International Journal of u- and e- Service, Science and Technology
Vol.10, No.7 (2017), pp.65-82
http://dx.doi.org/10.14257/ijunesst.2017.10.7.07

Validation in the Web Domain of a Formal Process to Evaluate the Usability of Software Applications: An Approach based on the Heuristic Inspection

Freddy Paz[1], César Collazos[2] and José Antonio Pow-Sang[1]

[1]*Pontificia Universidad Católica del Perú, San Miguel, Lima 32, Perú*
[2]*Universidad del Cauca, Popayán, Colombia*
fpaz@pucp.pe, ccollazo@unicauca.edu.co, japowsang@pucp.edu.pe

Abstract

The heuristic evaluation method is one of the most reported methods in the literature to evaluate the usability of software products. However, despite its wide acceptance by the scientific community, there is not a formal framework that can guide step by step to the specialists on how to execute the entire evaluation process. The literature only provides general recommendations that were established by the original author of the technique. However, through an in-depth review of the recent studies which describe the use of this method in the context of software applications, we have developed an own proposal. The new approach consolidates several perspectives from different authors who have used the method in diverse domains. This paper is focused on the validation of our new protocol to conduct heuristic evaluations. For this purpose, an experimental case study was carried out in the Web domain with the participation of twenty postgraduate students. The results establish that more accurate results are obtained when a structured process is followed.

Keywords*: usability, heuristic evaluation, assessment protocol, software development process, experimental case study*

1. Introduction

Nowadays, one of the most important quality attributes of the software products is the usability [1]. The several software alternatives that are offered in the current market for a specific purpose, especially in the e-commerce domain, force the companies to concern about the ease of use of their systems. If a website is difficult to use, the users look for another application immediately that allow them to achieve their goals with satisfaction. Given that companies are progressively adopting a Web model to reach more customers through new information and communication technologies (ICTs), the usability can be a crucial factor. An attractive design and an understandable interface can determine the user's decision to prefer certain websites instead others on Internet, where there are many websites that offer the same products and services and the distance to the competition is as short that can be found by a simple search.

The relevance of the usability had led to the development of some methods that allow specialists to evaluate this quality aspect of any technological interface [7]. The heuristic evaluation is one of the most widely used techniques in the context of software [28]. This recognized method allows the identification of a high number of usability problems by the use of minimal resources [21]. The purpose is to offer solutions to the identified problems in an iterative development process of the interface design.

Although the heuristic evaluation is one of the most commonly reported methods in the literature, there is no agreement on how to perform this usability evaluation process. The original author of the technique only provides general recommendations about the process [20]. However, there is not a specific protocol with well-defined activities that can guide

ISSN: 2005-4246 IJUNESST

step by step the usability evaluation process. In this work, we present the validation of a formal framework to conduct heuristic inspections for the context of software products, which was developed with basis on a previous systematic literature review [6].

This paper is structured as follows. In Section 2, we describe the main concepts and theories that were used for the development of the current study. In Section 3, we describe our new approach to assess the usability of software products. In Section 4, we specify the experimental design of the case study for the validation of our proposal. In Section 5, we present and discuss the results. Finally, the conclusions and opportunities for future works are established in Section 6.

2. Usability and Heuristic Evaluation

Usability is a widely used concept in the field of Human-Computer Interaction (HCI). The definition proposed by the ISO/IEC 9126 is framed in the Software Engineering (SE) and is related to the ease of use and learnability of a software product. In the standard, the usability is defined as a quality attribute that is reflected in "the capability of a computer system to be understood, learned, used and attractive to users" [16]. If the system fails to cover these aspects, it is highly expected that users will look for similar applications that despite offering the same functionality, provide a usable and understandable interface. In the current competitive environment, where there is a large number of available products for the same purpose, the usability becomes essential. The design of an intuitive interface that can allow a friendly interaction for the achievement of the user's goals is nowadays a relevant feature. The interface design must be usable enough to allow users to complete their objectives with effectiveness, efficiency, and satisfaction [15].

The importance of the usability in the computer systems has led to the appearance of a series of methods whose purpose is to determine if the interface of a particular software product is usable. These techniques are known as "usability evaluation methods" (UEM) and allow specialists to conclude systematically if the elements and properties of the interface design contribute to reach the desired degree of usability [3]. The test with users or user testing has always been one of the most employed methods to measure how usable the software interfaces are since it provides direct information about how people use the system to identify the exact problems with a specific interface [12]. However, in order to achieve accurate results, it can be needed the participation of a significant number of users for the application of this method in real scenarios, involving hundreds of users in some cases, depending on the magnitude of the software [25]. Although the participation of few users can be considered as a poor quality assessment, the software development team has to decide the number of users that will participate in the usability testing. The decision will be influenced by the budget and time that is available for the project. Given that the usability is an essential factor for the success of software products, an appropriate choice of the number of users will be based on the in-depth analysis of the costs and benefits.

The heuristic inspection is a usability evaluation method that was developed by Jakob Nielsen as an alternative to the time-consuming and effort-demanding user test [22]. The advantage of this technique over the regular testing with users is that allows the detection of the 75% of the usability problems that are present in a graphical interface involving the participation of only three to five evaluators. However, instead of representative users of the system, a small team of specialists is required for the execution of this technique. The heuristic evaluation method involves having usability experts who examine all the system interfaces and judge if they were properly designed with support on established principles (known as "heuristics") [21].

The usability evaluation based on heuristics has proved to be an effective method over time, in such a way that there are numerous studies in the literature which report its usage, especially, as part of the software development process [7]. The heuristic method provides specialists with the possibility to identify potential usability issues in an assessment where

only three evaluators are required. In comparison to other methods, where many resources are essential regarding users, costs and time, this technique turns out to be an economical option to find the most significant usability problems of a graphical user interface in a fast and accurate way.

2.1. Traditional Nielsen's Approach

The original Nielsen's proposal to carry out a heuristic evaluation does not establish concrete steps to follow. On the contrary, it only provides general recommendations that are used to guide the inspection process. The lack of a formal methodological procedure has led to several authors to establish particular ways of conducting this type of evaluation. However, in this study, we present the validation of a general protocol that has been developed considering the different current perspectives.

