

# Usability Testing as a Complement of Heuristic Evaluation: A Case Study

Braulio Murillo<sup>(✉)</sup>, Silvia Vargas, Arturo Moquillaza,  
Luis Fernández, and Freddy Paz

Pontificia Universidad Católica del Perú, Lima, Peru  
{bmurillov, silvia.vargas}@pucp.edu.pe,  
{amoquillaza, fernandezm.christian, fpaz}@pucp.pe

**Abstract.** The usability assessment of software systems is becoming more relevant, especially when it comes to web-based systems. A well-built and user-friendly website is capable of capturing a potential customer. There are different methods to evaluate usability, including heuristic evaluation and usability testing with users. In the present work, the critical functionalities of an airline's website were evaluated, such as the purchase of tickets, flight reservations, among others. First, a heuristic evaluation was performed and then a usability test with users. The evaluations were developed in an academic context and the participants were postgraduate students of a university. The problems detected in the first evaluation served to define specific tasks in the usability test with users. In this way, the results of the evaluations were complemented.

**Keywords:** Human computer interaction · Heuristic evaluation · Usability evaluation · Usability testing · Transactional web applications

## 1 Introduction

In software engineering, the term “Usability” is related to the ease of use of a software product [3]. Usability is a quality attribute that measures how easy the user interface is to use. It also includes methods to improve ease of use during the software design process [1]. Nowadays on the web, usability is a necessary condition for survival. If a website is difficult to use, people will stop using it. If the page does not clearly state what a company offers and what users can do on the site, people will stop using it. If users get lost on a website, they will stop using it. If the information on a website is difficult to read or does not answer the key questions of users, they will stop using it.

The first e-commerce law is that if users cannot find the product, they cannot buy it either [1]. In this paper we will evaluate the website of an airline using heuristic evaluation and a usability test with users as a complement [5]. The results obtained in both tests will be shown to compare them and provide some conclusions.

## 2 Related Work

Given the importance of usability, there are two types of methods to perform a usability assessment: inspection methods and test methods [4]. The difference between them lies in the person who applies them. In the first case the inspectors perform and in the second the users participate [2].

Heuristic evaluation is a well-known inspection technique that is widely used by usability specialists. It was developed by Nielsen as an alternative to user testing [3].

The usability test with users is a method of software products evaluation in which active users are representative of said products. The main purpose of this method is the identification of usability problems during the interaction of users and the system. This method allows collecting qualitative and quantitative data, as well as user satisfaction with the software product [7].

In the present work, a usability assessment is performed with the “Heuristic Evaluation” inspection method, which is based on the analysis of the ten heuristics defined by Dr. Jakob Nielsen, Latam Airlines web application [www.latam.com](http://www.latam.com).

We also perform a usability assessment with the user testing method to the same application. This evaluation is done as a complement to the heuristic evaluation previously performed by the same group. In [6] the advantages of performing heuristic and user evaluations as complementary studies are developed.

The present work has been developed under an academic context, during the months of June and July of the year 2016. All the participants have developed the tests with professionalism and ethical values.

## 3 Research Design

In order to test usability in the website [latam.com](http://latam.com), were used two methods that complement each other: heuristic evaluations and usability test users. The objective of this test and selection of the website were academics.

### 3.1 Description of the Web Site

The website corresponds to the new brand of LATAM Airlines Group: LATAM, which is a product of the consolidation of the two brands of the group that preceded it: LAN and TAM. LATAM is an airline based in the Republic of Chile, which operates in South, Central and North America, as well as the Caribbean, Europe and Oceania; with different destinations in more than 26 countries [8].

Due to the nature of the company, the website has numerous sections, among them: Purchase of tickets, Purchase of packages, Check-in Service, Charge Service, Flight Information, Offers Section, Press Room, Section for Investors, among many others. However, the main emphasis is on the first three sections because it is frequently used by passengers (Fig. 1).

