AHFE International

A Comparison between Usability Evaluation Methods: A Case Study in the E-Commerce Domain

Freddy Paz[1] and Freddy Asrael Paz[2]

[1]Pontificia Universidad Católica del Perú San Miguel, 15088, Peru
[2]Universidad Nacional Pedro Ruiz Gallo, Lambayeque, Peru

ABSTRACT

Usability is a quality attribute that nowadays must not be ignored by software development teams. Designing a software product that is difficult to use or understand, where the user is lost, and the achievement of the user's objectives is prevented, can cause users to choose to use similar alternatives. For this reason, Human-Computer Interaction specialists have developed a set of usability evaluation methods that allow problems to be identified to guarantee an appropriate level of user-friendliness of the application. However, there is a large disagreement between the authors as to which is the best method to use. In this study, the results of two of the most used methods to carry out usability evaluations of software products have been compared: the questionnaire and the heuristic evaluation. On the one hand, the questionnaire has turned out to be the method most used by the scientific community because it is easy to use and does not require much material preparation. However, most questionnaires have been aimed at being answered by end-users. In a previous work, we developed a questionnaire for UX practitioners that in this case has been used for the evaluation of an electronic commerce application. On the other hand, heuristic evaluation is a method where specialists verify compliance with certain principles in the design of the graphical interfaces of a software product. It is also widely used like questionnaires but requires more planning and execution time. In this research, it is possible to evidence the differences in the type of problems that can be identified by each of the methods. In conclusion, it is possible to mention that although a percentage of problems are identified that can only be identified with one or the other method, the recommendation is to use both methods as the results complement each other.

Keywords: Human-computer interaction, Assessment questionnaire, Heuristic evaluation, Comparative analysis, Usability case study

INTRODUCTION

Ensuring that a software product is understandable, usable, and attractive has become a relevant and essential aspect. Professionals involved in software development are increasingly concerned about the importance of designing products that allow users to achieve their goals with satisfaction (Craven and Booth, 2006). If the interaction experience with a product, system, or service does not meet the expectations of users, they stop using it, causing the failure of the new technological proposal that has been placed on the market (Quiñones and Rusu, 2019).

To offer quality software products, specialists have developed a set of methods and procedures that determine if the proposed graphical interfaces are appropriate. These proposals are known as usability evaluation methods and consist of a series of steps that allow determining the design characteristics of the graphical user interfaces and how these aspects contribute to achieving the user's goals (Fernandez et al., 2011). In the literature, it is possible to find different studies that address the multiple usability evaluation methods that have been proposed to date (Paz and Pow-Sang, 2016). From these methods, the most reported are usually the traditional tests with users, heuristic evaluations, and perception questionnaires (Salinas et al., 2020). Although it is not possible to establish that one method is better than another, it is possible to identify advantages in some of them. Heuristic evaluations are easy to perform and can be executed in a short time relative to other methods. It is for this reason that researchers frequently use this method to perform usability inspections of their software products. Likewise, the method can be applied to functional prototypes as well as sketches and mockups in the early phases of the development process (Díaz et al., 2017). On the other hand, questionnaires are also widely used due to their ease of application and elaboration (Sauro and Lewis, 2016). Unlike user tests where HCI specialists must develop several materials and prepare scenarios, in the application of questionnaires, users are only requested to answer a set of previously defined questions. Although users are also required to previously interact with the software product on which they are going to give their opinion through the selected questionnaire, this process can be carried out remotely without the need for the presence of users in a usability lab. The process can be performed in an asynchronous mode, making the request via email and providing an online form in which users must answer the questions that will allow specialists to determine the degree of perception regarding the level of usability of the product, system or service evaluated.

Both methods described the heuristic evaluations as well as the usability perception questionnaires are reliable instruments that can be used to improve the quality of a product's interaction interfaces. However, this research aims to show the similarity and differences in the results of applying one or another method. Although both methods are widely recognized and used by the scientific community, our intention is to highlight the degree of granularity that can be achieved with each of the methods, the advantages and distinctions. This work was elaborated to provide specialists with a better vision of the possibilities presented by each method, and in this way, they can make a more accurate decision for the development of their project.

THEORETICAL BACKGROUND

This section describes some concepts that have been relevant to carry out the research. Given that this research is focused on the heuristic evaluation and on the usability measurement questionnaire, the definition is detailed as follows, along with previous works that has allowed the development of this study.