Although the scholars have defined innovative mechanisms in the assessment process, the Nielsen's approach is still the most representative framework for the execution of heuristic evaluations in the context of software products. In some cases, the Nielsen's guidelines are wrongly interpreted by the authors, leading to errors that then are visible in the procedure is employed. In other cases, some improvements are considered in order to increase the accuracy of the obtained results. The general guidelines established as part of the original approach are presented as follows [21]:

- The assessment team must be defined by a set of five or at least three specialists. According to a statistical exploration of Nielsen [23], single evaluators find only 35% of the usability problems in the system interfaces. However, if the results from different evaluators are aggregated, it is possible to achieve better outcomes. The analysis of several projects shows that when more than five specialists take part of the evaluation, the percentage of new usability problems that are identified is low. Generally, the new findings are of little relevance, or most of them have been already recognized by the team of five evaluators. In the same way, if less than three evaluators participate of the assessment, there is a high probability of missing most of the critical problems in an interface.
- The evaluation team must employ the ten heuristics proposed by Nielsen for the detection of usability issues [19]. The specialists examine the interface in order to determine if each design element follows the established usability principles. In case a heuristic is infringed by the interface design, the issue is documented as a usability problem. The goal is to identify these problems and attend them as part of an iterative design process.
- The participation of usability specialists is explicitly recommended. Although the heuristic evaluation can be performed by people with little or no usability expertise, which is a disadvantage from the usability engineering perspective, the involvement of usability specialists as evaluators is highly preferable.
- At the beginning of the evaluation process, each usability specialist must examine the interface individually. The evaluator goes through the interface several times to determine if there are elements that infringe the set of principles. The outcome is a list of usability problems, annotated with references to the principles that are violated by the design. Only after all individual evaluations have been completed, the evaluators are allowed to communicate and consolidate their findings.

3. A Formal Process to Evaluate the Usability

The new approach to evaluate the usability of software applications was proposed with basis on a comprehensive analysis of all the recent studies which report the use of the heuristic inspection method. In a previous work [6], we conducted a systematic literature review to identify the most used methods to measure the degree of usability of software products. The heuristic evaluation was the third most relevant technique after the user

testing and the questionnaire. From 215 papers that were selected for the review process. Fifty of them support and describe the usage of this method in the software development process. The systematic review initially covered a spectrum of three years, considering studies from 2012 to 2015. However, the search results were updated to include in our analysis, papers that were published until November 2016. In order to define a formal process with clear instructions to follow, a total of seventy-one studies were examined. From the resulting papers of our systematic literature review, only those that illustrated in a way or another the use of the heuristic evaluation were considered. The list of papers is detailed in [6], but the new studies from 2016 that were incorporated to our analysis are [4], [5], [24], [29], [18], [14], [2] and [31]. The new evaluation process was structured in five phases according to the framework established by Hurtado *et al.* [27] for usability inspection methods. In accordance with the different perspectives that are established in the studies to carry out heuristic evaluations in the software context, we have developed a formal protocol that is presented as follows:

1. **Planning:** The proposal demands the role of *evaluation manager* who will be the responsible for leading the heuristic inspection of the software product. Evaluation managers must have a high level of expertise performing this type of evaluations since several of their planning decisions could affect the quality of the results. The purpose of this phase is to design the inspection by establishing some criteria such as the profile and the number of evaluators. The planning process is illustrated in Figure 1. The specific activities are:
   1.1. The evaluation manager must define the profile that the specialists must meet to participate as evaluators in the heuristic evaluation of the software system. According to the studies that were part of the systematic review, there are three type of professionals that are commonly considered: (1) usability novices, with general knowledge about computing but without special usability expertise, (2) single experts, who are usability specialists but not specialized in the software domain, and (3) double experts, with expertise in both usability and the domain of the system. Although, double experts are preferred, the decision will depend on the budget and costs that were estimated for the project, as well as on the availability of the possible evaluation team.
   1.2. The evaluation manager must determine the number of specialists that will constitute the assessment team of the software product. According to Nielsen, only from three to five evaluators are required to obtain valuable information about the usability of the interface design. Around 75% of the entire issues can be identified when three specialists participate in the evaluation. However, in those studies that state the use of the heuristic evaluation method, the reported number of inspectors is extremely varied. Although the majority of authors tends to follow the Nielsen's suggestion about the involvement of three to five specialists, there are scenarios in which the evaluation was conducted by only one specialist. In the same way, some studies report the participation of fifteen and even more evaluators. It seems apparent that the more specialists involved, the greater the number of problems that are identified. Nevertheless, from the entire set of usability issues that are present in an interface design, only the most severe of them are priority aspects. If most of the relevant problems can be detected with only three specialists, then the Nielsen's recommendation is an advantageous option to be considered, especially for small projects where the budget is limited. More evaluators should obviously be used in scenarios where usability is critical or when large workflows are expected.
   1.3. The evaluation manager should prepare a general description of the software product to be evaluated in terms of usability. This information will be used to provide the evaluation team with an overall idea about the system.

1.4. Finally, the evaluation manager should recruit the professionals that will form the assessment team according to the established criteria in step 1.1. This team of specialists will be responsible for examining the graphical system interfaces to identify design problems based on the usability heuristics.

