User Interface Evaluation of the CP's Website (Proposal)

Luís Cruz[1]

[1]*MAP-i, Joint Doctoral Programme in Computer Science*

January 4, 2014

Abstract

This document proposes an usability evaluation for the website of the company *CP – Comboios de Portugal*. One analytical method and one empirical method are going to be applied in this evaluation: *Heuristic Evaluation* and the *Usability Test*, respectively, both described in this document.

Through the Heuristic evaluation it was possible to find X issues in the user interface. As a result of the usability test a report complying with the CIF standards, an usability test plan, and a usability post-test questionnaire were produced. The usability test assessed the efficiency, efficacy and satisfaction measures of the system.

talk about

1 Introduction

This project aims to evaluate the user interface of CP.pt[1] — the official website of *CP - Comboios de Portugal, E.P.E.*

CP is a public portuguese company responsible for rendering national and international passenger rail services. In the year 2012, CP had 4690 employees, transported 122 million passengers and almost 8713 thousand metric tons (CP - Comboios de Portugal, 2012). They provide 3 main kinds of rail transportation services: *urban* in the cities of Oporto and Lisbon; *National* with regional services and the fast lines of *Alfa Pendular* and *Intercidades*; and *International*.

Through the website, CP's customers can check the timetables, buy tickets, get information about the available lines and special offers and read some news related with CP services. In order to buy tickets, the website provides the *netTicket* service, which requires the customers to have an account in their *myCP* service and it is only available for the long distance trains *Intercidades* and *Alfa Pendular*.

According to the website the graphical interface was optimally designed for windows with 800×600 pixels of resolution. A view of the website is depicted in the figure 1, using a window with the same resolution.

2 Users and Context

CP customers vary according to the service provided. Many college students, workers and pensioners use the regional and urban services for small and medium distances. Long distance services are more used by college students that are away from home, tourists, and executive workers. Unfortunately, no official document stating the segmentation of the CP.pt website's users was found.

It is noticeable that CP services have a lot more passengers during school time, which means that students are an important segment of CP's customers. Besides, most of the students have good experience with the WEB, so the CP.pt website is expected to be a great tool to them. Therefore, this usability evaluation will focus in the segment of college students, which might be portuguese citizens as well as foreigners that study or want to study in Portugal and are able to speak English.

Many scenarios can apply for the use of the website by students. Some times they leave the classes earlier and need a way of quickly check if there are other alternative trains that can take them home earlier. Also sometimes there is no direct train to their destination, so they have to catch another in the the middle of the traveling. Another scenario is when the weekend is over and the student has to buy

[1]Available at: `http://www.cp.pt`

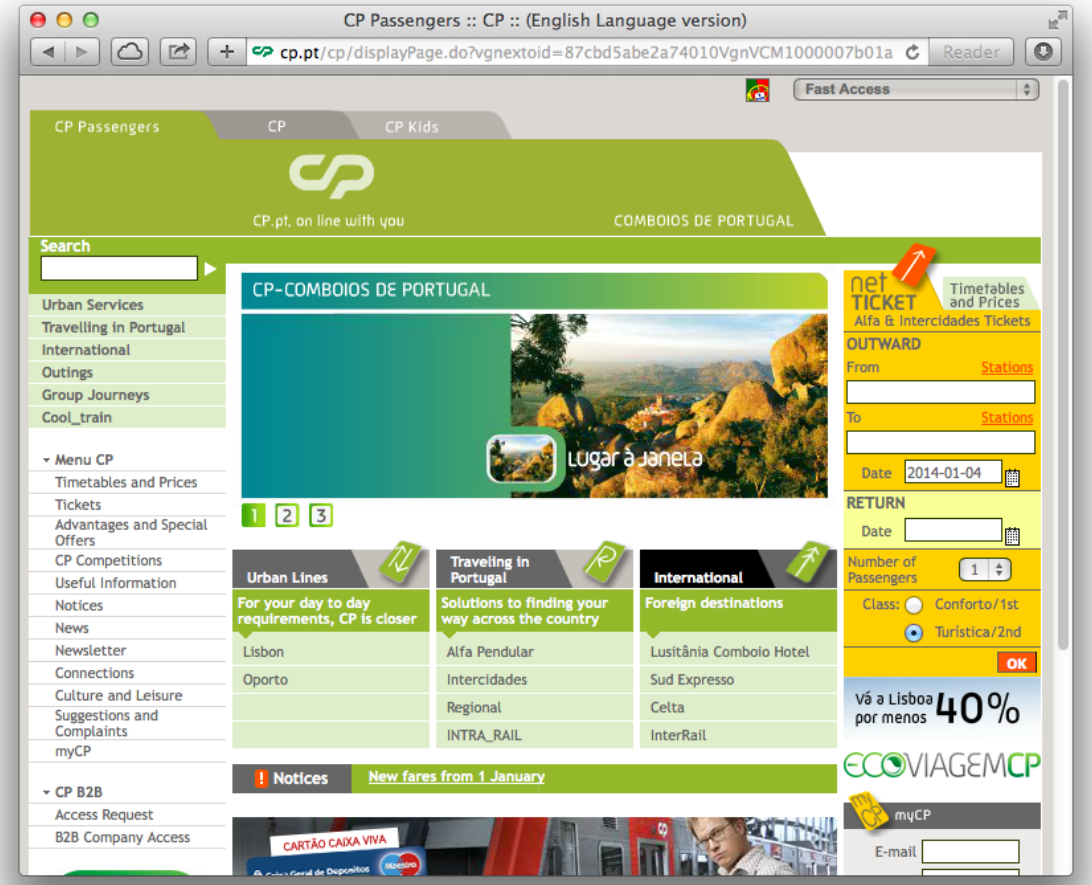

Figure 1: View of the main page of CP.pt website in a window with resolution 800 × 600 pixels, using the internet browser Safari Version 7.0.1.

his/her ticket from home to his/her university city. Buying it from the website is more convenient since the student can avoid wasting time in the ticket lines and can grant a seat for his/her trip. Therefore, the following two contexts are considered the most important in terms of usability:

- Check the timetable to find any suitable train for the trip and the respective prices.
- Buy a ticket for long distance trains with reserved seats.

