interface's elements allow the user to accomplish his/her tasks effectively, efficiently and pleasantly.

- **Usability related to Navigation**: relates to different user's access to the system functionalities. Under this perspective, the usability is satisfactory if the navigation options allow the user to accomplish his/her tasks effectively, efficiently and pleasantly.

- **Usability related to Structure**: relates to how the application structure is presented to the users. One understands for application structure as the associations between their components or modules. Under this perspective, the usability is satisfactory if the arrangement of the components and modules allow the user to accomplish his/her in an effective, efficient and pleasant way.

The initial version of the WDP technique was based on a set of heuristics that comprised the ten heuristics proposed by Nielsen (1994) and three additional heuristics adapted from Zhang et al. (1999). We associated this set of thirteen heuristics with the four Web design perspectives: Conceptual (C), Presentation (P), Navigation (N) and Structural (S). Table 1 shows the associations between heuristics and Web design perspectives in WDP version 1. In this table, the first ten heuristics represent the original set proposed by Nielsen (1994) and the last three correspond to the heuristics adapted from Zhang et al. (1999). The correlated pairs of heuristics-perspectives are marked with the ✓ symbol.

Table 1- Relationships between Heuristics and Design Perspectives in WDP v1

# H	Heuristics	Web Design Perspectives			
		C	P	N	S
1	Visibility of system status	✓	✓		✓
2	Matching between system and real world	✓			
3	User control and freedom			✓	
4	Consistency and standards		✓		
5	Error prevention	✓	✓	✓	✓
6	Recognition rather than recall	✓	✓		
7	Flexibility and efficiency of use			✓	
8	Aesthetic and minimalist project		✓		
9	Help users recognize, diagnose and recover from errors	✓	✓	✓	✓
10	Help and documentation	✓	✓	✓	✓
11	Minimize user's memory load and fatigue		✓	✓	
12	Visually functional design		✓		
13	Facilitate data entry				✓

For each pair heuristic-perspective, we pointed out hints to guide the use of the heuristic regarding the perspective's viewpoint. Figure 2 shows an extract of the technique, presenting examples of hints related to the Navigation Perspective.

This first version was assessed by a feasibility study described in the next section, in which we compared the number of defects found by inspection teams using WDP technique with the number of defects found by different inspection teams using a Heuristic Evaluation. The teams that applied the Heuristic Evaluation technique used the same set of thirteen heuristics (10 original heuristics from [Nielsen, 1994] + 3

heuristics adapted from [Zhang et al. 1999]). We called this set of thirteen heuristics HEV+, to differ from the original set of Heuristic Evaluation (HEV).

Usability related to Navigation

Heuristic: User control and freedom
- Evaluate if the interface allows the user to navigate easily using different steps of a given task.
- Evaluate if the interface has undo or redo functions or similar functions that allow the user to exit in case of wrong choices.
- Evaluate if the interface allows the user to return to the main flow of a task after a detour or after performing a secondary task.

Heuristic: Error prevention
- Evaluate if the interface prevents navigation errors, providing accessibility for the different types of users.
- …

Figure 2 - Extract of WDP v1

3.1. First Feasibility Study

Following the experimental methodology, we performed the first feasibility study in June 2006. This study addressed the question "Is the WDP technique feasible regarding the number of detected defects?" To answer this question we compared the WDP technique to the HEV+ as follows:

Subjects: The subjects were represented by twenty students from a senior-level undergraduate software engineering course at UFRJ. All subjects signed a consent form and filled out a characterization form, with answers scaling from 1 (no experience) to 5 (experience in several software industry projects), concerning their expertise with applications design, development tools and Internet usage in general. The characterization data were analyzed and each subject was characterized as having High, Medium or Low experience according to the information given by each one of them. The subjects were equally distributed in four teams, as summarized on Table 2.

Materials: All teams applied the WDP and HEV+ techniques to evaluate the usability of a real Web application called JEMS[1] (Journal and Event Management System). The JEMS application is used to support the process of conference creation, users' registration, submission, revision, acceptance and publishing of papers from conferences and journals sponsored by Brazilian Computer Society.