Heuristic Evaluation

The heuristic evaluation process is an inspection method directed by specialists in the area of Human-Computer Interaction (HCI) who, due to their degree of expertise, are capable of identifying design problems in a set of graphical user interfaces (Nielsen, 1994a). The method consists of a group of three to five specialists reviewing each of the evaluated interfaces in search of non-compliance with specific design principles (Nielsen, 1994b). Usually, the ten design principles proposed by Nielsen are used in a heuristic evaluation (Nielsen, 1994c). However, there are scenarios where these heuristics do not completely cover all the characteristics that must be evaluated for certain categories of software products (Paz et al., 2014). There are categories of emerging software that have their own nature, elements, and characteristics that are not currently considered by the Nielsen's heuristics. For example, in an e-commerce application there are aspects related to the 1) transaction, 2) security, 3) ease of access to product information, 4) cultural considerations, 5) availability of optimal search and filters tools, 6) access to means and channels of contact with the company, 7) availability of a post-sale service, among others, which influence the user experience, and affect the degree of perception of the usability of a software system. This fact has led many specialists to propose their own set of heuristics and guidelines that allow them to cover all the factors and aspects that impact the perception of usability for a certain type of application (Jiménez et al., 2012).

Given the gap in the research that exists with the traditional Nielsen heuristics, in a previous work, we developed and experimented with a new set of usability principles for transactional websites (Paz et al., 2015). This novel and validated proposal, in addition to address the typical aspects established by Nielsen's heuristics, focuses on characteristics that are relevant for a person who uses an e-commerce website. The new approach is composed of fifteen heuristics, and they were developed through an analysis of the user's objectives in a purchase process, and through a study in which their expectations and perspectives were analyzed regarding what an e-commerce website must provide (Paz, 2014). This investigation led to the addition of evaluation elements to the traditional Nielsen proposal, such as cultural aspects, navigability, the visibility of purchase elements, security, the reliability of financial transactions, the possibility of access to after-sales service as well as support and help.

Another identified aspect about the heuristic evaluation method is that the process was not correctly defined in the literature (Paz et al., 2017). Although Jakob Nielsen establishes a set of guidelines and considerations to consider for the execution of the method, clearly defined steps are not established. This scenario motivated to propose in a previous work (Paz, 2019), a framework with specific activities, roles and process models, which formalizes the heuristic evaluation method and involves five phases: (1) Planning, (2) Training, (3) Evaluation, (4) Discussion and (5) Report. Although this new execution protocol defines clear steps to be considered, it is not an unbending framework, it can be adapted to the needs of each evaluation team. In general terms, this proposal establishes in the first place, a planning phase in which

the manager defines the goal of the evaluation, the scope, the team of specialists, the heuristics and the materials to be employed. Subsequently, there is an optional training phase for the specialists in this same process. In this way, the evaluators are clear about the steps they must follow throughout the evaluation process. Then, the evaluators individually carry out the identification of usability issues in the design of graphical user interfaces using the selected heuristics. This is performed through a verification of compliance with the heuristic principles. If the graphical interface fails any of the heuristics, then this failure is classified as a usability problem. Once the individual evaluation has been carried out, discussion phase is proposed and the members of the evaluation as a team must consolidate the obtained results. Subsequently, it is proposed a problem qualification process to determine those problems that are the most critical and have the greatest impact on the level of usability of the software product. Finally, a report is elaborated to communicate the results as well as establish solution proposals and highlight the positive aspects of the interfaces. In this case study, this new evaluation approach has been used in conjunction with the new transactional heuristics.

Usability Questionnaire

Usability perception questionnaires represent a method widely used by specialists to determine the ease-of-use degree of an interface proposal (Quezada et al., 2021). The practicality with which the method can be applied makes it one of the most preferred methods by specialists. Although users must interact with the software product as a previous step, the time the method takes to apply is reasonably short, and it can also be used remotely. The purpose of a questionnaire is to capture the users' perception of the interface design. Even the team of specialists should not worry about preparing the questionnaire, since there are multiple proposals available in the literature that can be used (Sauro and Lewis, 2016). A positive aspect of the questionnaires is that they are validated and have been tested in multiple scenarios to guarantee the reliability of the data obtained. Usually, they have a defined Likert scale and an associated formula that allows obtaining the total number value on the level of usability of the software product. Quantifying usability opens up the possibility of making comparisons and determining by exact numbers whether one design proposal is better than another or exactly how much we require to achieve a desired degree of usability.