3 Usability Evaluation Methodology

The evaluation will be taken using two paradigms: *Analytical* and *Empirical*.

Analytical methods do not need to involve users — they are based on inspection methods. Some well known analytical methods are the *Heuristic Evaluation* (HE) proposed by Nielsen and Molich (1990), the *Cognitive Walkthrough* (Wharton et al., 1994) and its variant *Streamlined Cognitive Walkthrough* (Spencer, 2000).

Empirical methods involve the user in the evaluation process through Usability tests, involving *observation* and *query* techniques, through *controlled experiments*, in a more scientific approach, or even *questionnaires*, *focus groups*, etc.

In this evaluation, the used analytical method will be the *Heuristical Evaluation* and the empirical method will be the *Usability Test*. These methods are described in the next sections.

4 Heuristic Evaluation

The elected analytical method for this evaluation was the *Heuristic Evaluation*, because it is cheap, intuitive, easy to motivate people to do it and provides that useful results can be obtained (Nielsen and Molich, 1990).

4.1 Methodology

This method proceeds by having a small set of evaluators judging the system according to some general principles of interaction design, *usability heuristics*. It has been shown that a number of evaluators between 3 and 5 provides good results and that there is no point in having more than 10 evaluators (Nielsen and Molich, 1990). Nielsen (1995a) proposed the 10 most important usability heuristics for User Interface Design:

- Visibility of system status
- Match between system and the real world
- User control and freedom
- Consistency and standards
- Error prevention
- Recognition rather than recall
- Flexibility and efficiency of use
- Aesthetic and minimalist design
- Help users recognize, diagnose, and recover from errors
- Help and documentation.

Heuristic evaluation was originally developed for evaluators who had some knowledge in usability but who were not necessarily usability experts (Nielsen and Molich, 1990), however, it has been showed that the method is also very effective for expert evaluators (Nielsen, 1992).

In order to group the findings of this heuristic evaluation, the severity ranking proposed by Nielsen (1995b) with a scale from 0 to 4:

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.

In this evaluation, two experienced evaluators were asked to focus on the interface elements related with the following top-level user stories:

- As an unregistered user I want to check time tables and prices for traveling from a given train station to another one.
- As an unregistered user I want to buy a ticket for the fast train services *Alfa Pendular* and *Intercidades* (it may include a registration process).

4.2 Results

After completing the heuristic evaluation 11 problems were found. They were sorted in table 1 according to the severity rating.

number of problems

#	Problem	Severity Rating	Violated Heuristic(s)
1	The "Timetable and Prices" form is very similar with the form for buying tickets. Non expert users might not notice the difference.	3	Error prevention.
2	There is an auto complete feature in the "Timetable and Prices" and netTicket forms, which whenever the user misspells a letter of the station name, the system may complete with another station's name, and the user has to hit the backspace button one more time than usual.	3	Error prevention; Help users recognize, diagnose, and recover from errors.
3	The seat selection is made in a uncommon way. One has to first click on the previous seat and then on the desired seat.	3	Consistency and standards.
4	The services of trains are defined as acronyms which are not used by the users (eg., IC means Intercidades)	2	Consistency and standards; Match between system and the real world.
5	When an unregistered user wants to buy a ticket after choosing a specific train, the system asks him/her to register but after finishing the registration, he/she has to search again for the same train.	2	Recognition rather than recall.
6	In order to check the price in the timetable it is necessary to click in a link which when clicked replaces itself into the respective price, but only for one train at a time.	2	Flexibility and efficiency of use.
7	When searching for a traveling ticket to buy, the user selects whether is traveling in first or second class in the beginning of the search. It is defined as second class by default and might not be perceptible. Through the next steps the user cannot change it.	2	User control and freedom.
8	In the buying process there is no back button, besides the one provided by the browser that is not recommended by the system since he presents an alert warning that the user is about to leave the page.	2	User control and freedom.
9	In the timetable results, there is a clickable down arrow that does not do anything	1	Aesthetic and minimalist design; Consistency and standards.
10	There is a Time input field when in the form "Timetables and Prices", which might be unnecessary and provides no useful filtering in the results.	1	Aesthetic and minimalist design.
11	In the timetable results there is a number identifying the row that does not provide any useful information	1	Aesthetic and minimalist design.

Table 1: Problems found in the Heuristic Evaluation.

5 Usability Test

The usability test was carefully reported using the Common Industry Format (CIF) (Stanton, 2006) in a document entitled *Usability Test Report* which is provided in the appendix A. This section briefly describes the usability test, but for further details please refer to the usability test report.

Furthermore, in the scope of this test, an *Usability Test Plan*, and a *Post-Test Questionnaire* were designed, being all available in the appendix of the usability test report (see appendix A). Additionally, all material and data can be downloaded in this assignment project's website[2].

5.1 Participants

According to Nielsen (2012), 5 users is the most adequate size of the sample and provide good insights of the usability of a product. However, when a quantitative study of the usability is performed, at least 20 users should participate. The author also stated that for low-overhead projects, it is often optimal to test 2 users per study.

[2]This assignment is published in a website available at `http://paginas.fe.up.pt/~luiscruz/cp_usability/`

The CIF standards provided an useful framework to get quantitive measures of usability. However, due to the academic nature of this study and the small budget, two users made part of this test which can provide some useful insights about the usability. Two college students complying ages between 21 and 26 years were invited to participate. They usually have to travel in order to get to their universities and all participants had a moderate level of experience using WEB applications.

5.2 Methodology

The users were invited to perform some tasks regarding some of the most used features in the website:

1. Select the English Version of the application.
2. Find the schedule for a trip from Braga to Aveiro.
3. Find a cheap trip from Braga to Aveiro.
4. Buy a ticket from Braga to Porto.