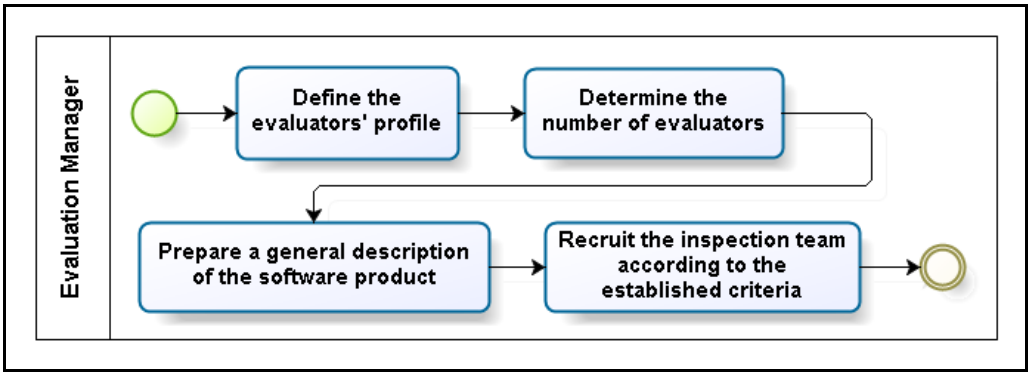

Figure 1. Planning Phase of the Usability Evaluation Process

2. **Training:** If the professionals that were recruited for the usability evaluation of the software product have little or no experience in the process to follow or in the main concepts of usability, they must receive appropriate training on these topics. All the team members should become familiar with the framework we are proposing for obtaining accurate results. The general description performed in step 1.3 along with free exploration of the system interfaces will help the evaluation team to determine the proper set of usability heuristics that must be considered for the inspection. The training process is illustrated in Figure 2, and the specific tasks are described as follows:
   2.1. A special session of training must be prepared for each member of the usability assessment team that has little or no previous experience performing heuristic evaluations. The instruction must be oriented to explain and describe the main concepts of usability and the fundamental purpose of the evaluation method.
   2.2. All members of the evaluation team must be aware of this entire framework that will guide the usability inspection of the software product. The evaluation manager should inform about the workflow explaining each phase along with the goals and the criteria that must be followed in each step.
   2.3. The overall description of the software product that was developed in step 1.3 will be used to provide the team a general idea of the upcoming inspection to be conducted. The specialists can perform a free exploration of the system interfaces to obtain relevant information about the software that can afterward be employed as a reference for the selection of the appropriate set of heuristics.
   2.4. The evaluation team must come together to determine the most appropriate set of usability heuristics according to the specific software domain. Several works state that the ten usability principles proposed by Nielsen fail to cover relevant aspects of usability when they are used to evaluate new emerging categories of software systems such as e-commerce applications [17], transactional websites [9], applications for touchscreen-based mobile devices [30], videogames [13], banking systems [26], *etc.* There is enough evidence in the literature proving that the conventional Nielsen's heuristics are no longer suitable for the kinds systems that currently exist [10]. The problem with the traditional guidelines is that they do not consider features which are specific to the application domain. For instance, an e-learning platform must be analyzed by considering how the established usability degree contributes to obtain the desired educational goals [20]. In this sense, there are several proposals to evaluate the usability in line

with the perspective of the particular domain. The team will be responsible for selecting the most appropriate heuristics according to the software product.

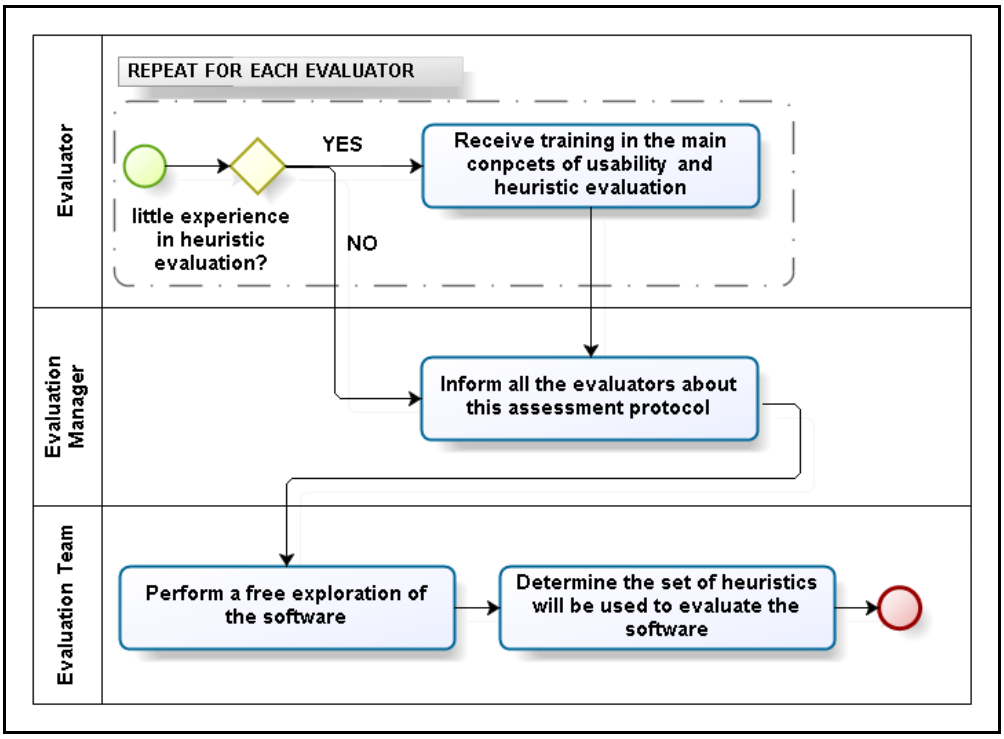

Figure 2. Training Phase of the Usability Evaluation Process

3. **Evaluation:** In this phase, the team members proceed to examine individually each of the system interfaces that were considered for evaluation. The professionals that were selected for this activity must work alone to ensure independent and unbiased results. The purpose of this phase is for the evaluators to identify infringements to the established heuristics which are produced because of a wrong interface design. Subsequently, the infractions to the usability principles are categorized as usability problems. The final result of this phase is a list of usability issues for each member of the team that can be documented using a template.

   3.1. The heuristic evaluation involves having a group of specialists who determine if the graphical user interfaces of a software product are usable and have been correctly designed. In order to perform this inspection, the specialists must use a set of usability principles. The goal is to identify those elements from the interface design that infringe the usability guidelines. In this sense, each evaluator must examine carefully and individually the selected interfaces to identify the usability problems that the proposed design presents. The reason for an individual assessment is the impartiality and neutrality of the results. Depending on the expertise of each evaluator, the inspection process can differ one from another. Specialists with experience in the use of the established heuristics will identify the problems at first, and then they will associate the identified issues with the corresponding infringed heuristics. However, novice evaluators will initially become familiar with the usability principles to begin with the examination of the system interfaces. Novice evaluators have two ways to proceed: (1) they can identify problems as they review each principle, or (2) they can review the entire list of principles and identify problems according to what they have learned. The findings should be documented in a template, completing the following information for each identified problem:

(a) ID, (b) definition, (c) comments/explanation, (d) occurrence examples, (e) infringed heuristic and (f) screenshots that can evidence the problem. The output of this phase is a list of usability problems for each member of the evaluation team. The process is illustrated in Figure 3.