There is an aspect that the professionals have not yet studied concerning the usability questionnaires. Most of the proposals are aimed at end-users of a software product, and there are very few approaches that can be used by specialists in Human-Computer Interaction to determine by themselves the level of usability of a software product. Given this scenario, in a previous work (Paz and Granollers, 2019), we developed a new questionnaire that involves 60 items and addresses the main sub-attributes of the concept of usability. The intention of this novel assessment instrument was to be used by specialists allowing them to answer about their perception of usability considering their vast experience in design and user experience. The new proposal was elaborated through an exhaustive analysis of the ten Nielsen heuristics (Nielsen,

1994c), and the sixteen design rules established by Tognazzini (2014). In this comparative study, this new questionnaire was used for the usability evaluation.

EXPERIMENTAL DESIGN

The intention of this study was to compare the difference between applying a heuristic evaluation and a perception questionnaire to measure the usability of an e-commerce website. To conduct this case study, three specialists in the field of Human-Computer Interaction were contacted, who agreed to participate in the study voluntarily. About the specialists, it is possible to establish that all have the same level of expertise and a similar profile since they all have a master's degree in Computer Science or similar careers and have participated as evaluators in multiple usability consultancies. Before beginning the usability assessment, the specialists received and signed an informed consent agreement. The evaluation process was carried out in an asynchronous way. The specialists were given a reasonable time of one week to perform and execute the evaluation and subsequently send us their findings. A virtual session via Google Meet was organized and held in which we explained to the specialists the evaluation route to be completed.

First, the evaluators had to spend time going through the website through its different sections. A defined time was not established for this activity since the purpose was for the specialists to feel completely familiar with the application in a way in which they could later indicate their findings. Then, they performed the usability evaluation following the formal evaluation protocol (Paz, 2019) and by employing the heuristics for transactional websites (Paz, 2014). The training phase was not required since all the specialists had experience in performing this type of inspection and in the use of guidelines. Although the evaluators independently and asynchronously identified the design issues that affect usability, there was a virtual meeting where they were able to consolidate the results and establish the severity of each identified problem. After this meeting, independently they proceeded to complete the evaluation questionnaire (Paz and Granollers, 2019). Finally, the results were processed for analysis.

The case study was carried out in the e-commerce domain for a website in Peru. The company's core business is footwear, its web address will remain anonymous because the purpose of this research is totally academic. The usability evaluation was executed from January 03 to January 10, 2022. The templates for the inspections were designed with the support of the *HCI, Design, User Experience, Accessibility, and Innovation Technology Research Group* (HCI-DUXAIT), which is assigned to the *Department of Engineering* and is part of the *Pontifical Catholic University of Peru* (PUCP).

ANALYSIS OF THE RESULTS

The results of the application of both methods demonstrate that there are similarities in the findings and that any of the methods can be used with

Table 1. Top 6 of the most severe usability problems identified during the HE.

ID	Problem Description	Severity
P01	Error during the purchase process. Sometimes it is not possible to finalize the purchase due to errors in the operation.	4.00
P02	The website does not present online communication channels to support the user during the purchase process.	3.67
P03	It is not possible to identify a purchase history section.	3.67
P04	Inconsistency in language. Despite being a website in Spanish, many phrases are displayed in English.	3.33
P05	The system does not correctly save the entered information and forces the user to re-enter data.	3.33
P06	The system does not make a correct use of the space distribution to display the graphical elements.	3.33
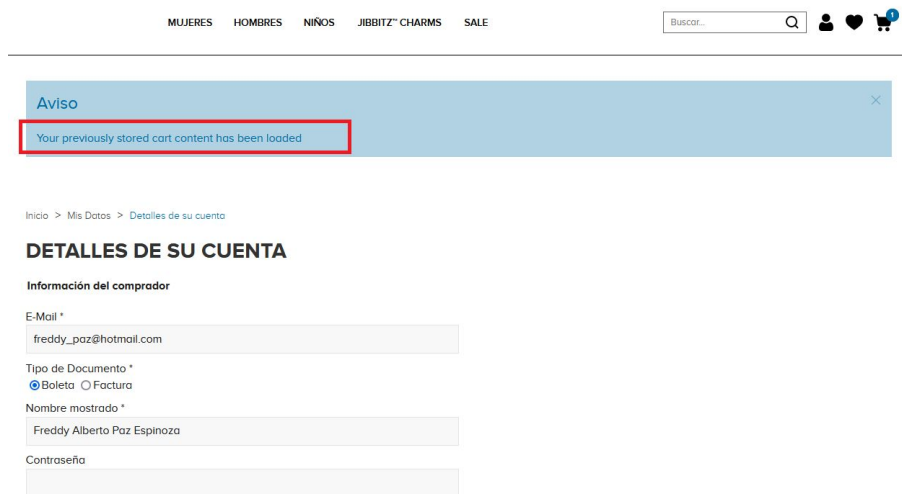
Figure 1: Inconsistency in the language in the GUI of the evaluated website.