As suggested in (Mitchell, 2007) it is beneficial to get the participant's opinion during the test session. The usability test combined both *observation* and *query*.

The participants were asked to use the think aloud behavior (TA), describing every step they made during the tasks. The moderator was directly observing the participant while taking some notes using the *Data Logging Form*, and after the completion of each task, users were interviewed in a casual way, trying to answer some questions clearly stated in the *Usability Test Plan* document. In addition, the screen and audio were also recorded for indirect observation.

All this setup intended to evaluate the system in terms of *effectiveness*, *efficiency*, and *user satisfaction*:

- Effectiveness
  - Completion Rate
  - Unassisted completion rate
  - Number of assistances
  - Number of steps made differently
  - Back Button hits
  - Errors
- Efficiency
  - Task time
  - Completion rate efficiency
- User satisfaction
  - SUS scale

After the completion of each task, users were interviewed in an informal way, trying to answer some questions clearly stated in the *Usability Test Plan* document. Moreover, after the whole test, participants gave some feedback through the *Post-Test Questionnaire*.

5.3 Main Results

All users were able to successfully complete the given tasks, and the assisted completion rate was 75%.

The first task, selecting the english version, was very simple and users took just 8 seconds to complete. This task proved that the feature is well implemented, and was important to make the participant comfortable with the test environment. The task 4, buying a ticket, was more complex and took longer for the users to complete, with an average of 6 minutes and 27 seconds. The efficiency was also measured through the completion rate efficiency, resulting in a mean of $38\%/min$.

The time results suggest an interesting fact in the times for task 2 and task 3. The complexity of these tasks is very similar and the steps to complete are identical. However, task 2 took in average 2

minutes and 38 seconds, while task 3 reduced this time to 1 minute and 9 seconds. This can be an indicator that the system might be well suited for second time visitors.

Some times the user was confused and made wrong steps, but easily recognized and hit the back button. This inefficiency was reported with an average of 2.5 hits per user during the whole test.

The SUS questionnaire scored 60 points. In (Sauro, 2011), the average SUS for 500 systems is 68, which means that this website has a satisfaction level considered below average. According to a SUS qualitative classification proposed in (Bangor et al., 2009), the overall satisfaction perceived from users is **good**.

Observed issues and users's feedback One very important contribution of the interviews and the post-test questionnaire was to get the most from the user's opinion. This procedure successfully provided some spots in the features that need improvement.

Users are satisfied with the graphical appearance of the website. However, they stated that is too much information in the view.

An important issue was the fact of the "Timetable and Prices" form is not visible but is very similar to the *netTicket* form which is visible. Users frequently made the mistake of using the wrong form. Another component users didn't like was the table results. They felt the need of a more readable table that provided some filtering, at least for price range and line service.

Another difficulty faced by first time users, was the feature of selecting a seat for the ticket. The user interaction was not intuitive and users had some negative remarks about that.

5.4 Improvements and Pilot Testing

Although some statistical analysis was performed, the sample of users is not representative due to its small size (two individuals). This means that, although some important considerations were taken and which might help to improve the usability of the website, the measures reported cannot be considered as valid. The state of this study can be seen as the starting point for a rigorous usability test with more participants. If a quantitative measure of usability is required, at least 20 users should take part of the test (depending on the kind of results) (Nielsen, 2012).

Nevertheless, this experiment allowed to perform pilot tests and validate the design of the test: some aspects of the test were iteratively improved, and some other might need some changes in future work.

Some of the changes made were regarding the interview for each task. The interview guidelines tried to answer the same generic questions in different tasks, which did not provide any extra information. This was improved from the first to the second participant. Another small fix that had to be made was to increase the size of the fields in the post-test questionnaire, since they were too small.

In the first attempts for the screen and audio recording, each task was being recorded separately. This was leading to an interruption of the flow of the test, because the moderator had to start and stop the recording each time. Thus, it was decided to record the session in a single take.

One of the main difficulties was the think aloud technique. It is hard to make a participant feel really comfortable to explain every single step. Also, it depends on the profile of the participant, some participants might be more quiet than others. Another aspect to improve is the fact of users sometimes want to prove they have good skills, and do not want to make mistakes, which makes harder to reproduce the usual behavior of the user.

Regarding the Data Logging Form, some measures did not provide useful measures and could be removed in future experiments in order to simplify the observation process. This is the case of *time start*, *time end*, and *number of negative remarks*. Regarding *negative remarks*, it is important to write down this feedback in the notes field, but the number of remarks was not very useful in this case.

Another change that could be made is related with the interviews made after each task. They are very useful, but some of the questions made were not.

6 Conclusions and Future Work

This document reported the usability evaluation of CP.pt website. The *Heuristic Evaluation* was very effective, it allowed to find 11usability problems, from which 3 were considered to be high priority fixes. FIXME Besides, it was very cheap and easy to prepare. Before proceeding to the usability test, this issued should have been fixed.

The usability test provided was carefully prepared and two college students participated in the experiment. An usability test plan with some guidelines, a post-test questionnaire, a SUS questionnaire, and

a data logging form were developed to get a very detailed information about the usability. The results suggest that website is very effective, since all users were able to correctly complete all the tasks, with an assisted completeness rate was of 75%.

The SUS instrument allowed to classify the user's satisfaction of the site as good, with a score of 60. However, this result is below average, which is standardized as 68.

, interview and questionnaire

Although the size of the sample was to small, the experiments with these 2 participants, provided some useful insights of the usability and served as a pilot test, allowing to improve the test procedure and tools. As future work, after fixing the spotted issues, the designed usability test should be conducted with a more representative sample.