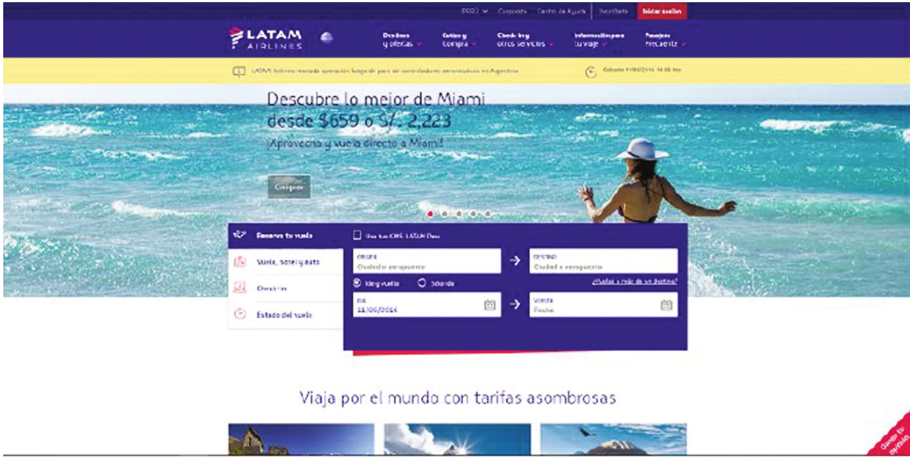


Fig. 1. Main page of the evaluated web site: [latam.com](http://latam.com)

The purpose of the application is to offer its customers a computer tool for its main services: purchase of tickets, flight check-in and flight status display. In this sense, the website has a design oriented to emphasize these services and to link them to each other.

### 3.2 Study Design

The purpose of this paper was to compare heuristic evaluations and user usability tests based on a web transactional system. This work was developed in two moments. First a heuristic evaluation was carried out and then a user usability test was developed.

## 4 Heuristic Evaluation

### 4.1 Participants

The heuristic evaluation was performed using the Nielsen’s methodology analyzing the ten usability principles “heuristics”. The evaluation was performed by four evaluators, computer engineers, three of them master’s students and one doctoral student.

### 4.2 Phases

This section describes the steps used to perform the heuristic evaluation. These are described below.

First phase: Each participant performed alone an evaluation of the product and recorded all the results found in their respective reports.

Second phase: A moderator, who was previously selected, facilitated a meeting where the evaluators were able to unify the results obtained by briefly explaining the problems they found. As a result, a clean and unified listing of the problems encountered was obtained.

Third phase: Each evaluator independently rated the severity and frequency of each one of the problems of the unified listing. With the values of severity and frequency was calculated the criticality:  $\text{criticality} = \text{severity} + \text{frequency}$ .

Fourth phase: A coordinator, who was previously selected, performed the activities of this phase, calculating the averages and standard deviations of the three previously calculated values: severity, frequency and criticality of each problem. With the results, was established a ranking of the problems found.

The severity was evaluated according to the rating proposed by Nielsen [9], in which 0 means “I don’t agree that this is a usability problem at all” and 4 means “Usability catastrophe: imperative to fix this before product can be released”.

The frequency was evaluated according the rating of the Table 1.

**Table 1.** Frequency ratings

Note	Frequency
0	<1%
1	1–10%
2	11–50%
3	51–90%
4	>90%

### 4.3 Data Analysis and Results

A total of thirty seven usability problems were identified, which were categorized by the participants who performed the heuristic evaluation. Only for the heuristic “Recognition rather than recall” wasn’t found non-compliance. In Table 2, it can be seen the times that each unfulfilled heuristic.

Table 3 shows the identified problems sorted descending by severity, without considering the frequency. Of the total of problems found by the evaluators, almost 50% of them resulted in a severity value greater than or equal to 2.50, that is, they tend to be greater or catastrophic.

Thus, the most severe problem identified by the evaluators is that the system displays a blank screen after log off a user, which could be perceived as unsafe for the end user. Also, the evaluators have considered severe that during the ticket purchase’s process, the system doesn’t provide the option to return to the previous step, moreover, it doesn’t allow to save the information already entered so that forcing a backward the information is lost.

Other severe problems are that the system displays blank screens before certain options, other options don’t even work, displays error messages that aren’t understand and have broken links. Almost all screens don’t have the option of help for the user.