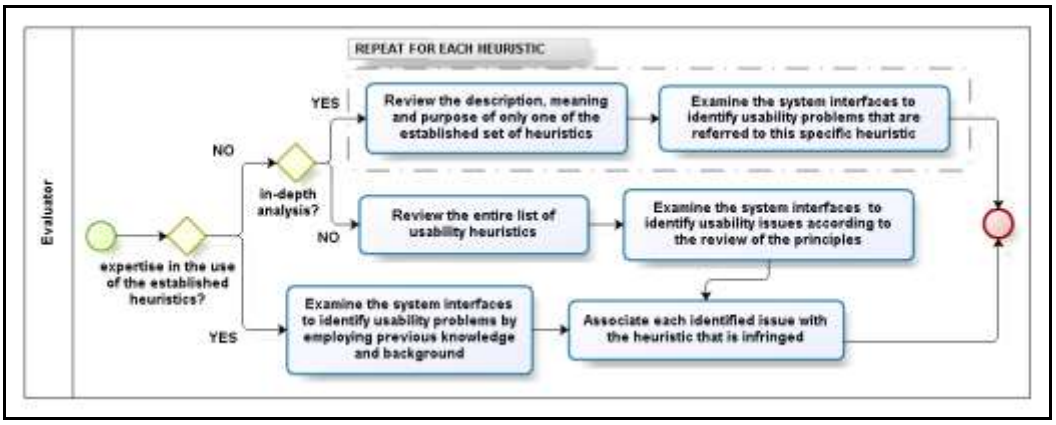

Figure 3. Evaluation Phase of the Usability Evaluation Process

4. **Discussion:** Once the individual evaluations are completed, the team must prepare collaboratively a single list of usability problems. In order to reach a consensus, the team must organize a meeting where all findings must be discussed. In this session, the team will discard the repeated issues and those which indeed do not represent a usability problem. The session will be led by a moderator whose responsibility will be to promote a productive discussion and an open debate, where the participation of each team member is equal and valued. The evaluation process is described as an algorithm in Figure 4.

4.1. The moderator verbalizes the usability issue.
4.2. The evaluator who identified the issue provides more details about the problem, even with the help of screenshots.
4.3. The team will discuss if the issue is indeed a usability problem. For this, each evaluator will give an opinion about the identified problem.
If the group agrees that it is a usability problem **then,**
4.3.1. The team has to determine if the issue was identified by more than one evaluator.
If the issue was identified by more than evaluator **then,**
4.3.1.1. The evaluators who identified the issue must discuss to determine if they have indeed identified the same incidence or if they are making reference to different aspects.
If they are making reference to the same issue **then,**
4.3.1.1.1. They must prepare collaboratively a definition and description of the issue.
Else
4.3.1.1.2. The new incidence should be classified as usability problem to be discussed afterwards.
4.3.1.2. The team has to determine if the issue, in the way it was written, express indeed the incidence to which it is making reference.
If the problem is written properly **then,**
4.3.1.2.1. The problem is included in the final list.
Else
4.3.1.2.2. The team collaboratively must prepare a definition and description of the identified problem.
Else
4.3.1.3. Go to step **4.3.1.2.**
Else
4.3.2. The evaluator who identified the issue should decide after having listening to the different points of view if he still believes it is a usability problem.
If the evaluator persists in the idea that the issue is a usability issue **then,**
4.3.2.1. Go to step **4.3.1.2.**
Else

Figure 4. Discussion Phase of the Usability Evaluation Process

5. **Report:** Once the evaluation team has established the single list of problems, they must rate individually the severity and frequency of each problem that is in the list. After this activity, they must come together again to average the individual scores. An analysis of the standard deviation will be convenient to determine if there are significant differences in the scores that were given to each identified problem. If the results are sorted by severity and frequency, it is possible to identify the most critical problems that could have a negative impact on user satisfaction. Finally, the team must document all findings in a report providing possible solutions to the problems that have been identified. The process is illustrated in Figure 5.
   5.1. Once the evaluation team has defined the single list of problems, the evaluation manager must send the unified list to all the members. The purpose is that each evaluator could then rate the severity and frequency of the identified problems.
   5.2. The evaluators proceed to rate individually the severity and frequency of each problem that was established in the single list. The degree of severity is related to the impact of the problem on the user experience. Given that the usability is nowadays a key factor for the success of a software product, some problems could have a devastating effect on the popularity of the system. The purpose of this activity is to support prioritizing and decision-making in order to attend, in the first instance, the most relevant problems. In the same way, the frequency is related to how often the problem appears in the interface. Some authors have

defined their own scales to rate the severity and frequency of the problems. However, the most reported approach in the literature is the Nielsen's proposal [21]. The severity ratings established by Nielsen are presented in Table I, and an own proposal to rate the frequency is presented in Table II [12].

Table 1. Severity Ratings for Usability Problems

Rating scale	Severity of the usability problem
0	I don't agree that this is a usability problem at all.
1	**Cosmetic problem only**: need not be fixed unless extra time is available on project.
2	**Minor usability problem**: fixing this should be given low priority.
3	**Major usability problem**: important to fix, so should be given high priority.
4	**Usability catastrophe**: imperative to fix this before product can be released.

Table 2. Frequency Ratings for Usability Problems

Rating scale	Frequency of the usability problem
0	< 1%
1	1 – 10%
2	11 – 50%
3	51 – 90%
4	> 90%

5.3. The evaluators must meet again to average the individual scores of severity and frequency of the problems. This activity will allow the evaluation team to find the most critical problems, with high priority to be solved in the next iteration of the interface design process.

5.4. The evaluators must calculate the standard deviation of the individual scores in the severity and frequency to determine if the opinions are one from each other markedly different. If the value of the standard deviation is high, then there are significant differences in the perspective of how severe or frequent the problem is. In these cases, the evaluation team must discuss the reasons for the assigned values in order to establish a consensus regarding how critical the problem is.

