

Comparing the Effectiveness and Accuracy of New Usability Heuristics

Freddy Paz, Freddy Asrael Paz and José Antonio Pow-Sang

Abstract Nielsen's usability heuristics are the most recognized assessment tool to conduct heuristic evaluations. However, in a previous study we demonstrated these principles fail to cover all aspects of usability in the emerging categories of software products. The current generation of applications is embedded of special features that are not considered by the conventional principles. For this reason, we have developed a new set of usability heuristics that provide accurate results when are used to evaluate transactional Web applications. In this paper, we present a comparative study, in which the effectiveness of our new proposal and the Nielsen's approach are contrasted. For this purpose, two groups of students were trained in a different set of principles. Subsequently, they were requested to conduct a heuristic evaluation using the approach that was assigned. Then, the results were compared. The analysis establishes that our new proposal covers more features, is more understandable and is perceived as easy to use. Although the promising results, some improvements and more experiments are required in other scenarios.

Keywords Human computer-interaction · Heuristic evaluation · Usability heuristics · Comparative study · Experimental case study

F. Paz (✉) · J.A. Pow-Sang
Pontificia Universidad Católica Del Perú, San Miguel, Lima 32, Peru
e-mail: fpaz@pucp.pe

J.A. Pow-Sang
e-mail: japowsang@pucp.edu.pe

F.A. Paz
Universidad Nacional Pedro Ruiz Gallo, Lambayeque, Peru
e-mail: freddypazsifuentes@yahoo.es

1 Introduction

Usability is considered one of the most important factors in the development of software products. This quality attribute establishes the degree in which end users of a specific application believe that using a particular graphical interface would be free of effort. Nowadays, usability is a necessary condition for the success of any product, especially in the context of Software Engineering. Many companies have failed in developing applications by not considering usability guidelines. The lack of a user-centered approach has led to the demise of numerous software products, because of their complexity, difficulty to use, unclearness or hardness to understand.

Given the current importance of usability in software products, several evaluation methods had emerged as a result. These techniques were developed to determine if a graphical user interface meets appropriate levels of usability. One of the best known and widely used techniques for this purpose in HCI is the heuristic evaluation. This usability inspection method involves the participation of three to five specialists who judge whether each dialogue element follows established principles called “heuristics” [1]. According to this method, evaluators must use a list of heuristics to identify usability issues of the user interface design that need to be solved. The ten usability heuristics proposed by Nielsen [2], are considered as the most recognized assessment tool to conduct a heuristic evaluation. However, these principles provide inaccurate results when they are used to evaluate non-traditional software applications [3].

Nowadays, there are new categories of systems such as mobile-based applications, videogames, augmented reality applications and virtual worlds. This new generation of products is embedded of special features that were not considered during the development of the conventional principles [4]. Nielsen’s approach fails to cover all usability aspects that are currently present in the new emerging types of software products [5]. The transactional Web applications are not the exception, and for this reason, we have previously developed fifteen new specialized heuristics that are capable of providing more accurate results in this domain [6].

This paper describes a validation process of the new heuristics in an academic environment. The intention of this work is to provide specialists with a tool which can be used to evaluate effectively the usability of transactional Web applications. For this purpose, we conducted a comparative study, in which the results of both proposals, the new heuristics and the traditional Nielsen’s approach, are compared.

2 Usability and Heuristic Evaluation

Usability is a quality attribute that extends its concept not only to software products but also to electronic devices. The ISO 9241 standard provides a broad definition that can be applied to any technological interface: “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” [7]. Similarly in the context of Software Engineering (SE), the ISO 9126-1 standard provides another definition of usability: “the capability of a software product to be understood, learned and liked by the user when it is used under specified conditions” [8]. Although these concepts are oriented to different types of products such as hardware and software, both emphasize the design of an intuitive interface that allows users to achieve their purposes easily. User satisfaction is usually the result of a successful interaction. This positive user experience (UX) can only be achieved by employing user-centered design techniques and usability evaluation methods during all phases of the development process.

The relevance of usability has led to the emergence of several techniques that allow specialists to evaluate this quality attribute in software products. These usability evaluation methods are defined as “procedures composed of well-defined activities to collect data related to the interaction between the end user and a software product, in order to determine how the particular properties of this application contribute to achieving specific goals [9].” According to Holzinger [10], these methods can be classified into two groups: inspection methods (which involve the participation of usability specialists), and test methods (which involve the participation of end users).