**Table 2.** Unfulfilled heuristics

ID	Heuristic	Problems that non-compliance the heuristics	Number the problems that non-compliance the heuristics
N1	Visibility of system status	P10, P12, P14, P17, P18, P27, P30, P31	8
N2	Match between system and the real world	P4, P22, P26, P35	4
N3	User control and freedom	P3, P16, P28, P37	4
N4	Consistency and standards	P2, P5, P7, P8, P11, P15, P19, P24	8
N5	Error prevention	P1, P6, P20, P23, P29, P32, P33	7
N7	Flexibility and efficiency of use	P25	1
N8	Aesthetic and minimalist design	P9	1
N9	Help users recognize, diagnose, and recover from errors	P21, P34	2
N10	Help and documentation	P13, P36	2

**Table 3.** Ranking of the more severe problems

ID	Problem	Average severity
P17	The screen goes blank on log off	3.25
P26	Button return to the previous screen on the Visa payment screen doesn't work	3.25
P3	In the ticket purchase's process, in several steps it isn't possible to return to the previous ones, it's returned to option 1: Date	3.00
P18	The screen goes blank by selecting the prize icon	3.00
P22	The name of the page doesn't match when you enter the Claims book option	3.00
P32	Broken links	3.00
P34	The share option on Facebook doesn't work, shows a technical error	3.00
P37	"Where you find it" option doesn't work	3.00

On the other hand, the Table 4 shows the identified problems sorted descending by criticality. It's observed that the maximum value of criticality is 6, this means that the problems encountered don't drastically affect the functionality of the system.

The problem of greater criticality is that the system doesn't provide the user with the option to return on any of the pages of the ticket purchase's process. This means that this problem is the most severe and most frequent.

**Table 4.** Ranking of the more critically problems

ID	Problem	Average severity	Average frequency	Average critically
P3	In the ticket purchase's process in several steps it isn't possible to return to the previous ones, it's returned to option 1: Date	3.00	3.00	6.00
P32	Broken links	3.00	2.75	5.75
P13	There isn't help option that can guide the user in case he doesn't understand the interface	2.75	2.75	5.50
P19	The system behaves differently when using different browsers	2.75	2.75	5.50
P16	The back button isn't displayed properly	2.25	3.00	5.25
P23	The system returns to home when choosing the country of origin from different pages: the page of claims book and the page of help	2.75	2.50	5.25
P26	Button return to the previous screen on the Visa payment screen doesn't work	3.25	2.00	5.25
P35	There is poorly worded information	2.50	2.75	5.25
P10	There is the option to see rates in other currencies, but it doesn't work	2.75	2.25	5.00
P22	The name of the page doesn't match when you enter the claim's book option	3.00	2.00	5.00

Other critical issues include broken links, lack of system-wide help and poorly worded information on several pages. One problem worth highlighting is that the system has different behaviors when using different browsers.

Finally, the problems P17 and P26, although they were evaluated as very severe, were not the most critical because they didn't have a high frequency of occurrence.

## 5 Usability Testing

### 5.1 Test Purpose

The purpose of performing the usability test in the LATAM application is that the user may encounter problems when using it by performing certain previously defined tasks. The tasks have been established based on the result of the heuristic evaluation performed in the previous stage. Problems with severity greater than 3 (P17, P26, P3, P18) that are directly related to the acquisition of passages were selected, since this is the most important functionality of the application and for them, activities that could cover these problems were defined.

**5.2 Test Design**

From the heuristic inspection that was performed in the previous stage, the problems were taken with severity greater than 3. With them, the problem, task and context matrix is elaborated, as shown in Table 5. Of the list of problems, those that emphasize the transactional functionalities, that is, in the purchase of air passages, on other functionalities of the web site were chosen.

**Table 5.** Matrix of problems, task and context

ID	Problem	Task	Context
P26	Button return to the previous screen on the Visa payment screen doesn't work	1	Step 5 of ticket purchasing, Payment section. When user want to go back to last screen by clicking in Return link, system redirects to a error page with no possibility of leave
P3	In the ticket purchase's process in several steps it isn't possible to return to the previous ones, it's returned to option 1: Date	1	When user returns to a previous step, no matter being in step 2 or step 3, always go back to step 1 (date)
P10	There is the option to see rates in other currencies, but it doesn't work	1	In the Flight reservation option, step 3, Price, Option "See rates in other currencies" exists, but it does not make any change in prices or rates
P13	There isn't help option that can guide the user in case he doesn't understand the interface	1	In none of the options is shown the help option with the interface elements designed
P8	The information of flights available in selection of two sections or more sections, is different	2	When user selects to fly to more than one destination, but only place two flights, the available flight information is different (limited) on the selected dates than shown if you enter by purchasing a flight to a single destination
P1	By placing missing information (without return flight), user can continue with the process	2	When user buys a ticket and does not select the return flight, he can select the Continue button At the initial screen of the purchase process, user can also select the Continue button

**5.3 Participants**

The four participants were students of the computer science masters of the PUCP, male, whose ages ranged from 28 to 47 years.