5.5. A usability evaluation must be conceived as a productive activity whose results are positive suggestions to improve the interfaces of the software product, in addition to the corresponding usability problems. Therefore, a relevant task for the evaluation team is to provide possible solutions to the design problems that were identified in the system interfaces.

5.6. A usability evaluation should not be perceived as a negative criticism of the system interfaces. Although the main purpose of this inspection is to identify usability problems, the evaluation team must also remark the positive aspects of the proposed interfaces. If there are good design decisions, these must be highlighted by the evaluators.

5.7. The evaluation team must prepare a final report which describes all findings and results of the inspection process. The document must contain the following information: a description of the evaluated software product, the methodology that was used for the evaluation, the unified list of usability problems, the profile of the evaluation team members, the infringed heuristics, the most

critical issues, the possible solutions to the problems, and the positive elements of the interfaces.

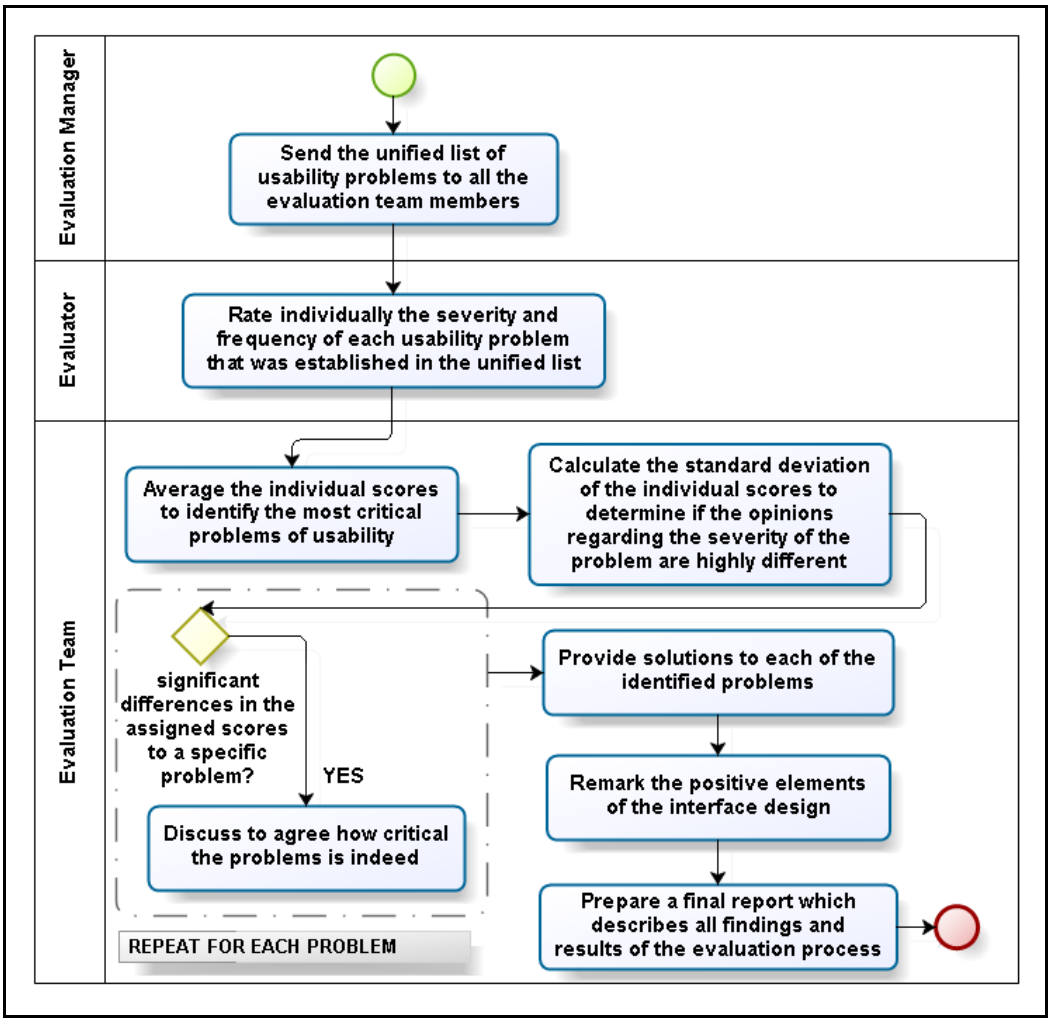

Figure 5. Report Phase of the Usability Evaluation Process

4. Experimental Case Study

In order to validate the new methodological process based on the heuristic inspection to evaluate the usability of software products, we performed a comparative study, where the results of using the traditional Nielsen's approach were contrasted with the application of the new formal evaluation protocol. The case study was conducted in an academic context with the participation of twenty post-graduate students of the Pontifical Catholic University of Peru (PUCP), located in Peru. The students were randomly selected from a mandatory subject about "Human-Computer Interaction" of the Master's Degree Program in Informatics Engineering. Their expertise regarding the knowledge in Computer Science was similar since all of them have attended the same courses of the curricula. However, none of them had previous experience performing a usability evaluation.

The validation of our new proposal was carried out with the voluntary participation of the students. They kindly agreed to contribute without expecting compensation for their involvement in the case study. In this way, we previously informed that their performance would not affect their grades in the course to minimize external factors that could affect

the validity of the results. The experimental study, whose design is illustrated in Figure 6, was conducted in November 2016.