The purpose of the usability evaluation methods is to identify aspects in the interface design that can negatively affect the usability of a software system. These issues can be directly identified by specialists, or by observing the interactions between the software and the end users. A test method can be always supplemented with interviews in which users are asked about their opinion on the interface [11].

Heuristic evaluation is an inspection method that was developed by Nielsen [12]. This technique involves the participation of a small group of evaluators who examine all graphical user interfaces (GUIs) of the software application to determine if all elements of the design follow usability principles called “heuristics” [1]. This method must be conducted by professionals in HCI. In case a heuristic is infringed by the user interface design, the issue is classified as a usability problem. There are many protocols to conduct a heuristic evaluation. However, we have considered the following proposal [11]:

STEP 1: Each evaluator works independently for one or two hours. During this time interval, these specialists should examine all graphical user interfaces of the system to determine if all heuristic principles are followed. The result of this phase is an individual list of usability issues per evaluator, in which each design problem is related to one heuristic that is infringed by the interface.

STEP 2: When all specialists have completed their individual list of usability issues, they should come together to elaborate a single list. In this activity, there should be a consensus between all inspectors to determine if all issues, that were identified, indeed represent a usability problem. Additionally, the team should establish the best way to describe each issue. Finally, they should determine the correct association between the identified problem and the principle is not meet.

Table 1 Rating scale for severity and frequency

Rating	Severity	Frequency (%)
0	I do not agree that this is a usability problem at all	<1
1	Cosmetic problem only: need to be fixed unless extra time is available on project	1–10
2	Minor usability problem: fixing this should be given low priority	11–50
3	Major usability problem: important to fix, so should be given high priority	51–90
4	Usability catastrophe: imperative to fix this before product can be released	>90

STEP 3: Once the single list of usability issues is established, it must be sent to each specialist. In this phase, all evaluators should individually estimate the severity and frequency of each problem that was defined in the single list. The severity is related to the impact of the design problem in the use of the system: in case it occurs, will it be easy or difficult for users to overcome? Likewise, the frequency is related to the number of times each problem becomes visible in the interface. For this study, the scales proposed by Nielsen were considered. The ratings for severity and frequency are presented in Table 1.

STEP 4: As a final step, a member of the team must calculate the criticality (severity + frequency) of each usability problem, and average the individual scores in order to analyze the results. All the evaluation team should elaborate a final report. In this document, specialists must propose possible solutions to the identified issues.

3 A Comparative Study

In heuristic evaluations, the ten usability principles proposed by Nielsen are the most commonly used approach [2]. This list of broad rules is considered a traditional assessment tool by specialists. However, there is enough evidence in the literature stating that these heuristics provide inaccurate results when they are used to evaluate the new categories of software products that are available nowadays [13–15]. Current applications are embedded of emerging features that were not considered during the development of the traditional heuristics. In Web domain, systems incorporate new attributes such as sophisticated designs, extra functionality, and real-time processing. Software products are constantly evolving, and for this reason, an updated assessment tool, that could cover all the new aspects of usability, is required.

In a previous work [16], we conducted a heuristic evaluation to a transactional Web site with the participation of recognized specialists in the field. The purpose

was to identify important aspects related to usability that are not considered by the conventional heuristics in this specific domain. The results demonstrated that Nielsen's approach fails to deal with usability issues related to culture, design, transaction, and functionality. In this way, we developed a new set of principles in order to provide specialists with a tool that will consider all the embedded features of the transactional Web applications for a successful evaluation.

In this paper, we compare the effectiveness and accuracy of both proposals. These two different approaches are presented as follows.

3.1 Nielsen's Usability Heuristics

Nielsen's principles are the most known guidelines to perform heuristic evaluations. According to the author of this approach, these principles are relatively broad and can be applied to any user interface [17]. Given that these heuristics allow specialists to find usability problems during early phases of the software development, these issues can be solved as part of an iterative design process. Nielsen's usability heuristics are [2]: (N1) Visibility of system status, (N2) Match between system and the real world, (N3) User control and freedom, (N4) Consistency and standards, (N5) Error prevention, (N6) Recognition rather than recall, (N7) Flexibility and efficiency of use, (N8) Aesthetic and minimalist design, (N9) Help users recognize, diagnose, and recover from errors, and (N10) Help and documentation.