A Usability Test Report

The original document is available at `http://paginas.fe.up.pt/~luiscruz/cp_usability/`

Usability Test Report of CP's Website (Proposal)

Tested by: Luís Cruz[1]

[1]*MAP-i, Joint Doctoral Programme in Computer Science*

Tested: January 20, 2014

Common Industry Format for Usability Test Report v1.1
Date prepared: January 27, 2014
Prepared by: Luís Cruz (luiscruz@fe.up.pt)

Executive Summary

This document intends to report the usability test for the website *CP.pt* in the scope of the unit course of Interactive Computing Systems(ICS) of the MAP-i doctoral programme. The usability test was conducted by inviting two potential users to accomplish four tasks, using the *think-aloud* technique, while using some behavior analysis/acquisition methods. The participant's activity was recorded through screen and audio cast, and the moderator observed the participant by taking notes in the provided Data Logging Form. After each task, a short interview had place, and after the whole test, users were asked to answer a customized 30 questions survey, as well as the System Usability Scale (SUS) questionnaire.

The recruited participants were screened in order to focus on college students that need to travel to their universities.

With this study some usability issues were spotted, but the results were satisfactory. All the users were able to complete all the tasks, and besides some inefficiencies in some steps, the completion rate efficiency was $38\%/min$. The overall website satisfaction of the users was perceived as *good*, assessed using the SUS scale that scored 60 out of 100. The feedback obtained from users through the interviews helped to find some issues in the user interface that need special attention.

1 Introduction

1.1 Product Description

CP.pt is the official website of emphCP - Comboios de Portugal, E.P.E, the public portuguese company responsible for rendering national and international passenger rail services.

CP customers vary according to the service provided. Many college students, workers and pensioners use the regional and urban services for small and medium distances. Long distance services are more used by college students that are away from home, tourists, and executive workers. Unfortunately, no official document stating the segmentation of the CP.pt website's users was found.

It is noticeable that CP services have a lot more passengers during school time, which means that students are an important segment of CP's customers. Besides, most of the students have good experience with the WEB, so the CP.pt website is expected to be a great tool to them. Therefore, this usability evaluation will focus in the segment of college students, which might be portuguese citizens as well as foreigners that study or want to study in Portugal and are able to speak English.

Many scenarios can apply for the use of the website by students. Some times they leave the classes earlier and need a way of quickly check if there are other alternative trains that can take them home earlier. Also sometimes there is no direct train to their destination, so they have to catch another in the the middle of the travelling. Another scenario is when the weekend is over and the student has to buy his/her ticket from home to his/her university city. Buying it from the website is more convenient since the student can avoid wasting time in the ticket lines and can grant a seat for his/her trip.

From the features implemented in the website, the following were tested:

- Choose between Portuguese or English versions
- Check timetables for a trip.
- Buy fast train tickets, being the features that were tested.

1.2 Test Objectives

The aim of the test was to validate the usability of the main features of the CP.pt — finding the most suitable train and buying tickets. It is important that these tasks are easy to learn.

Representative users were asked to complete some tasks, measures were taken of effectiveness, efficiency and satisfaction, and some notes about the users' opinion were taken in order to have ideas for some improvements that can be made.

2 Method

2.1 Participants

This test had 2 participants, both college students with ages between 21 and 27 years old. Regarding the small size of the sample, 100% of the participants were female.

They are intermediate level users, that frequently use WEB applications and already have experience in other transportation company's websites (e.g., *Ryanair*). This characteristics are expected in the population of college students, except for the gender that should be in equal portion, and the ages that could be more representative — students enter to college in ages starting from 17 years old.

2.2 Context of Product Use in The Test

2.2.1 Tasks

The tasks that the participant has to accomplish are the following:

1. Select the English Version of the application.
2. Find the schedule for a trip from Braga to Aveiro.
3. Find a cheap trip from Braga to Aveiro.
4. Buy a ticket from Braga to Porto.

These tasks are described in more detail in the *Usability Test Plan*, available in the appendix B. For each task, all the steps were defined in order to describe how the task can be efficiently accomplished.

These tasks were selected with the criteria of being features that are expected to be of the most important use. Every transportation website has these features and add great value to the customers, so it is important that *CP.pt* meets the users needs.

All the completion and performance criteria are also described in the Usability Test Plan (see appendix B).

2.2.2 Test Facility

The test was made in a study room at the faculty. The moderator is sitting next to the participant in order to make the observation and query and give assistance. The screen and audio were recorded using the tool QuickTime Player 10.3 which is invisible to the user and does not affect the user experience.

2.2.3 Participant's Computing Environment

According to (Prakash, 2013), the most common resolution used in WEB is 1366×768. In this experiment a 13 inches RGB screen with approximately the same resolution was used: 1440×990. For interaction with the application the participant used an Apple laptop keyboard, and an Apple Magic mouse. Some other devices were available if the user didn't feel comfortable with these devices, however, they were not used.

The browser Safari 7.0.1 with default settings was used with an internet connection which had an average download and upload speeds of $4Mbit/s$ and $1Mbit/s$, respectively.

2.2.4 Test Administrator Tools

A *Data Logging Form* (see C) was designed providing the moderator with a tool to record some notes about each task of each participant. The form has some fields for a few variables and a space to take some notes for the post-task interview.

All the task information is provided in the *Usability Test Plan*, available in the appendix B, defining a script about how the moderator should conduct the experiment is provided. All the steps that are necessary to finish a task are clearly described as well as some guidelines with the important questions for the post-task interview.

After the test, the participants were asked to answer a post-test questionnaire based on (Henry and Thorp., arch) provided in appendix D and the SUS questionnaire available in the literature (Brooke, 1996).

As it was already mentioned, during the experiments, the screen and voice were recorded using the tool QuickTime Player 10.3.

2.3 Experimental Design

Two college students were tested. For each of the four tasks, described above, the completeness, number of assistances, number of steps made differently, back button hits, number of errors, and the duration of the task were recorded. The completion rate, unassisted completion rate, and completion rate efficiency were also calculated as .

2.3.1 Procedure

The participants were informed that the usability of CP's website was being tested, to find out whether it met the needs of users. They were told that it was not a test of their abilities. They were asked to sign a consent form.

Participants were given introductory instructions. The evaluator reset the state of the computer before each task, and provided instructions for the next task.