## 5.4 Materials

The following materials were developed for the Usability Test:

**Confidentiality Agreement:** It is a consent document where the evaluated user manifests his voluntary intention to participate in the usability test at a certain place and date.

**Previous Indications:** In order to help to participant and a brief description of the stages that will be followed in the test is made.

**Pre-test Questionnaire:** It is a questionnaire of demographic type and serves to obtain information to classify the evaluated user. In this way, the responses of the evaluated users can be contextualized.

**Post-test Questionnaire:** It allows to obtain additional information that complements the observation made during the execution of the tasks assigned.

**Task List:** It is a document that describes in detail the activities that the user will perform for each task defined. In this list, the user must detail, when required, information that is requested as a backup of what was done in the application.

For this test, the tasks have been built on the following scenarios: In the first task the user need to look for alternative (double) tickets to Miami. The dates of the holidays are from August 1 to 14, 2016 and there is a budget of S/4,700 destined for the tickets. Since it is high season, the user must utilize some promotion with which you can purchase tickets with the allocated budget (in that sense, you must use economic rates, etc.). In the second task, the user must change the flight schedule: now you want to spend three days in Miami and the rest of the time in New York. For this, the option “Multiple destinations” must be used.

**Task Compliance Observation Sheet:** It is used by the evaluator to detail the fulfillment of each activity of the task, the time spent and the pertinent observations.

## 5.5 Usability Testing Process

The test was performed individually, each user had at his disposal an evaluator who accompanied him in the process.

Each participant was presented the Confidentiality Agreement and a list of previous Indications. Each participant gave their consent and signed the indicated documents. Subsequently, each participant was given the pretest, which was filled immediately. Then, each participant was given the task list and some general inquiries were acquitted.

The recording of the interaction was started, and the user was left in front of the browser, and each evaluator took note of the Observations of Compliance of the Tasks on what he was observing.

Finally, each participant was given the post-test questionnaire, which ended the execution of the test.



## 5.6 Data Analysis and Results

### Task 1 Results

Hits presented:

- Users were able to select the options indicated.
- Those who had previous knowledge of the application, were already more familiar, so they performed tasks faster compared to less experienced users.
- Users can obtain the prices of the cheapest tickets.
- Even when users experience complications, they did not have to seek help from the system.

Inconveniences presented:

- The functionality to see the different exchange rates cannot be used correctly.
- The system is not easy to use for people who are not experienced in this type of applications.
- The system cannot complete the payment through the credit card option when the steps indicated in task 1 are performed.

### Task 2 Results

Hits presented:

- Users were able to select the options indicated, those who did not have much knowledge of the application were already more familiar.

Inconveniences presented:

- Since the users had gone through another previous task, where the functionality was different, confused the new interface for the selection of flights.
- Most of the error messages presented clearly indicate why they were presented, but users before proceeding with the process do not identify which information they lacked to complete.
- In a particular case, an error occurred that wiped all data on the screen but did not tell the user the reason for it.
- Since it is not possible to select route rates, users are not sure to continue with the purchase.
- There is no help in the system, which is necessary for users who do not have much experience in the use of similar applications.

### Data Analysis: Observations

In general, users did not use system help. In Task 1 most users had trouble getting the rate in other currencies when using Chrome. Most users had trouble getting to the VPOS payment window. In addition, in Task 2, most users had trouble selecting destinations. Finally, all had problems to select the flight with scales more economic since that information is not shown.