3.2 New Heuristics for Transactional Web Sites

The new set of heuristics was developed because of the need for a tool that could provide accurate results when it is used to evaluate the usability of transactional Web applications. Although Nielsen's principles are still valid in this domain, there are aspects of usability that are not considered by this traditional approach. We used the methodology proposed by Rusu et al. [18] that defines a systematic procedure to establish new heuristics for interactive systems. After two iterations in which a literature review and some experimental case studies were performed, we obtained a list of fifteen usability principles: (F1) Visibility and clarity of the system elements, (F2) Visibility of the system status, (F3) Match between system and user's cultural aspects, (F4) Feedback of transaction, (F5) Alignment to Web design standards, (F6) Consistency of design, (F7) Standard iconography, (F8) Aesthetic and minimalist design, (F9) Prevention, recognition and error prevention, (F10) Appropriate flexibility and efficiency of use, (F11) Help and documentation, (F12) Reliability and quickness of transactions, (F13) Correct and expected functionality, (F14) Recognition rather than recall, and (F15) User control and freedom. A more detailed description of our proposal can be found in previous works [6, 16].

4 Research Design

4.1 Participants

This study involved the participation of forty-five undergraduate students in their final year of Engineering in Computing at the National University “Pedro Ruiz Gallo” (UNPRG). They were randomly chosen from two different sections of the same course (*Usability Engineering*). As part of the program activities, students had to learn the main concepts of usability as well as the methods to evaluate this quality attribute. This fact allowed us to train students in heuristic evaluations. Although students were from two different sections, they were not mixed. In order to perform a comparative study, the assessment process was explained using a different tool in each section. The traditional Nielsen’s proposal was used in Section I. In the same way, the new set of heuristics for transactional Web applications was employed in Section II. This distribution is presented in Table 2. Students had little or no previous experience in this topic. Given that they attended the same courses of the academic program, we can establish they had a similar background.

Students were informed about this research. All they agreed voluntarily to participate without expecting any compensation. Before conducting the experiment, students were notified that the quality of their reports and answers would not affect their grades in the course. The experiment was performed in January 2016.

4.2 Study Design

Our empirical study was focused on a comparative analysis of the results of both approaches. For this purpose, we have considered the experimental design illustrated in Fig. 1.

First, all participants were trained in the main concepts of usability and heuristic evaluations. In order to avoid personal preferences, we described the method using a different assessment tool in each section. Section I was trained with the ten traditional Nielsen’s heuristics. Similarly, section II was trained with the new fifteen usability heuristics for transactional Web sites. Subsequently, participants examined

Table 2 Instrument-Subject Distribution

Section I	Section II
Nielsen’s usability heuristics	New usability heuristics for transactional web sites
Total subjects: 21 (3 Teams of 7 participants)	Total subjects: 24 (3 Teams of 8 participants)

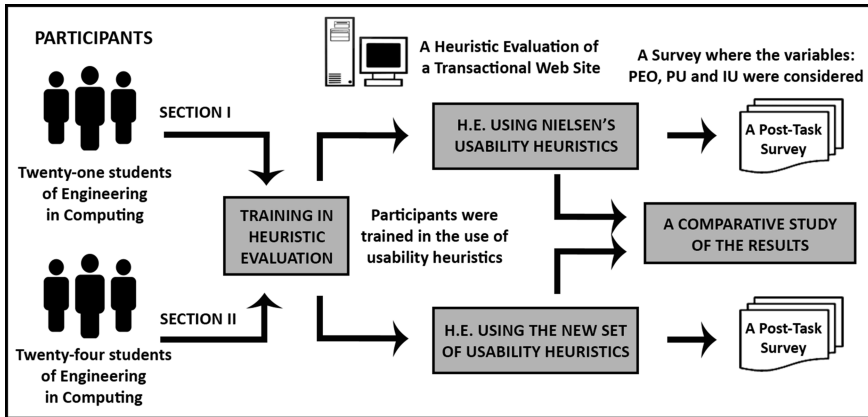


Fig. 1 Experimental design

all the user interface design of an e-Commerce Website using the heuristics that were assigned to their section. For this activity, *HotelClub.com*, an online Web site for hotel booking was selected. Students spent about two hours inspecting the graphical user interfaces of the system. As a result of this process, each student reported a list of usability issues. Each problem was detailed according to following parameters: (a) problem ID, (b) problem definition, (c) comments/explanation, (d) occurrence examples, (d) infringed heuristic, and (e) screenshots.