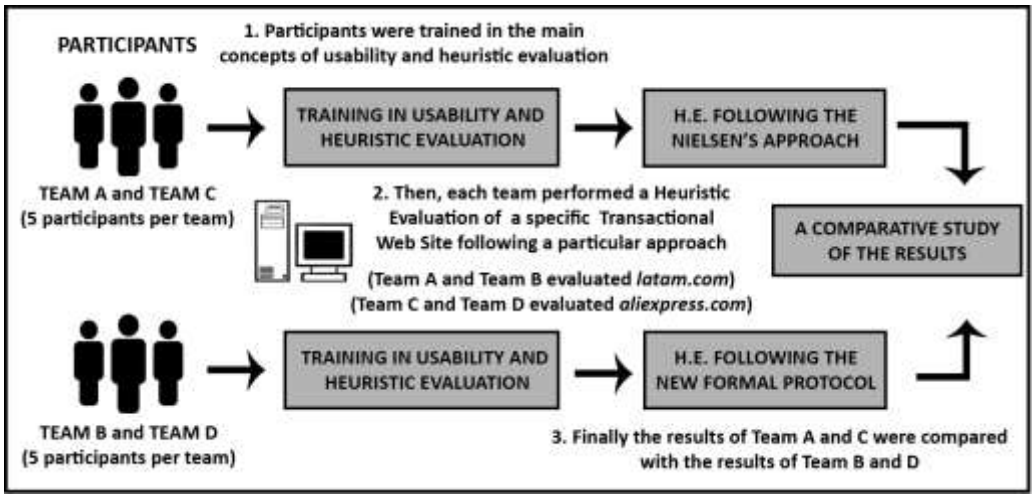

Figure 6. Research Design

The execution of the heuristic evaluation method was one of the scheduled activities as part of the course in which the students were enrolled. This academic scenario facilitated the conduction of a validation process for the new usability evaluation framework. First, all the students were trained in the main concepts of usability. Once they became familiar with the fundamental definitions and the purpose of the usability evaluation methods, the students were divided into four teams of five people. Two of the teams were trained in the heuristic evaluation method according to the original approach proposed by Nielsen [20]. The other two teams were trained with basis on the new formal process for the usability evaluation of software products, which is described in this study. Followed by this step, each team was requested to perform a heuristic evaluation. Two e-commerce websites were selected for this activity: the website of an airline (*www.latam.com*) and the website of a retail store (*www.aliexpress.com*). In order to obtain results that are comparable, each selected website was assessed by both approaches, the traditional Nielsen's proposal and our new framework for usability evaluations. The instrument-subject distribution for the heuristics evaluations is presented in Table 3. Team A and C, that received training on the conventional guidelines, assessed respectively *latam.com* and *aliexpress.com*. In the same way, team B and D whose training was focused on the new usability evaluation approach, examined respectively the same websites. The evaluations were performed independently by each team in three sessions of three hours (a total of nine hours per team). As final step, once the teams completed the entire process of evaluation, the results were analyzed and compared.

Table 3. Instrument-Subject Distribution of the Experiment

Evaluation Team	Website	Evaluation Process
Team A	www.latam.com	Nielsen's traditional approach
Team B	www.latam.com	New evaluation process
Team C	www.aliexpress.com	Nielsen's traditional approach
Team D	www.aliexpress.com	New evaluation process

5. Analysis of the Results

The students proceeded to evaluate the usability of the assigned websites according to the established approaches. Following the general recommendations proposed by Nielsen, teams A and C employed the ten traditional usability heuristics for user interface design. However, teams B and D organized a meeting to determine the proper set of principles for the specific domain of e-commerce. After a brief analysis of the web systems features, the students determined that the conventional heuristics were not an accurate tool to evaluate all the usability aspects that the graphical user interfaces presented. A quick review of the literature allows the teams to conclude that the fifteen usability heuristics proposed by Paz *et al.* [11] were a better alternative to cover elements related to culture, interactivity, real-time processing, transactional processes, and system functionality, that in some way affect the degree of usability of the software product. The set of principles that was employed by each team is described in Table 4.

Table 4. Sets of Usability Heuristics Employed by Each Team

Evaluation Team	Usability Heuristics
Team A	Usability principles proposed by Nielsen (ten usability heuristics for the interface design of nonspecific software products).
Team C	
Team B	Usability principles proposed by Paz (fifteen usability heuristics for the interface design of transactional websites).
Team D	

The results of the heuristic evaluations provided by both approaches, were compared in three categories: (1) the number of identified problems, (2) the severity of the identified problems, and (3) the number of errors in associations. The final reports from all teams were consolidated in order to perform the comparative analysis and the discussion of the results.

Table 5 presents the number of identified problems by each team as well as the severity average of the issues. According to the results, more usability problems are detected in the interface design when our new methodological procedure is employed. The execution of the heuristic inspection method by following predefined and clear steps allows a more in-depth analysis of the usability. In addition, the problems identified by the teams who used the new evaluation process turned out to be more critical for the system success. The new protocol allows specialist to perform the usability assessment in a more structured way, in which all findings are discussed to obtain accurate and relevant results. The agreement of an appropriate set of principles for the evaluation is another factor that can impact on the outcomes of the inspection. The new valuation process has been developed in a way that can be used by any professional in the field of Computer Science. Table 6 and 7 describes the five most critical problems that were only identified by those teams who used our new approach. The results establish that these issues are relevant and should not be overlooked in a usability evaluation. However, when evaluators only follow general guidelines, these features are not considered. The quality of the results increases if a formal methodology is used to carry out this type of inspection.

Table 5. Results of the Heuristic Evaluations

Evaluation Team	Number of Identified Problems	Severity Average of the Identified Problems
Team A	22	2.37
Team B	37	2.65
Team C	17	2.38
Team D	28	2.80

Table 6. Five of the Most Critical Problems that were only Identified by the Team B and not by Team A

ID	Problem Definition	Severity
P17	The system displays a blank screen when the user logs off.	3.25
P26	The button that allows users to return to the previous step during the payment process with VISA cancels the entire process.	3.25
P13	There is no help system to support users with clear instructions on how to use the website.	3.00
P35	There are misspellings in the displayed information.	3.00
P19	The system is not adapted to be used with different Web browsers and screen dimensions.	3.00

Table 7. Five of the Most Critical Problems that were only Identified by the Team D and not by Team C

ID	Problem Definition	Severity
P04	The option to log out is difficult to recognize.	3.3
P02	The help system displays only displays the information in English.	2.8
P03	There is not an option to change the currency and display the prices according to the local context.	2.3
P14	The system interfaces have a different design in some sections.	2.0
P10	Several options of the system force the user to register.	2.0

Another aspect that was examined in this comparative analysis is the number of correct and incorrect associations that are performed when a specific approach is used. When the evaluators identify usability problems, they associate the issue with the infringement of a particular heuristic. If the principles are misinterpreted, the evaluators have detected false positives or problems that are indeed related to other aspects. A valid association is when the inspectors associate correctly the identified usability problem with the heuristic that is infringed by the graphical user interface [8]. In the same way, a wrong association is when the inspector specifies the infringement of a principle that is not actually related to the issue that was identified. In order to identify the correct and incorrect associations, we requested the participation of two external professors of the National University of Cauca, located in Colombia. These specialists analyzed the association of each problem that was reported by the teams. According to the results of this analysis that are presented in Table 8, more correct associations are performed when a formal methodology is employed. The execution of a discussion phase allows the evaluation team to have a detailed examination of each identified issue for the achievement of better findings.