The participant could ask for assistance and make questions whenever they find necessary, in order clarify any part of the task. All assistances were logged by the moderator. Also there was no time limit for the task completions, but if the moderator feels that the participant is stuck in some part, he was allowed to give some hints if properly logged.

After each task the moderator conducted a small interview trying to answer some crucial questions provided in the Usability Test Plan (see appendix B).

In the end of the test, a post-test questionnaire based on (Henry and Thorp., arch) was given to the participants (available in appendix D). This questionnaire has a Likert response format in a scale from 1 to 5, and intends to provide some insights about the way the design should be changed to make it more suitable to the user.

In addition to this questionnaire, the users are asked to fill the SUS questionnaire (Brooke, 1996).

The participants were non remunerated voluntaries and during each test session only the moderator and a participant were present.

2.3.2 Participant General Instructions

The instructions were given personally by the moderator to each participant. The test session proceeds by having only one user in the room with a moderator. Whenever the user needed help he/she could simply ask for help.

The participant was asked to use the *think-aloud* technique, describing every step he/she makes during the tasks.

2.3.3 Participant Task Instructions

Before starting each task the moderator explained what it was expected to accomplish in the following task. The task instructions are very simple, being described in a short sentence, as stated in section 2.2.1.

2.4 Usability Metrics

2.4.1 Effectiveness

For measuring effectiveness the following measures were considered:

Completion Rate the percentage of participants that correctly finished each task.

Unassisted Completion Rate the percentage of participants that correctly finished each task without assistance.

Number of Assistances The average of assurances given in each task.

Number of steps made differently number of steps that the participants made differently to the ones described in the test plan.

Back Button hits The average of times the user hit the Back Button in the browser.

Errors The number of times a user had to repeat parts of the task.

2.4.2 Efficiency

Efficiency was accessed by measuring the following parameters:

Task time The average time the users took to correctly complete each task.

Completion rate efficiency mean completion rate/mean task time (in minutes).

2.4.3 Satisfaction

Satisfaction is a subjective measure that correlates with the user's motivation to use a product. The standardized instrument *System Usability Scale* (SUS) provides a 10 item questionnaire with five-scale responses that can be converted into a score (Brooke, 1996). After the test the users answered this questionnaire.

Also a post-test questionnaire based on (Henry and Thorp., arch) was given to the participants (available in appendix D). Although this questionnaire does not provide a metric result, it is an useful tool to obtain user's feedback.

3 Results

3.1 Data Analysis

3.1.1 Data Scoring

The method by which the data collected were scored should be described in sufficient detail to allow replication of the data scoring methods by another organization if the test is repeated. Particular items that should be addressed include the exclusion of outliers, categorization of error data, and criteria for scoring assisted or unassisted completion.

3.1.2 Data Reduction

The combined results of all tasks show the mean results for the effectiveness and efficiency metrics. Based on the data acquired in each task, combined results were derived with mean results for effectiveness and efficiency metrics.

3.1.3 Statistical Analysis

The method by which the data were analyzed should be described in sufficient detail to allow replication of the data

3.2 Presentation of the Results

Effectiveness, Efficiency and Satisfaction results must always be reported

3.2.1 Performance Results

Task 1 This task was extremely easy for all participants. Every user completed the task correctly with no assistances and no inefficiency. The results observed in task 1 are reported in table 1. **Completion rate efficiency** is $13\%/s$. Although this was a trivial task, it was important for the user to get familiar with the test setup and procedure.

Table 1: Results observed in **task 1**.

Participant ID	Completion	Number of Assistances	Number of steps made differently	Back Button hits	Errors	Time
1	1	0	0	0	0	0'08"
2	1	0	0	0	0	0'08"
Mean	1	0	0	0	0	0'08"
Std Dev	0.0	0.0	0.0	0.0	0.0	0'00"
Min	1	0	0	0	0	0'08"
Max	1	0	0	0	0	0'08"

Task 2 The results observed in task 2 are reported in table 2. **Completion rate efficiency** is $38\%/min$.

Table 2: Results observed in **task 2**.

Participant ID	Completeness	Number of Assistances	Number of steps made differently	Back Button hits	Errors	Time
1	1	1	1	2	3	3'50"
2	1	0	2	1	0	1'25"
Mean	1	0.5	1.5	1.5	1.5	2'38"
Std Dev	0.0	0.5	0.5	0.5	1.5	1'13"
Min	1	0	1	1	0	1'25"
Max	1	1	2	2	3	3'50"

Task 3 The results observed in task 3 are reported in table 3. **Completion rate efficiency** is $87\%/min$. Although the task was very similar to the previous one, we can see that, in average, users needed less time to accomplish it (mean time was 1 minute and 9 seconds, which is a considerable decrease). This might be an indicator that the system is a lot easier for second time visitors.

Table 3: Results observed in **task 3**.

Participant ID	Completeness	Number of Assistances	Number of steps made differently	Back Button hits	Errors	Time
1	1	0	0	0	0	0'50"
2	1	0	0	0	0	1'28"
Mean	1	0	0	0	0	1'09"
Std Dev	0.0	0.0	0.0	0.0	0.0	0'19"
Min	1	0	0	0	0	0'50"
Max	1	0	0	0	0	1'28"

Task 4 The results observed in task 4 are reported in table 4. **Completion rate efficiency** is $16\%/min$. Since this task was more complex, it took a lot more than the others: in average, users took 6 minutes and 27 seconds to complete.

Table 4: Results observed in **task 4**.

Participant ID	Completeness	Number of Assistances	Number of steps made differently	Back Button hits	Errors	Time
1	1	3	0	0	0	6'45"
2	1	0	1	2	0	6'08"
Mean	1	1.5	0.5	1	0	6'27"
Std Dev	0.0	1.5	0.5	1.0	0.0	0'19"
Min	1	0	0	0	0	6'08"
Max	1	3	1	2	0	6'45"

Combined Results The overall results of efficiency and effectiveness are combined in the table 5. It is possible to see that all users were able to finish all the tasks. These results suggest that some of the tested features are not very intuitive, since the back button was hit by users with an average of 2.5 times. Besides, 25% of the experiment tasks were completed with assistance. The average time to complete is 2 minutes and 35 seconds, however, this measure is lowered by the fact that task 1 was really simple and took 8 seconds for the users to complete. In addition, the mean **completion rate efficiency** is $38\%/min$ — ignoring the data from first task, this measure would decrease to $29\%/min$.