Next, several teams were randomly organized in both sections. Three teams of seven students were formed in Section I. Likewise, three teams of eight students were formed in Section II. The purpose of each team was to elaborate a single list of issues into a final report. In this document, each group had to specify the average rating for severity, frequency and criticality of each identified issue.

Finally, a post-task survey was employed to identify the students' perceptions about the use of heuristics with regard to the following variables: PEO (perceived ease of use), PU (perceived usefulness) and IU (intention to use). The items of the survey were formulated using a 5-point Likert scale, where 1 was referred to a negative perception of the construct, and 5 to a positive opinion (see Appendix A).

5 Data Analysis and Results

In this section, we present the results of the comparative study. Both approaches were contrasted in the following categories: (1) number of identified issues, (2) errors in associations, and (3) students' perceptions.

5.1 Number of Usability Issues

We consolidated all reports to determine the number of usability issues that were identified by each approach. The results are illustrated in Fig. 2.

The conclusions of this analysis are:

- A total of 25 usability problems were identified by the participants who used the Nielsen's usability heuristics.
- A total of 39 usability problems were identified by the participants who used the new set of usability heuristics for transactional Web sites.
- There are 10 usability problems that can only be identified by the use of the Nielsen's usability heuristics.
- There are 24 usability problems that can only be identified by the use of the new set of usability heuristics for transactional Web sites.
- There are 15 usability problems that can be identified by both approaches.

According to the validation process proposed by Rusu et al. [18], if more usability problems are identified with the new proposal in a comparative study (with Nielsen's heuristics), the results can be considered favorable. However, it is still necessary to analyze the quality of the problems that were identified by our new approach. In Table 3, we present the five most critical problems that only were

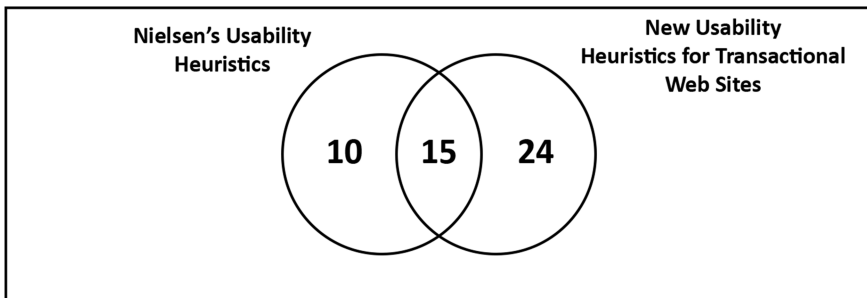


Fig. 2 Number of usability issues identified by each approach

Table 3 Five most critical problems that were only identified by the new approach

Problem ID	Problem definition	Criticality
P18	Limited payment options	5.6
P12	The Website does not display the hotel prices in the local currency of user	5.0
P24	Places of error are no easily recognizable	5.2
P14	Few currency types available	4.4
P25	Recommended destinations are not optimized to user's preferences	4.0

identified by the new usability heuristics. Considering that the maximum score for criticality is 6.0, then the identified issues by the new heuristics are relevant.

5.2 Errors in the Associations

Another aspect that we considered in this comparative analysis was the percentage of wrong associations. When the evaluators identify usability issues, they must specify the heuristic has been infringed. However, in some cases, the heuristic is misunderstood by the inspectors. They establish that a particular principle is not followed, when in fact, the heuristic that was infringed is another. To perform this analysis, we have defined the following concepts:

Valid association: When the inspector associates correctly the identified usability issue with the heuristic that is infringed by the graphical user interface.

Wrong association: When the inspector specifies the infringement of a heuristic that is not actually related to the usability issue that was identified. In Tables 1 and 2, this category represents how many times participants chose this heuristic incorrectly to justify the finding of a usability issue.

The percentage of wrong associations are presented in Table 4 for Nielsen's usability heuristics, and Table 5 for New Heuristics for Transactional Web Sites. The results show that fewer errors are made when our new proposal is used. Possibly, the descriptions of our principles are more understandable than the established ones by Nielsen. However, more case studies are required to generalize these conclusions.