Procedure: The context of the inspection comprised sixteen use cases involving three actors (Conference Chair, Reviewer and Paper Author). The teams had three weeks to complete the task and they were free to organize themselves to reach their goals, which were: to inspect all scenarios related to the sixteen use cases and prepare a defect report. Detection activity was accomplished in a totally asynchronous way, that is, each inspector had freedom to accomplish that task at the moment he found more convenient, respecting the planning and schedule of his team. After the individual detection, the collection activity took place, where discrepancies were grouped in order to purge replicated discrepancies. Finally, it was generated a list of defects found by each team.

---

[1] https://submissoes.sbc.org.br

Data Collection: All inspections reports, with no duplicate discrepancies, were delivered on time and none of them was discarded. However, only 5 of 10 subjects sent back the follow-up questionnaire with comments regarding the WDP technique. To simplify evaluation, we have considered all the discrepancies in the teams' final lists as real defects, since there was no defects oracle regarding JEMS usability. The number of defects reported by each team is shown in Table 2.

Table 2 - Teams, experience level, employed technique and reported defects

Teams	Experience Level	Technique	# Reported Defects
A	High to Medium	HEV+	62
B	Medium to Low	HEV+	34
C	High	WDP	130
D	Low	WDP	121

Results and Lessons Learned: Analyses of the defects reports gave us some important feedback. From Table 2 we can notice that:

- Teams A and B, despite having similar experience levels, present a significant difference in the number of reported defects.
- Teams C and D, despite differing largely on their experience levels, do not present a significant difference in the number of reported defects.

Team A found almost twice the number of defects found by team B. Assuming that the HEV+ technique is more dependent on inspectors' experience level, this difference may be directly associated with the difference in experience levels between teams A and B. It seems to have a strong correlation between inspectors' experience level and number of defects when applying the technique.

Conversely, teams C and D reported almost the same number of defects. Based on the same reasoning used above, we hypothesize that the WDP technique is not as dependent on inspectors' expertise as a pre-requirement to reach good inspection results. It seems that there could be a weak correlation between inspectors' expertise and number of defects when applying the WDP technique, making it less sensitive to inspectors' experience level than the HEV+ technique.

We did not establish any relationship between the number of defects and the time spent by inspectors, since the long period of inspection and the asynchronous detection process made it impracticable to measure the inspection time precisely. In addition, the experimental object was a real Web application (JEMS), so it was not possible to anticipate all the possible defects one could find (defects oracle), and also if the defects were real or false positives. Due to this, we could not make a detailed analysis of the teams' defects list. For the reasons mentioned above, we decided not to apply any statistical analysis on the quantitative data obtained in the first study.

Finally, the analysis of the five follow-up questionnaires provided us with the feedback summarized in Table 3.

Table 3 - Summary of inspectors' comments on WDP Technique

No.	Comment	Inspector				
		1	2	3	4	5
1.	The technique is feasible.	✓	✓	✓	✓	✓
2.	It is easy to apply the technique.	✓	✓	✓	✓	✓
3.	It is easy to understand Heuristics definitions.	✓	✓	✓	✓	✓
4.	It is easy to understand the Conceptual perspective.	✓	✓	✓	✓	
5.	It is easy to understand the Presentation perspective.	✓	✓	✓	✓	✓
6.	It is easy to understand the Navigation perspective.	✓	✓	✓	✓	✓
7.	It is easy to understand the Structural perspective.			✓		
8.	The Heuristic/Perspective pair facilitates the defect detection.	✓	✓			✓
9.	Technique proposes too many items to check, increasing the inspection time.		✓			
10.	New comments on Heuristic/Perspective pairs are needed.				✓	
11	Some Heuristic/Perspective pairs seem to overlap, making more difficult the classification of the defects.			✓	✓	✓
12.	A rearrangement between Heuristics and Perspective is needed, since some relationships are redundant or do not make sense.					✓

4. Redesigning the WDP Technique

The quantitative and qualitative analyses of the first feasibility study have motivated us to review the WDP technique as follows:

- Conceptual and Structural Perspectives' definitions were reviewed.
- The structural perspective role was modified, because we noticed that this perspective was misunderstood for the inspectors.
- The relationships among heuristics and design perspectives were reviewed in order to reduce redundancy and purge irrelevant ones.
- The heuristics' interpretations regarding each design perspective were modified with the aim of make it clear what kind of defects that relationship relates to.
- The training material was improved with examples associated with heuristics/perspectives pairs from real Web applications.