Table 5: Combined Performance Results.

Participant ID	Completion Rate (%)	Unassisted Completion Rate (%)	Number of Assistances	Number of steps made differently	Back Button hits	Errors	Time
1	100	50	4	1	2	3	2'53"
2	100	100	0	3	3	0	2'17"
Mean	100	75	2	2	2.5	1.5	2'35"
Std Dev	0	25	2.0	1.0	0.5	1.5	0'18"
Min	100	50	0	1	2	0	2'17"
Max	100	100	4	3	3	3	2'35"

3.2.2 Satisfaction Results

Table 6: SUS results

Participant ID	Total Score	SUS Score
1	25	62.5
2	23	57.5
Mean	24	60.0
Min	23	57.5
Max	25	62.5

The mean SUS score achieved is 60 (out of 100). Using the *Adjective Rating Scale* proposed by Bangor et al. (2009), detailed in the figure 1, the website was perceived by the participants as **good**.

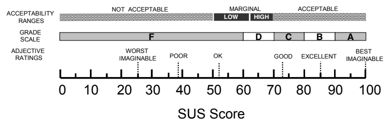

Figure 1: A comparison of the adjective ratings, acceptability scores, and school grading scales, in relation to the average SUS score. Source: (Bangor et al., 2009)

Table 7: Post-test questionnaire results

Question	User 1	User 2	Mean
1. The homepage is attractive	2	4	3
2. The overall site is attractive	2	4	3
3. The site's graphics are pleasing	3	4	3.5
4. The site has a good balance of graphics versus text	3	4	3.5
5. The colors used throughout the site are attractive	3	4	3.5
6. The typography is attractive	3	4	3.5
7. The homepage's content makes me want to explore the site further	3	3	3
8. It is easy to find one's way around the site	4	2	3
9. You can get information quickly	3	2	2.5
10. It is fun to explore the site	2	2	2
11. It is easy to remember where to find things	4	4	4
12. Information is layered effectively on different screens	3	4	3.5
13. The homepage is attention-getting	3	3	3
14. Information is easy to read	3	3	3
15. Information is written in a style that suits me	3	4	3.5
16. Screens have the right amount of information	2	4	3
17. The site effectively communicates the company's image	4	4	4
18. Information is relevant	4	4	4
19. The site is designed with me in mind	3	3	3
20. The site's content interests me	4	5	4.5
21. The site's content would keep me coming back	4	5	4.5
22. The site has characteristics that make it especially appealing	3	3	3
23. The site reflects progressive, leading edge design	2	3	2.5
24. The site is exciting	2	3	2.5
25. The site is well-suited to first-time visitors	2	1	1.5
26. This site is well-suited to repeat visitors	2	4	3
27. The site has a clear purpose	3	4	3.5
28. It is always clear what to do next	3	2	3.5
29. It is clear how screen elements work	2	5	3.5
30. Mistakes are easy to correct	4	1	2.5

The results of the post-test questionnaire are presented in the table 7. From all the answers the most relevant notions we get are that users agree that the website has relevant content that meets the users' needs and that the provided information is relevant. However, they state that it is not well-suited for first-time visitors, and navigating through the website is not fun.

3.3 Observation and Interview Results

As mentioned earlier, after each task, a small interview was performed in order to get the most from user's opinion. Besides, the moderator was standing next to the participant in order to find out the cause of some user's errors or inefficiencies. It allowed to spot some problems that can be solved in order to improve the website usability.

Some remarks made from the users was that the website is not well organized.

Task 1 - Select the English Version Users understood clearly what they had to do, and recognized the british flag as a familiar icon for choosing the english version.

Task 2 - Find a train from Braga to Aveiro During this task some usability problems were perceived. The "Timetable and Prices" form is not easy to find. Even experienced users made the mistake of using the NETTicket form which only provides trains for some services. Also, one user that is used to travel in the Urban service lines complained that train connections that take some time (e.g. 30 min) were not available, but when she travels usually prefers these ones for being cheaper. Also, there was some information that they found necessary to ask in the ticket office: discount prices and in which line do they have to switch in the middle of the the trip.

Task 3 - Find a cheap train from Braga to Aveiro In this task, users complained about the timetable results table. They found annoying having to click in a link every time they want to check one train's price, and suggested to introduce some filters according to price, service line, etc.

Task 4 - Buy a ticket from Braga to Porto The major difficulty was to select the seat, it was not intuitive, but after knowing how to use it they were able to complete the task with no problem. During the registration process, users didn't understand the form asking information about the most preferred lines. Furthermore, there is an important issue that should be fixed: when an unregistered users searches for a ticket and decides to buy it, the system redirects to the registration process; however, after registering, the user has to search again for the train, which is very inefficient and might be frustrating for the user.

The screen and audio recording of the test including interviews is available at this usability project's website (url: `http://paginas.fe.up.pt/~luiscruz/cp_usability/`).

A Consent and Recording Release Form

The present *Consent and Recording Release Form* was adapted from a template available at *Usability.gov* website[1].

CONSENT AND RECORDING RELEASE FORM

I agree to participate in the study conducted and recorded by the MAP-i student Luís Cruz.

I understand and consent to the use and release of the recording by Luís Cruz. I understand that the information and recording is for research purposes only and that my name and image will not be used for any other purpose. I relinquish any rights to the recording and understand the recording may be copied and used by Luís Cruz without further permission.

I understand that participation in this usability study is voluntary and I agree to immediately raise any concerns or areas of discomfort during the session with the study administrator.

Please sign below to indicate that you have read and you understand the information on this form and that any questions you might have about the session have been answered.

Date: ____________
Please print your name: ______________________________
Please sign your name: ______________________________
Thank you!
Your participation is kindly appreciated.