### Data Analysis: Post-test Questionnaire

Table 6 shows that the general appreciation of users with respect to the page evaluated is positive. They emphasize the ease of navigation and the possibility of re-using the portal. The two points that received the lowest rating was the fact that they were able to complete the tasks and overall satisfaction with the portal. This could be considered complementary, since not being able to complete a task can influence the satisfaction of use. In addition, users 1 and 2 stand out for having extreme opinions regarding the page, the first with an extremely negative opinion and the second with an extremely positive opinion. What is recommended to validate these results is to increase the number of evaluated users.

**Table 6.** Results of Post-test questionnaire

Question	Average result	
1 Fulfill tasks	2,75	Neutral
2 Sufficient and complete information	3,5	Agree
3 Easy-to-understand information available	3,25	Neutral
4 Required information easy to find	3,25	Neutral
5 Information found useful	3,5	Useful
6 Portal easy to navigate	3,75	Easy
7 Orientation on the portal	3,5	Easy
8 Satisfaction with the portal	2,75	Neutral
9 Will use the portal again	3,75	Agree

## 6 Conclusions

It emphasizes the heuristic evaluation as a tool for evaluation of usability of wide use, given its advantages in time and cost versus analysis with participation of end users. This evaluation is an expert analysis to determine if the elements of an interface comply with widely accepted principles such as Nielsen heuristics.

In the heuristic evaluation, a significant amount of usability problems have been found. It is important to mention that an expert has been used less than recommended by Nielsen. When evaluating large or multi-functional applications, it is advisable to pre-define a scope to focus the experts on what may be of interest in the evaluation. A heuristic evaluation will provide better results when focused on detecting relevant aspects for the client.

According to the above, the problems found, which concentrate on problems of lack of help, broken links and consistency errors, are detected in the heuristic evaluation, shows problems in the process of migration or changes to this LATAM platform.

With reference to the site as a whole, apart from the problems detected, in general, it is emphasized that the site meets the objectives of each functionality evaluated, problems in general are not blocking but can hinder the user's tasks. Besides, the overall evaluation of the site is positive, since it meets most of the objectives for which

it has been built, and improvement is pending on the basis of the problems detected, as well as the completion of the process of migration or change through which the indicated site is passing.

For those reasons, it concluded that two methods of usability evaluation which are heuristic evaluations and usability test users, are complement each other. When they are used together, let a better focus in user, a better diagnostic and analysis about the object of study and increase a better feedback, among other advantages.

It is suggested, for usability tests with users, to increase the number of users evaluated, with a more heterogeneous sample, in order to obtain more results that validate the conclusions.

## References

1. Nielsen, J.: Usability 101: introduction to usability (2012). <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>. Accessed 29 June 2016
2. Holzinger, A.: Usability engineering methods for software developers. *Commun. ACM* **48**(1), 71–74 (2005)
3. Paz, F., Paz, Freddy A., Pow-Sang, J.A.: Evaluation of usability heuristics for transactional web sites: a comparative study. In: Latifi, S. (ed.) *Information Technology: New Generations. AISC*, vol. 448, pp. 1063–1073. Springer, Cham (2016). doi:10.1007/978-3-319-32467-8\_92
4. Paz, F.; Pow-Sang J.A.: Usability evaluation methods for software development: a systematic mapping review. In: 8th ASEA (2015)
5. Paz, F., Pow-Sang, J.A.: Current trends in usability evaluation methods: a systematic review. In: 2014 7th International Conference on Advanced Software Engineering and Its Applications (ASEA), pp. 11–15, December 2014
6. Paz, F., Villanueva, D., Pow-Sang, J.A.: Heuristic evaluation as a complement to usability testing: a case study in web domain. In: 2014 Tenth International Conference on Information Technology: New Generations (ITNG), pp 546–551, April 2015
7. Paz, F., Villanueva, D., Rusu, C., Roncagliolo, S., Pow-Sang, J.A.: Experimental evaluation of usability heuristics. In: 2013 Tenth International Conference on Information Technology: New Generations (ITNG), pp. 119–126, April 2013
8. lan.com: Nace LATAM: La nueva marca que adoptarán LAN, TAM y sus filiales. <http://www.lan.com/destinosudamerica/notadeprensa/nace-latam-la-nueva-marca-que-adoptaran-lan-tam-y-sus-filiales/>. Accessed 11 June 2016
9. Nielsen, J.: Severity ratings for usability problems (1995). <https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/>. Accessed 29 June 2016