After a detailed analysis of the first study's results, we observed that the original set proposed by Nielsen covers the purpose of the three additional heuristics adapted from Zhang et al. (1999). Therefore, we eliminated the three additional heuristics, reducing the heuristics to the original set proposed by Nielsen (1994). The hints associated with the eliminated pairs "heuristic-perspective" were replaced by other hints in different pairs "heuristic-perspective". With these modifications, a new version of WDP technique (WDP2) was generated. This version is detailed in Section 4.1.

4.1 WDP Second Version

Table 4 shows all the associations between heuristics and the design perspectives (Conceptual – C, Presentation – P, Navigation – N, Structural – S) in the WDP2. One can notice the differences between the pairs "heuristic-perspective" in version 2.

Table 4 - Relationships between Heuristics and design perspectives in WDP2

#H	Heuristics (original set proposed by [Nielsen 1994])	Web Design Perspectives			
		C	P	N	S
1	Visibility of system status	✓	✓		✓
2	Matching between system and real world	✓	✓		✓
3	User control and freedom			✓	
4	Consistency and standards	✓	✓		
5	Error prevention		✓	✓	
6	Recognition rather than recall		✓		✓
7	Flexibility and efficiency of use		✓	✓	
8	Aesthetic and minimalist project		✓		
9	Help users recognize, diagnose and recover from errors	✓	✓	✓	
10	Help and documentation	✓	✓	✓	

Having Table 4 as a starting point, one inspector can decide if s/he wants to start the inspections by (1) exploring the perspectives for each heuristic OR (2) analyzing the heuristics for each perspective. Tables 5, 6, 7 and 8 show the hints on how to interpret each of WDP2´s heuristics in the light of each design perspective.

Table 5 - Heuristics related to Conceptual (C) Perspective

#H	Hints
1	Evaluate if users can easily understand the system status – that is, if the system status is represented using terms, symbols, words or sentences of the problem domain.
2	Evaluate if the interface uses the problem domain language. Evaluate if the definition of specific domain terms (words or symbols) is accessible.
4	Evaluate if problem domain terms (words or symbols) are presented to the user in a consistent way.
9	Evaluate if error messages provide recovery steps or procedures easily understood by users. Evaluate if error messages are simple and use terms, symbols, words or sentences of problem domain.
10	Evaluate if help and documentation uses terms, symbols, words or sentences of problem domain. Evaluate if the interface provides some kind of help to accomplish complex tasks or to understand non trivial concepts.

Table 6 - Heuristics related to Presentation (P) Perspective

#H	Hints
1	Evaluate if the system status is always visible to the user
2	Evaluate if the information is presented in a logical and natural order Evaluate if terms (words or symbols) presented by the system follow the real world conventions Evaluate if system options are structured and presented to the user in a logical and natural order
4	Evaluate if the terms, graphs and symbols of the interface are consistent Evaluate if the interface adhere to the adopted standards for layout and controls Evaluate if the interface is consistent for equivalent tasks
5	Evaluate if required data are clearly defined to the user Evaluate if data entry format is clearly defined to the user Evaluate if information are presented in a balanced and natural way according to

	the problem domain Evaluate if the interface make it easy to distinct between different tasks and data
6	Evaluate if system interface allows the user to identify key information during a task Evaluate if it is easy to recognize/identify data already provided Evaluate if it is easy to recognize/identify the correct option to reach a goal
7	Evaluate if the arrangement of interface elements minimize users' physical effort Evaluate if the arrangement of interface elements minimize users' visual search Evaluate if the interface supports tasks repeated very often Evaluate if the interface assists the input of simple or complex data Evaluate if the interface provides search mechanisms to assist the input of required data.
8	Evaluate if the relevant data of a task are visible and clear
9	Evaluate if the error messages are presented in a consistent way according to the adopted standards. Evaluate if the user's attention is directed to the error messages Evaluate if the alternative solutions of an error are clear to the user.
10	Evaluate if the user is able to identify the available help options Evaluate if help and documentation aids the user to accomplish a task, as a series of concrete steps, for example.