[1] (url: `http://usability.gov`)

B Usability Test Plan

The plain original document can be accessed at `http://paginas.fe.up.pt/~luiscruz/cp_usability/`

Usability Test Plan for CP's Website

Luís Cruz[1]

[1] *MAP-i, Joint Doctoral Programme in Computer Science*

January 22, 2014

1 Introduction

This document serves as the session moderator's script, outlining the series of activities participants undertake during the usability test. This document was iteratively designed through some pilot tests, in order to get the most from the participants.

Four tasks are described with the detailed steps for completing them, and for each of them the outlines for the post-task interviews are provided.

2 Tasks

The computer is recording the screen and audio and set with an open browser to: `http://www.cp.pt/`.

In this section the tasks are detailed with all the necessary steps in order to help the moderator to help the user. However, unless the user needs some help he/she is not supposed to read it.

During each task, the moderator reports some variables in the provided Logging Form Sheet, and whenever a task completes a small interview is proceeded between the moderator and the participant.

Task 1 - Select the English Version In this task the user has to change the language of the application to English:

i. Go to the CP.pt homepage (url: `http://www.cp.pt/`)
ii. Select the british flag on the top bar.

Completion Criteria: The website has to be in the English language.

Post-task Interview:

1.1. Is this the first time you visit this website?
1.2. Please give me your first impressions about the layout and design of the website.
1.3. How easy or difficult was it for you to accomplish this task?
1.4. Was there something specific that made this task easy or difficult?

Task 2 - Going from Braga to Aveiro In this task the user is asked to find when is the next train from Braga to Aveiro, how much is the ticket and where is it necessary to switch lines. In order to accomplish this task, the user has to:

i. Go to the CP.pt homepage
ii. Find the form through the tab "Timetables and Prices" of the right panel.
iii. Introduce the details of the trip, from Braga to Aveiro in the present day. Optionally, the time can be set, and it is not necessary to specify the return trip.
iv. Select one of the listed trips, by selecting "see" and/or "detail" links.
v. Report the time of departure, the price of the ticket, the duration and how long will the passenger have to wait in the station, and when is the departure of the second train.

1

Completion Criteria: The user is able to say when is the departure, when the train arrives to the destination and where is the train scale.

Post-task Interview:

2.1. How easy or difficult was it for you to accomplish this task?
2.2. Was it easy to find the form?
2.3. Did you like the way you had to input the trip details?
2.4. Did you clearly understand the results table?
2.5. Which part of the task you found more confusing?
2.6. When you clicked in the details of your trip, did you find it easy to check when you have to switch trains?
2.7. Do you feel that you need to ask some extra information to the ticket officer?

Task 3 - Find a cheap ticket from Braga to Aveiro This task is similar to the previous task 2 but this time, the passenger only has 10 euros to spend in the trip. The steps are the same, but in this case the user cannot select a fast train IC or AP, because they are more expensive.

Post-task Interview:

3.1. How easy or difficult was it for you to accomplish this task?
3.2. Which part of the task you found more confusing?
3.3. When you clicked in the details of your trip, did you find it easy to check when you have to switch trains.
3.4. What is your opinion of how price information is displayed?

Completion Criteria: The user chose a train from Inter-regional or Urban services and he/she is able to say when is the departure, when the train arrives to the destination, and how much is it going to cost.

Task 4 - Buy a ticket from Braga to Porto In this task the user will be using the netTicket feature to buy a ticket for the train *intercidades* (IC) from Braga to Porto. This will include the registration in myCP service and the task ends before the payment step in order to simplify the test setup. This task intends to analyse how an unregistered user manages to buy a ticket. This might be a difficult task.

The steps are the following:

i. Go to the CP.pt homepage.
ii. Find the form netTicket in the right panel.
iii. Insert in the *from* field, the value *Braga*, and in the field *To* the value *Porto - Campanhã*. Specify the dates for the trip, including return.
iv. Choose the two trains for the round trip and click "continue" to proceed to the next step.
v. Since the user is not registered yet, the user clicks on the link "Register"
vi. The user fills the registration form submits.
vii. A new form asking for Preferred/Most used Service appears. It is optional, the user can skip it.
viii. A new form asking if CP can use the email for newsletter. The user can now finish the registration.
ix. The user has to resubmit the trip information.
x. Now the user identifies the passenger with his/her name and ID Card number
xi. Select one seat different from the default when available. In the end click "Confirm"

2

Completion Criteria: The user is on stage 5 "Payment" of the buying process. In the description of the page, the details of a trip from *Braga* to *Porto - Campanhã* are summarized.

Post-task Questions:

4.1. How easy or difficult was it for you to accomplish this task?
4.2. Which part of the task you found more confusing?
4.3. What did you think about the registration process?
4.4. When you had to select the seat for your ticket, the interface was familiar? Did you have any trouble understanding how it works?
4.5. What is your about opinion how the price information is displayed?
4.6. In the whole task was there any steps that you found unnecessary? Which?

3 System Usability Scale

After all the tasks have been completed, the moderator asks the participant to answer the *System Usability Scale* (SUS), trying to record his/her immediate response to each item.

4 Post-test Questionnaire

In order to have a more detailed impression of the participants, the *Post-Test Questionnaire* is given to the participants.

3

C Data Logging Form

The moderator of the usability test observes the behavior of the participant while taking notes in the Data Logging Form. Each task needs one Data Logging Form.

Logging Form Sheet
CP.pt Usability Test
MAP-i Doctoral Programme
Responsible Moderator: Luís Cruz
22 Jannary 2014

Participant ID: _____
Date: _____
Time: _____

Task _____	Performance	Observations and Comments
Completeness	Correct / Incorrect	
Assists	*Tally*	
Negative Remarks	*Tally*	
Back Button Hits	*Tally*	
# of steps made differently	*Tally*	
# of Errors	*Tally*	

Task start trigger: _____ Task end: _____ Time to complete: _____

Based on the interview guided by the questions defined in the *Usability Test Plan* document, the following notes were taken:

_____
_____
_____
_____
_____
_____
_____
_____

10

D Post-Test Questionnaire

The moderator of the usability test observes the behavior of the participant while taking notes in the Data Logging Form. Each task needs one Data Logging Form.