certainty for a usability evaluation. The heuristic inspection allowed the detection of 26 usability issues. Table 1 highlights the six most critical identified problems as well as their average severity score. To calculate the severity of a problem, a scale from 0 to 4 was used, where 0 means that the problem is not severe at all, and 4 means that is imperative to fix before software product can be released. Most of the problems are related to an incorrect and unexpected operation of the application, cultural issues, interface design, and lack of essential elements that should be considered in an e-commerce system.

In Figure 1, one of the most representative problems of the evaluated system can be observed, related to the inconsistency in the language. In this case, for instance, many of the international companies with branches in Latin America usually use generic software solutions that are already preconfigured to be used in other environments, different than the local ones. It is due to this and an inappropriate configuration that design errors such as "inconsistency in language" and issues associated with cultural aspects such

as currency, icons, availability of payment methods and expressions that still need to be adapted to local scenarios. Companies must concern about offering websites that are indeed aligned with the needs and expectations of the users of the context to which they are addressing.

Regarding the usability evaluation questionnaire, it is divided into 15 sections. Each section is made up of approximately 3 to 5 verification questions. For each question, the user must answer affirmatively, negatively, neutrally or indicate that if the aspect to be evaluated does not apply for the evaluation. Since this instrument has been designed to allow the quantification of the level of usability, the affirmative answers award 1 point, while the neutral ones 0.5 points, and the negative ones 0 points. Those questions that do not apply are removed from the final qualification of the software product. In this way it is possible to obtain a total score of 60 points. Table 2 shows the total value obtained by each specialist in the categories of the questionnaire. With basis on the results, it can be established that there is a relationship in the findings. The evaluators agree on the scores, which is evidence of the correlation and reliability of the evaluation instrument, and it can also be observed there are categories related to the results obtained in the heuristic evaluation. For example, there is a relationship between the problems related to the non-existence of consumer support channels and the low score in the help and documentation category. Likewise, the functional errors of the application are reflected in the low scores obtained in the categories of user control, flexibility, efficiency and help users to recover from errors. Finally, cultural problems are reflected in the score for the category of connection between the system and the real world. However, one important aspect that is observed in the results is related to degree of granularity and the types of problems identified. Given that in the heuristic evaluation is not defined a set of specific aspects, problems with a broader spectrum are detected, going deeper than generally and commonly is considered.

CONCLUSION AND FUTURE WORKS

Usability is a highly relevant quality attribute today for the development of software products. It is for this reason that various usability evaluation methods have emerged. Two of the most widely used methods are the heuristic evaluations and the perception questionnaires. This study describes an example of the application of both methods in the inspection of an electronic commerce website. To carry out the experimental study, the support of usability specialists was requested. Likewise, the evaluation instruments were defined, including a defined formal process, a set of heuristics and a new questionnaire proposal. These tools were developed and validated in previous works and demonstrated their effectiveness in different and multiple contexts.

The results show that both methods offer similarities in the findings. It is possible to appreciate that the usability problems identified through the heuristic inspection are reflected in the low scores obtained from those categories that address related aspects. However, through heuristics it is possible to find a broader and more diverse spectrum of problems. The questionnaire,

Table 2. Scores obtained by the specialists in the categories of the questionnaire.

Category	Specialists		
	S1	S2	S3
C1. Visibility and system state	2/5	1.5/5	1/5
C2. Connection between the system and the real world, metaphor usage and human objects	2/4	2/4	3/4
C3. User control and freedom	0/3	1/3	0.5/3
C4. Consistency and standards	4/6	4/6	4.5/6
C5. Recognition rather than memory, learning and anticipation	3/5	3.5/5	2/5
C6. Flexibility and efficiency of use	0/6	2/6	1.5/6
C7. Help users recognize, diagnose and recover from errors	1/4	1.5/4	0.5/4
C8. Preventing errors	1.5/3	1/3	2.5/3
C9. Aesthetic and minimalist design	3/4	2/4	2/4
C10. Help and documentation	0/5	0.5/5	0/5
C11. Save the state and protect the work	2/3	2.5/3	2/3
C12. Color and readability	4/4	4/4	3/4
C13. Autonomy	1/3	0.5/3	0/3
C14. Defaults	0/3	0/3	0/3
C15. Latency reduction	1/2	0/2	0.5/2
Total Score	24.5/60	26/60	23/60

on the other hand, allows us to confirm aspects that are commonly reported in usability evaluation reports. Nevertheless, both methods are reliable and can be used with certainty, in a simple way, and as a complement in a software development process. As future work, it is expected to conduct case studies in other scenarios that allow generalizing the obtained results. Future test scenarios could involve other categories of software products and distinct profiles of specialists.