Table 7 - Heuristics related to Navigation (N) Perspective

#H	Hints
3	Evaluate if the interface allows the user to navigate easily through different steps of a task. Evaluate if the interface has undo/redo or similar functions that allow the user to exit in case of wrong choices. Evaluate if the interface allows the user to return to the main flow of a task after a detour or after performing a secondary task
5	Evaluate if the interface prevents navigation errors, tat is, if the options available defines clearly which results or states will be reached.
7	Evaluate if the interface provides different ways of accessing the main tasks. Evaluate if the interface provides accelerator keys or short-cuts when a user performs the main tasks Evaluate if the accesses provided by the interface minimize the user's physical effort.
9	Evaluate if the system shows how to access alternative solutions presented on error messages.
10	Evaluate if the interface provides an easy way to access help and documentation of a specific task

Table 8 - Heuristics related to Structural (S) Perspective

#H	Hints
1	Evaluate if the interface allows the user to identify which module is being executed
2	Evaluate if the modules organization corresponds to the user's perceived reality
6	Evaluate if the interface allows the user to identify which module has to be used to reach a goal

Since we were unable to establish in the first feasibility study any relationship between the number of detected defects, the total number of existing defects and the time spent by inspectors, we decided to plan and carry out a second feasibility study, detailed in Section 4.2.

4.2 Second Feasibility Study

The experimental study was carried out during November - December 2006 using WDP2. This feasibility study addressed the question "Is the WDP2 technique efficient and effective for Usability inspection?", where efficiency and effectiveness are defined as follows:

- Efficiency: the ratio between the number of detected defects and the time spent in the inspection process.
- Effectiveness: the ratio between the number of detected defects and the total number of existing (known) defects.

To answer this question we compared the efficiency and the effectiveness of WDP2 with the efficiency and the effectiveness of the original Heuristic Evaluation technique (HEV). This controlled experiment is detailed in [Conte et al. 2007] and a summary is presented below:

Subjects: Fourteen subjects were chosen by convenience from three different courses offered at UFRJ (three undergraduate students, nine M.Sc. students and two PhD students). All fourteen subjects signed a consent form and filled out a characterization form that measured their expertise with software inspections, software usability and software engineering in general. The characterization data was used to organize subjects into three groups:

- **Group 1**: represents the control group and includes three subjects (having high expertise in software inspections) who applied an *ad-hoc* inspection technique.
- **Group 2**: six subjects who applied the HEV technique.
- **Group 3**: five subjects who applied the WDP technique.

The control group was created mainly because:
- Three subjects already knew about the research being conducted;
- There was no defects oracle against which the defects found by inspectors could be checked. So, this group was used to check the defects found by the others groups and identify the genuine defects and false positives.

Material: The object of this study was the same application used in the first study (JEMS system). However, due to the large number of use cases, we only selected the use cases associated with the role 'Reviewer'. All the remaining tasks needed to simulate a real conference (e.g. conference creation, deadlines setup, papers submission) were carried out by the authors.

Procedure: All subjects received proper training regarding JEMS and respective technique and had the same time (4 days) to perform the inspection. At the end of the inspection activity, all subjects returned a worksheet with the defects encountered and the total time spent in the inspection and the follow up questionnaire filled out.

Data Collection: All defects' worksheets were checked for incorrect or missing information, and one worksheet (Group 2) was discarded because the subject did not accomplish the work on time. No outlier was eliminated. At the end of the inspection there was a meeting, attended by the control group and two authors, to analyze all discrepancies found for each group. The attendees had no clue about which technique was used and the results of this meeting were as follows:

- A list classifying discrepancies as real defects or false positives.
- A list of replicated defects, that is, equivalent defects detected by different inspectors.
- A list with the known defects, totalizing 90 defects, and 48 false positives.

Results and Lessons Learned: The results of the statistical analysis indicated that WDP2 was significantly more effective than and as efficient as Nielsen's Heuristic Evaluation [Conte et al. 2007]. Table 9 summarizes the results:

Table 9- Effectiveness and efficiency per technique

Tech.	Avg. Time (min)	Efficiency Avg. Defects/Hour	Effectiveness Avg. Defects	Defects ratio %
HEV	139.2	5.34	12.4	13.78
WDP2	216	5.72	20.6	22.89

As planned, we computed two indicators – an effectiveness indicator (Defects ratio), computed as the average number of defects found per group (Avg. Defects) divided by the total number of known defects (90); and an efficiency indicator (Avg. Defects/Hour), computed as the average number of defects found per group divided by the amount of time spend per group. Results show that both techniques were not very effective at finding the total number of defects; however the WDP2 technique was nearly twice as effective as the HEV technique. Their efficiency was similar.

Thus, in terms of our research question:

"Is the WDP2 technique efficient and effective for Usability inspection?"