Table 8. Correct and Incorrect Associations per Heuristic and Approach

Team A and Team C (Nielsen's Approach)				Team B and Team D (New Evaluation Process)			
	Total Problems	**Correct Associations**	**Incorrect Associations**		**Total Problems**	**Correct Associations**	**Incorrect Associations**
N01	9	6	3	T01	8	7	1
N02	5	3	2	T02	4	4	0
N03	7	5	2	T03	6	5	1
N04	7	4	3	T04	3	3	0
N05	1	1	0	T05	4	4	0
N06	2	1	1	T06	7	5	2
N07	3	2	1	T07	3	3	0
N08	1	1	0	T08	2	2	0
N09	1	1	0	T09	6	5	1
N10	3	2	1	T10	5	1	4
-	-	-	-	T11	2	2	0
-	-	-	-	T12	2	2	0
-	-	-	-	T13	6	6	0
-	-	-	-	T14	4	2	2
-	-	-	-	T15	4	3	1
Total	**39 (100%)**	**26 (66.67%)**	**13 (33.33%)**	**Total**	**66 (100%)**	**54 (81.82%)**	**12 (18.18%)**

6. Conclusions and Future Works

The heuristic evaluation is one of the most recognized techniques to assess the usability of software products. In a previous study, we noticed this method was commonly reported in the literature. However, despite its popularity, there was not a formally defined process that allows specialists to execute this technique in a structured way. We could determine that each researcher performs a different interpretation of the original recommendations that were established by Jakob Nielsen. In this sense, we did an exhaustive review of the different perspectives the authors establish to carry out this method. A systematic review of the recent studies which report the use of the heuristic evaluation method allowed us to consolidate the different perspectives and define a methodical process. The intention was to develop a proposal that can be easily used by any professional in the field of Computer Science. The framework was implemented, and in this study, we present its validation.

The validation of our proposal to measure the usability of software products involved a comparative study in which the results of the use of the new process were contrasted with the results of applying the general guidelines of the original author of the technique. The experimental case study was performed with the participation of twenty postgraduate students of the Master's Degree Program in Informatics Engineering from the Pontifical Catholic University of Peru, who voluntarily agreed to contribute with this research. The results showed that more usability problems can be detected when a formal methodology guides the inspection process. Moreover, the issues that were identified by the teams who employed our new proposal were more critical for the system success. Important aspects of usability can only be considered if a formal and structured procedure is followed and a proper set of principles is selected. Traditional Nielsen's heuristics fail to cover relevant aspects of the e-commerce domain such as real-time processing, culture, functionality, and interactivity. Therefore, the selection of the usability principles will have a substantial impact on the quality of the final results. Finally, we highlight the fact that the evaluators establish more valid associations between identified issue with infringed principles when the new usability evaluations process is used.

Although the results are positive for our new proposal, more validations are required to generalize its usefulness, ease of use, effectiveness and efficiency. The evaluation process must be proved in other domains such as applications for mobile devices, videogames, e-learning platforms and specialized software. The execution of more experimental cases studies will allow to improve the proposed methodology in contribution to the scientific community of the Human-Computer Interaction area.

Acknowledgments

This paper is a revised and expanded version of the paper entitled *Validation of a Usability Evaluation Protocol based on the Heuristic Inspection Method: An Experimental Case Study in the Web Domain* presented at *9th International Conference on u- and e- Service, Science and Technology, Jeju Island, Korea, December 21-23, 2016.*

References

[1] A. Dhouib, A. Trabelsi, C. Kolski and M. Neji, "A classification and comparison of usability evaluation methods for interactive adaptive systems", Proceedings of the 9th International Conference on Human System Interactions (HIS 2016), Portsmouth, United Kingdom, **(2016)**.
[2] A. Dirin and M. Nieminen, "Relevance of UCD education to software development – recommendation for curriculum design", Proceedings of the 8th International Conference on Computer Supported Education (CSEDU 2016), Rome, Italy, **(2016)**.
[3] A. Fernandez, E. Insfran and S. Abrahão, "Usability evaluation methods for the web: A systematic mapping study", Information and Software Technology, vol. 53, no. 8, **(2011)**, pp. 789-817.
[4] B. A. Myers, A. J. Ko, T. D. LaToza, and Y. Yoon, "Programmers are users too: Human-centered methods for improving programming tools", Computer, vol. 49, no. 7, **(2016)**, pp. 44-52.
[5] C. Altin Gumussoy, "Usability guideline for banking software design", Computers in Human Behavior, vol. 62, **(2016)**, pp. 277-285.
[6] F. Paz and J. A. Pow-Sang, "A systematic mapping review of usability evaluation methods for software development process", International Journal of Software Engineering and Its Applications, vol. 10, no. 1, **(2016)**, pp. 165-178.
[7] F. Paz and J. A. Pow-Sang, "Current trends in usability evaluation methods: A systematic review", Proceedings of the 7th International Conference on Advanced Software Engineering and Its Applications (ASEA 2014), Hainan, China, **(2014)**.
[8] F. Paz, F. A. Paz and J. A. Pow-Sang, "Comparing the effectiveness and accuracy of new usability heuristics", Proceedings of the 7th International Conference on Applied Human Factors and Ergonomics (AHFE 2016), Orlando, FL, USA, **(2016)**.
[9] F. Paz, F. A. Paz and J. A. Pow-Sang, "Evaluation of usability heuristics for transactional web sites: A comparative study", Proceedings of the 13th International Conference on Information Technology: New Generations (ITNG 2016), Las Vegas, NV, USA, **(2016)**.
[10] F. Paz, F. A. Paz and J. A. Pow-Sang, "Experimental case study of new usability heuristics", Proceedings of the 17th International Conference on Human-Computer Interaction (HCI International 2015), Los Angeles, CA, USA, **(2015)**.
[11] F. Paz, F. A. Paz and J. A. Pow Sang, "Usability heuristics for transactional web sites", Proceedings of the 11th International Conference on Information Technology: New Generations (ITNG 2014), Las Vegas, NV, USA, **(2014)**.
[12] F. Paz, F. A. Paz, D. Villanueva and J. A. Pow-Sang, "Heuristic evaluation as a complement to usability testing: A case study in web domain", Proceedings of the 12th International Conference on Information Technology: New Generations (ITNG 2015), Las Vegas, NV, USA, **(2015)**.
[13] G. F. Tondello, D. L. Kappen, E. D. Mekler, M. Ganaba and L. E. Nacke, "Heuristic evaluation for gameful design", Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion (CHI PLAY 2016), Austin, TX, USA, **(2016)**.
[14] I. Nascimento, W. Silva, B. Gadelha, and T. Conte, "Userbility: A technique for the evaluation of user experience and usability on mobile applications", Proceedings of the 18th International Conference on Human-Computer Interaction (HCI International 2016), Toronto, Canada, **(2016)**.
[15] ISO, "Ergonomic requirements for office work with visual display terminals (vdts), -- part 11: Guidance of usability", International Organization for Standardization, Geneva, Switzerland, ISO 9241-11:1998, **(1998)**.
[16] ISO, "Software engineering – product quality – part 1: Quality model", International Organization for Standardization, Geneva, Switzerland, ISO 9126-1:2001, **(2001)**.
[17] J. Díaz, C. Rusu and C. Collazos, "Experimental validation of a set of cultural-oriented usability heuristics: e-Commerce websites evaluation", Computer Standards & Interfaces, vol. 50, **(2017)**, pp. 160-178.