Table 4 Errors in the associations for the Nielsen's usability heuristics

Usability heuristic	Total number of problems	Valid associations	Wrong associations	Percentage of wrong associations (%)
N1	2	1	1	50
N2	2	0	2	100
N3	1	1	0	0
N4	13	6	7	54
N5	2	1	1	50
N6	1	0	1	100
N7	1	0	1	100
N8	2	1	1	50
N9	1	1	0	0
N10	0	0	0	0
Total	25	11	14	56

Table 5 Errors in the associations for the new usability heuristics

Usability heuristic	Total number of problems	Valid associations	Wrong associations	Percentage of wrong associations (%)
F1	4	4	0	0
F2	1	1	0	0
F3	3	3	0	0
F4	0	0	0	0
F5	3	3	0	0
F6	6	5	1	17
F7	1	1	0	0
F8	3	2	1	33
F9	4	4	0	0
F10	6	1	5	83
F11	1	1	0	0
F12	1	1	0	0
F13	5	5	0	0
F14	0	0	0	0
F15	1	1	0	0
Total	39	32	7	18

5.3 Perception Variables

The last aspect we examined in this comparative study was the post-survey. In this questionnaire, we evaluated three variables:

Perceived ease of use (PEU): The extent to which an evaluator believes that using a particular set of usability heuristics would be free of effort.

Perceived usefulness (PU): The extent to which an evaluator believes that a particular set of usability heuristics will achieve its intended objectives.

Intention to use (IU): The extent to which a reviewer intends to use a particular set of usability heuristics in the future. This construct is an intention-based variable for predicting the adoption in practice of the heuristics.

The results are presented in Table 6. Although there is an improvement in all aspects regarding the Nielsen's traditional heuristics, the differences were not highly remarkable. Therefore, further studies are needed.

Table 6 Comparison of the perception variables between the Nielsen's approach and the new usability heuristics

Approach	PEO	PU	IU
Nielsen's usability heuristics	3.02	3.74	3.62
Heuristics for transactional web sites	3.37	3.71	3.64

6 Conclusions and Future Works

Heuristic evaluation is a widely used method to determine the usability of software products. This technique involves a group of specialists judging if all elements of a graphical user interface follow specific guidelines called “heuristics”. Nielsen’s usability heuristics are the most recognized tools to perform heuristic evaluations in all domains. However, these principles fail to cover all aspects of usability that are currently embedded in the current generation of software products. Transactional Web applications are not the exception. Therefore, a new proposal was developed.

A new set of fifteen heuristics was developed in a previous study. In this paper, we present a contribution of experimental nature which describes the validation process in the e-Commerce Web domain of the new heuristics. The purpose of this study was to perform a comparative analysis of the conventional Nielsen’s approach and our new proposal. Forty-five students from two different sections were requested to perform a heuristic evaluation. The traditional heuristics were used in Section I by twenty-one students, and in the same way, the new set of heuristics for transactional Web sites was employed in Section II by twenty-four students. The accurate of the results was compared.

The results show that the new heuristics cover more usability aspects of this new specific domain of software. The number of issues was higher in the evaluations in which the new proposal was used. However, there were usability problems that were only detected by the traditional approach. Although the results are promising, some improvements and more experiments are required in other scenarios. This work is intended to provide specialists with an effective tool to support heuristic evaluations in the context of Transactional Web Applications.

Appendix: Survey Instrument

For each usability heuristic, please mark a cross [X] over the square which most closely matches your opinion. There are no “right” answers to this survey. For this reason, just give your honest opinion based on the experience that was gained in the case study.

- (A) **Ease of Use of the Usability Heuristics:** How easy to use was each usability heuristic? (Table 7)
- (B) **Usefulness of the Usability Heuristics:** The same pattern of question A was employed to measure the usefulness of each usability heuristic. How useful do you consider each usability heuristic is? (Table 8)
- (C) **Intention to Use the Usability Heuristics:** Would you use these heuristics in future evaluations to measure the level of usability of transactional Web applications? (Table 9)