The data showed that the WDP2 technique was effective for usability inspection, and provided similar efficiency to the HEV technique.

Regarding the qualitative analysis, the data provided the following trends:

- All subjects, except one, considered the WDP2 technique adequate for use. One subject complained about the difficulty of use, because of the large number of heuristic-design perspective pairs.
- One subject argued for a better description of the conceptual perspectives and four complained about the description of the structural perspective. They also reported difficulties to apply the structural perspective.
- Two subjects said that the WDP2 technique made the classification of discrepancies easier to do.

There were various threats that could affect the validity of our results. Regarding the internal validity, we considered three main threats: training effects, classification experience and time measurement. We controlled training effects by preparing equivalent training courses on techniques HEV and WDP2 using the same examples of discrepancies that could be found. We controlled classification experience by comparing the subjects' self classification with our own observation during the courses. Nevertheless, concerning time measurement, we asked subjects to be as precise as possible, but there is no guarantee that the time reported was accurate.

Regarding the external validity, we considered two issues: (1) the probability that students are not good surrogates for real inspectors, and; (2) the use of an academic environment that may not have simulated existing conditions in an industrial development environment. However, JEMS represents a real application and students

may to some extent present similar inspection skills to less experienced practitioners. Carver et al. [24] point out a series of benefits that researchers can gain from empirical studies using students.

Regarding conclusion validity, the small number of data points is not ideal from a statistical point of view. However small sample sizes are a known problem difficult to overcome. Even considering the limitation of the results due to the size of the sample used for the studies, we believe to be possible to associate a positive answer to the first question on Figure 1: "Are the results feasible and time well spent?". After analyzing the results of the second study results, it seemed that the structural perspective did not to fit in with usability evaluations of conventional Web software applications (conventional software applications that use Web's infrastructure for execution [Mendes et al. 2006]). However, it is still necessary to investigate if the structural perspective could be useful for usability inspections of non-conventional Web applications, such as Scientific Workflow-based or e-science applications.

5. Conclusion and Future Work

This paper presented a Web usability inspection technique based on Web design perspectives – WDP. The development of such inspection technique has been supported by a methodology based on experimentation. This experimental methodology organizes the activities in such way that encourages one to examine the larger issues, such as usefulness and cost-efficiency, regarding the technology under development first to avoid wasted effort with larger problems later [Shull et al. 2001]. The results obtained from the experimental studies seem to indicate the WDP2's feasibility and its possibility to be more effective than and as efficient as Heuristics Evaluation [Nielsen, 1994].

Further work entails to use the results of the formal experiment described here to go one step further on the experimental methodology, in order to also answer the second question of the methodology (see Figure 1): "Do the steps of the process make sense?". To investigate this question, we are planning an observational study to be accomplished in 2007. Its purpose is to observe usability inspectors applying the WDP2 technique and to collect data regarding how they apply the technique. The replication of the second feasibility study is being also considered, since we need to be more confident about the effective and efficiency of the WDP2 technique.

Acknowledgements

References

Abrahão, S., Condori-Fernández, N., Olsina, L., Pastor, O. (2003). "Defining and Validating Metrics for Navigational Models", Proceeding of the Ninth International Software Metrics Symposium (METRICS'03).

Baresi, L., Garzotto, F., Paolini, P. (2000), "From Web Sites to Web Applications: New Issues for Conceptual Modeling". In Proc. of the ER'00 Workshop on World Wide Web and Conceptual Modeling, Salt Lake City (USA).

Ceri, S., Fraternali, P., Bongio, A. (2000). "Web Modeling Language (WebML): a Modeling Language for Designing Web Sites". WWW9 Conference, Amsterdam.

Conallen, J. (2002), "Building Web Applications with UML", second edition Addison-Wesley, ISBN: 0-20-173038-3.

Conte T., Massolar, J., Mendes, E., Travassos, G., H. (2007). "Usability Evaluation Based on Web Design Perspective", First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007), Madrid, Spain (accepted for publication).

Conte T., Mendes, E., Travassos, G., H. (2005). "Development Processes for Web Applications: A Systematic Review", in Proceedings of XI Simpósio Brasileiro de Multimídia e Web - WebMedia, Poços de Caldas, Brasil, 2005. (In Portuguese).

Diaz, P., Montero, S., Aedo, I. (2004). "Modelling hypermedia and Web applications: the Ariadne Development Method", Information Systems.