Post-Test Questionnaire[1]

CP.pt Usability Test
MAP-i Doctoral Programme
Responsible Moderator: Luís Cruz
22 January 2014

Participant ID: _____
Date: _____
Time: _____

For our last activity I'm going to give you a short questionnaire that I would like you to fill out. The questionnaire will give you a series of statements about the CP.pt Web site. I would like you to rate your agreement with each statement.

For each statement, please circle a number to indicate the level to which you agree with each statement. If you feel a statement isn't relevant to your experiences with the Web site, please feel free to skip it.

1. The homepage is attractive — Disagree ○ ○ ○ ○ ○ Agree
2. The overall site is attractive — Disagree ○ ○ ○ ○ ○ Agree
3. The site's graphics are pleasing — Disagree ○ ○ ○ ○ ○ Agree
4. The site has a good balance of graphics versus text — Disagree ○ ○ ○ ○ ○ Agree
5. The colors used throughout the site are attractive — Disagree ○ ○ ○ ○ ○ Agree
6. The typography is attractive — Disagree ○ ○ ○ ○ ○ Agree

7. The homepage's content makes me want to explore the site further — Disagree ○ ○ ○ ○ ○ Agree
8. It is easy to find one's way around the site — Disagree ○ ○ ○ ○ ○ Agree
9. You can get information quickly — Disagree ○ ○ ○ ○ ○ Agree
10. It is fun to explore the site — Disagree ○ ○ ○ ○ ○ Agree
11. It is easy to remember where to find things — Disagree ○ ○ ○ ○ ○ Agree
12. Information is layered effectively on different screens — Disagree ○ ○ ○ ○ ○ Agree

13. The homepage is attention-getting — Disagree ○ ○ ○ ○ ○ Agree
14. Information is easy to read — Disagree ○ ○ ○ ○ ○ Agree
15. Information is written in a style that suits me — Disagree ○ ○ ○ ○ ○ Agree
16. Screens have the right amount of information — Disagree ○ ○ ○ ○ ○ Agree
17. The site effectively communicates the company's image — Disagree ○ ○ ○ ○ ○ Agree
18. Information is relevant — Disagree ○ ○ ○ ○ ○ Agree

[1] Adapted from *AIR, Human Factors and Research Design group's* work.

1

11

19. The site is designed with me in mind — Disagree ○ ○ ○ ○ ○ Agree
20. The site's content interests me — Disagree ○ ○ ○ ○ ○ Agree
21. The site's content would keep me coming back — Disagree ○ ○ ○ ○ ○ Agree
22. The site has characteristics that make it especially appealing — Disagree ○ ○ ○ ○ ○ Agree
23. The site reflects progressive, leading edge design — Disagree ○ ○ ○ ○ ○ Agree
24. The site is exciting — Disagree ○ ○ ○ ○ ○ Agree

25. The site is well-suited to first-time visitors — Disagree ○ ○ ○ ○ ○ Agree
26. The site is well-suited to repeat visitors — Disagree ○ ○ ○ ○ ○ Agree
27. The site has a clear purpose — Disagree ○ ○ ○ ○ ○ Agree
28. It is always clear what to do next — Disagree ○ ○ ○ ○ ○ Agree
29. It is clear how screen elements work — Disagree ○ ○ ○ ○ ○ Agree
30. Mistakes are easy to correct — Disagree ○ ○ ○ ○ ○ Agree

And that's it! Please return this questionnaire to the moderator.
Thank you for your collaboration!

2

12

References

Bangor, A., Kortum, P., and Miller, J. (2009). Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123.

Brooke, J. (1996). SUS - a quick and dirty usability scale. *Usability evaluation in industry*, 189:194.

Henry, S. L. and Thorp., J. (March). Wai site usability testing planning. Available at `http://http://www.w3.org/WAI/redesign/eval-sum.html`.

Prakash, S. (2013). Most common desktop/laptop and mobile screen resolution worldwide. available at `http://www.satya-weblog.com/2013/07/desktop-laptop-mobile-screen-resolution-most-common-worldwide.html`.

13

References

Bangor, A., Kortum, P., and Miller, J. (2009). Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123.

CP - Comboios de Portugal (2012). Relatório e contas consolidadas do grupo CP de 2012. Available at: `http://cp.pt/cp/displayPage.do?vgnextoid=360a4ff42133f310VgnVCM100000be01a8c0RCRD`. Accessed: January 4, 2014.

Mitchell, P. P. (2007). *A step-by-step guide to usability testing.* iUniverse, Inc.

Nielsen, J. (1992). Finding usability problems through heuristic evaluation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 373–380. ACM.

Nielsen, J. (1995a). 10 usability heuristics for user interface design. *Nielsen Norman Group, available at* *`http://www.nngroup.com/articles/ten-usability-heuristics/`*.

Nielsen, J. (1995b). Severity ratings for usability problems. *Nielsen Norman Group, available at* *`http://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/`*.

Nielsen, J. (2012). How many test users in a usability study. *Last modified June*, 4.

Nielsen, J. and Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 249–256. ACM.

Sauro, J. (2011). Measuring usability with the system usability scale (sus). Available at `http://www.measuringusability.com/sus.php`.

Spencer, R. (2000). The streamlined cognitive walkthrough method, working around social constraints encountered in a software development company. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 353–359. ACM.

Stanton, B. (2006). What is the cif? Industry USability Reporting, available at `http://zing.ncsl.nist.gov/iusr/documents/whatistheCIF.html`. Accessed: January 4, 2014.

Wharton, C., Rieman, J., Lewis, C., and Polson, P. (1994). The cognitive walkthrough method: A practitioner's guide. In *Usability inspection methods*, pages 105–140. John Wiley & Sons, Inc.