[18] J. M. I. de Carvalho, T. S. da Silva, and M. S. Silveira, "Agile and UCD integration based on pre-development usability evaluations: An experience report", Proceedings of the 18th International Conference on Human-Computer Interaction (HCI International 2016), Toronto, Canada, **(2016)**.
[19] J. Nielsen, "Enhancing the explanatory power of usability heuristics", Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA, **(1994)**.
[20] J. Nielsen, "How to conduct a heuristic evaluation", https://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation, **(1995)**.
[21] J. Nielsen, "Usability engineering", Academic Press, San Diego, CA, USA, **(1993)**.
[22] J. Nielsen and R. Molich, "Heuristic evaluation of user interfaces", Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Seattle, WA, USA, **(1990)**.
[23] J. Nielsen and T. K. Landauer, "A mathematical model of the finding of usability problems", Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems, Amsterdam, The Netherlands, **(1993)**.
[24] N. Alomar, N. Almobarak, S. Alkoblan, S. Alhozaimy, and S. Alharbi, "Usability engineering of agile software project management tools", Proceedings of the 18th International Conference on Human-Computer Interaction (HCI International 2016), Toronto, Canada, **(2016)**
[25] N. Bevan, C. Barnum, G. Cockton, J. Nielsen, J. Spool and D. Wixon, "The "magic number 5": Is it enough for web testing?", Proceedings of the Conference on Human Factors in Computing Systems, Fort Lauderdale, FL, USA, **(2003)**.
[26] N. Fierro and C. Zapata, "Usability heuristics for web banking", Proceedings of the 18th International Conference on Human-Computer Interaction (HCI International 2016), Toronto, Canada, **(2016)**.
[27] N. Hurtado, M. Ruiz, E. Orta and J. Torres, "Using simulation to aid decision making in managing the usability evaluation process", Information and Software Technology, vol. 57, **(2015)**, pp. 509-526.
[28] N. Vatankhah, K. T. Wei and S. Letchmunan, "Usability measurement of Malaysian online tourism websites", International Journal of Software Engineering and Its Applications, vol. 8, no. 12, **(2014)**, pp. 1-18.
[29] R. Cavichi de Freitas, L. A. Rodrigues, and A. Marques da Cunha, "AGILUS: A method for integrating usability evaluations on agile software development", Proceedings of the 18th International Conference on Human-Computer Interaction (HCI International 2016), Toronto, Canada, **(2016)**.
[30] R. Inostroza, C. Rusu, S. Roncagliolo, C. Jimenez and V. Rusu, "Usability heuristics validation through empirical evidences: A touchscreen-based mobile devices proposal", Proceedings of the 31st Conference of the Chilean Computer Science Society (SCCC 2012), Valparaiso, Chile, **(2012)**.
[31] R. Inostroza, C. Rusu, S. Roncagliolo, V. Rusu, and C. A. Collazos, "Developing SMASH: A set of smartphone's usability heuristics", Computer Standards & Interfaces, vol. 43, **(2016)**, pp. 40-52.

Authors

Freddy Paz, is part-time professor at Pontificia Universidad Católica del Perú (PUCP). His research interests include human-computer interaction and programming languages. He received his Master Degree in Informatics and Informatics Engineering from PUCP – Peru, and Pontificia Universidad Católica de Valparaíso (PUCV) – Chile. He is currently a doctoral student in Engineering. He has a BSc degree in System Engineering from Universidad Nacional Pedro Ruiz Gallo (UNPRG), Peru.

César Collazos, is full professor at Universidad del Cauca – Colombia. His research interests include human-computer interaction, software engineering education, collaborative learning, and gaming. He is Ph.D. in Computer Science from Universidad de Chile – Chile. He is currently head of the IDIS Research Group in Colombia, a senior member of IEEE Computer Society, and a member of AIPO (Latin American Association of Human-Computer Interaction).

José Antonio Pow-Sang, is full professor and executive director of the Postgraduate School at Pontificia Universidad Católica del Perú (PUCP). His research interests include empirical software engineering, software metrics, software engineering education, and human-computer interaction. He is Ph.D. in Informatics Engineering, and Master in Software Engineering from Universidad Politécnica de Madrid (UPM) – Spain. He is currently a senior member of IEEE Computer Society and a member of ACM.