Dix, A., Finlay, J., Abowd, G., Beale, R. (2004) "Human-Computer Interaction", Ed. Pearson/ Prentice Hall, Third Edition.

Fons, J., Pelechano, V., Albert, M., Pastor, O. (2003). "Development of Web Applications from Web Enhanced Conceptual Schemas", Conference on Conceptual Modeling (ER), Is International, 22nd, Il-Yeol Song, Stephen W. Liddle, Tok Wan Ling, Peter Scheuermann, Springer-Verlag, Lecture Notes in Computer Science, 3-540-20299-4, 2813.

Garzotto F., Matera M., Paolini P. (1999). "Abstract Tasks: a tool for the inspection of Web sites and off-line hypermedia", Proceedings of the tenth ACM Conference on Hypertext and hypermedia.

Gómez, J., Cachero, C., Pastor, O. (2001), "Conceptual Modeling of Device-Independent Web Applications", IEEE Multimedia, 8(2), April – June.

ISO (2000), "ISO/IEC FDIS 9126-1: Software Engineering - Product quality – Parts 1-4".

Kappel, G., Michlmayr, E., Pröll, B., Reich, S., Retschitzegger, W. (2004). "Web Engineering - Old wine in new bottles?", Web Engineering, Springer LNCS 3140.

Kitchenham, B. (2004), "Procedures for Performing Systematic Reviews", Keele technical report SE0401 and NICTA technical report 0400011T.1 .

Koch, N., Kraus, A. (2002). "The expressive Power of UML-based Web Engineering". Second Int. Worskhop on Web-oriented Software Technology (IWWOST´02).

Matera, M., Rizzo, F., Carughi, G. (2006). "Web Usability: Principles and Evaluation Methods", In "Web Engineering" (Eds: Emilia Mendes and Nile Mosley), Springer.

Mafra, S. N., Barcelos, R. F., Travassos, G. H. (2006) "Aplicando uma Metodologia Baseada em Evidência na Definição de Novas Tecnologias de Software". In: XX SBES, Florianópolis, SC, Brasil, 2006.

Mendes, E. and Kitchenham, B. (2004), "Protocol for Systematic Review". Available at http://www.cs.auckland.ac.nz/emilia/srspp.pdf. Last accessed by 05/10/2005.

Mendes, E., Mosley, N., Counsell, S. (2006). "The Need for Web Engineering: An Introduction", In "Web Engineering" (Eds: Emilia Mendes and Nile Mosley), Springer.

Mian, P., Conte, T., Natali, A., Biolchini, J., Mendes, E., Travassos, G. (2005), "Lessons Learned on Applying Systematic Reviews to Software Engineering", 3rd Intern. Workshop "GUIDELINES FOR EMPIRICAL WORK" in the Workshop Series on Empirical Software Engineering (WSESE 2005).

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

Offutt, J., "Quality attributes of Web software applications", IEEE Software, 19(2), March/April 2002.

Olsina, L., Covella, G., Rossi, G. (2006). "Web Quality", In "Web Engineering" (Eds: Emilia Mendes and Nile Mosley), Springer, 2006.

Pastor, O. (2004). "Fitting the Pieces of the Web Engineering Puzzle" – Invited Talking, XVIII Simpósio Brasileiro de Engenharia de Software (SBES), Brasília, October.

Pfleeger, S. L. (1999). "Albert Einstein and Empirical Software Engineering", IEEE Computer.

Polson, P., Lewis, C., Rieman, J., Wharton, C. (1992). "Cognitive Walkthroughs: a method for theory-based evaluation of users interfaces", International Journal of Man-Machine Studies.

Pressman, R. (2000). "What a tangled Web we have", IEEE Software, Jan – Feb.

Schwabe, D., Rossi, G., Barbosa, S. (1996). "Systematic Hypermedia Design with OOHDM", ACM Conference on Hypertext.

Shull, F., Carver, J., Travassos, G. (2001), "An Empirical Methodology for Introducing Software Processes", In: Proceedings of the Joint 8th European Software Engineering Conference (ESEC) and 9th ACM SIGSOFT Symposium on the Foundations of Software Engineering (FSE-9).

Zhang, Z., Basili, V., Shneiderman, B. (1999). "Perspective-based Usability Inspection: An Empirical Validation of Efficacy", Empirical Software Engineering: An International Journal, vol. 4(1), March.