Table 7 Form to measure the ease of use of each heuristic

Heuristic	Very Difficult	Difficult	Neutral	Easy	Very Easy
NX			X		

Table 8 Form to Measure the Usefulness of Each Heuristic

Heuristic	Completely useless	Useless	Neutral	Useful	Completely useful
NX					X

Table 9 Scale to Measure the Intention to Use the Heuristics in the Future

Definitely not	Probably not	Neutral	Probably yes	Definitely yes
			X	

References

1. Nielsen, J.: Usability inspection methods. In: Proceedings of the Conference Companion on Human Factors in Computing Systems (CHI), pp. 413–414. Boston, Massachusetts, USA (1994)
2. Nielsen, J.: 10 Usability Heuristics for User Interface Design. <https://www.nngroup.com/articles/ten-usability-heuristics/> (1995), Accessed 3 Mar 2016
3. Rusu, C., Rusu, V., Roncagliolo, S., Apablaza, J., Rusu, V.Z.: Design, user experience, and usability: design discourse: 4th International Conference, DUXU 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2–7, 2015, Proceedings, Part I, chap. User Experience Evaluations: Challenges for Newcomers, pp. 237–246. Springer International Publishing, Cham (2015)
4. Rusu, C.: Usability versus user experience: theory and practice. In: Proceedings of the XV International Conference on Human Computer Interaction (Interacción), pp. 1–2. Puerto de la Cruz, Tenerife, Spain (2014)
5. Paz, F., Villanueva, D., Rusu, C., Roncagliolo, S., Pow-Sang, J.A.: Experimental evaluation of usability heuristics. In: Proceedings of the 2013 Tenth International Conference on Information Technology: New Generations (ITNG), pp. 119–126. Las Vegas, NV, USA (2013)
6. Paz, F., Paz, F.A., Pow-Sang, J.A.: Experimental case study of new usability heuristics. In: Marcus, A. (ed.) Design, User Experience, and Usability: Design Discourse, Lecture Notes in Computer Science, vol. 9186, pp. 212–223. Springer International Publishing, Berlin (2015)
7. ISO: Ergonomic requirements for office work with visual display terminals (VDTs). ISO IEC 9241-11, International Organization for Standardization, Geneva, Switzerland (1998)
8. ISO: Software Engineering-Product Quality-Part 1: Quality Model. ISO IEC 9126-1. International Organization for Standardization, Geneva, Switzerland (2001)
9. Fernandez, A., Insfran, E., Abrahão, S.: Usability evaluation methods for the web: a systematic mapping study. *Inf. Softw. Technol.* **53**(8), 789–817 (2011)
10. Holzinger, A.: Usability engineering methods for software developers. *Commun. ACM—Interact. Design Child.* **48**(1), 71–74 (2005)
11. Paz, F., Paz, F.A., Villanueva, D., Pow-Sang, J.A.: Heuristic evaluation as a complement to usability testing: a case study in web domain. In: Proceedings of the 2015 12th International Conference on Information Technology—New Generations (ITNG), pp. 546–551. Las Vegas, NV, USA (2015)

12. Nielsen, J., Molich, R.: Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI), pp. 249–256. Seattle, Washington, USA (1990)
13. Jimenez, C., Rusu, C., Roncagliolo, S., Inostroza, R., Rusu, V.: Evaluating a Methodology to Establish Usability Heuristics. In: Proceedings of the 2012 31st International Conference of the Chilean Computer Science Society (SCCC), pp. 51–59. Valparaiso, Chile (2012)
14. Inostroza, R., Rusu, C., Roncagliolo, S., Jimenez, C., Rusu, V.: usability heuristics for touchscreen-based mobile devices. In: Proceedings of the 2012 Ninth International Conference on Information Technology: New Generations (ITNG), pp. 662–667. Las Vegas, NV, USA (2012)
15. Díaz, J., Rusu, C., Pow-Sang, J.A., Roncagliolo, S.: A cultural-oriented usability heuristics proposal. In: Proceedings of the 2013 Chilean Conference on Human—Computer Interaction, pp. 82–87. Temuco, Chile (2013)
16. Paz, F., Paz, F.A., Pow-Sang, J.A., Collantes, L.: Usability heuristics for transactional web sites. In: Proceedings of the 2014 11th International Conference on Information Technology: New Generations (ITNG), pp. 627–628. Las Vegas, NV, USA (2014)
17. Nielsen, J.: Usability Engineering, 1st edn. Academic Press, San Diego (1993)
18. Rusu, C., Roncagliolo, S., Rusu, V., Collazos, C.: A methodology to establish usability heuristics. In Proceedings of the 4th International Conferences on Advances in Computer-Human Interactions (ACHI), pp 59–62. Gosier, Guadeloupe, France (2011)