ACKNOWLEDGMENT

REFERENCES

Craven, J. and Booth, H. (2006). Putting awareness into practice: practical steps for conducting usability tests, Library Review, 55(3), 179–194.

Díaz, J., Rusu, C. and Collazos, C. (2017). Experimental validation of a set of cultural-oriented usability heuristics: e-Commerce websites evaluation, Computer Standards & Interfaces, 50, 160–178.

Fernandez, A., Insfran, E. and Abrahão, S. (2011). Usability evaluation methods for the web: A systematic mapping study, Information and Software Technology, 53(8), 789–817.

Jiménez, C., Rusu, C., Roncagliolo, S., Inostroza, R. and Rusu, V. (2012). Evaluating a Methodology to Establish Usability Heuristics. 31st International Conference of the Chilean Computer Science Society (SCCC 2012), 51–59.

Nielsen, J. (1994a). Usability inspection methods. Conference Companion on Human Factors in Computing Systems, 413–414.

Nielsen, J. (1994b). Usability Engineering. Massachusetts: Morgan Kaufmann.

Nielsen, J. (1994c). 10 Usability Heuristics for User Interface Design. https://www.nngroup.com/articles/ten-usability-heuristics/

Paz, F. (2014). Heurísticas de usabilidad para sitios web transaccionales. Pontificia Universidad Católica del Perú. https://tesis.pucp.edu.pe/repositorio/handle/20.500.12404/5399

Paz, F. (2019). Método para la evaluación de usabilidad de sitios web transaccionales basado en el proceso de inspección heurística. https://tesis.pucp.edu.pe/repositorio/handle/20.500.12404/9903

Paz, F. and Granollers, T. (2019). Redesign of a Questionnaire to Assess the Usability of Websites. In: Ahram, T., Karwowski, W., Pickl, S., Taiar, R. (eds) Human Systems Engineering and Design II. IHSED 2019. Advances in Intelligent Systems and Computing, vol 1026. Springer, Cham.

Paz, F. and Pow-Sang, J.A. (2016). A Systematic Mapping Review of Usability Evaluation Methods for Software Development Process, International Journal of Software Engineering and Its Applications, 10(1), 165–178.

Paz, F., Collazos, C. and Pow-Sang, J.A. (2017). Validation in the Web Domain of a Formal Process to Evaluate the Usability of Software Applications: An Approach based on the Heuristic Inspection, International Journal of U- and E-Service, Science and Technology, 10(7), 65–82.

Paz, F., Paz, F.A. and Pow-Sang, J.A. (2015). Experimental Case Study of New Usability Heuristics. In: Marcus, A. (eds) Design, User Experience, and Usability: Design Discourse. Lecture Notes in Computer Science, vol 9186. Springer, Cham.

Paz, F., Paz, F.A., Pow-Sang, J.A. and Collantes, L. (2014). Usability Heuristics for Transactional Web Sites. 11th International Conference on Information Technology: New Generations (ITNG 2014), 627–628.

Quezada, P., Cueva, R. and Paz, F. (2021). A Systematic Review of User-Centered Design Techniques Applied to the Design of Mobile Application User Interfaces. In: Soares, M.M., Rosenzweig, E., Marcus, A. (eds) Design, User Experience, and Usability: UX Research and Design. HCII 2021. Lecture Notes in Computer Science, vol 12779. Springer, Cham.

Quiñones, D. and Rusu, C. (2019). Applying a methodology to develop user eXperience heuristics, Computer Standards & Interfaces, 66, 103345.

Salinas, E., Cueva, R. and Paz, F. (2020). A Systematic Review of User-Centered Design Techniques. In: Marcus, A., Rosenzweig, E. (eds) Design, User Experience, and Usability. Interaction Design. HCII 2020. Lecture Notes in Computer Science, vol 12200. Springer, Cham.

Sauro, J. and Lewis, J.R. (2016). Quantifying the User Experience: Practical Statistics for User Research, 2nd edition, Massachusetts: Morgan Kaufmann.

Tognazzini, B. (2014). First Principles of Interaction Design (Revised & Expanded). https://asktog.com/atc/principles-of-interaction-design/