

# Testing Website Usability in Spanish-Speaking Academia through Heuristic Evaluation and Cognitive Walkthroughs

**María Paula González**

(National Council of Scientific and Technical Research CONICET, Argentina  
Computer Science and Eng. Department – Universidad Nacional del Sur, Argentina  
[mpg@cs.uns.edu.ar](mailto:mpg@cs.uns.edu.ar))

**Toni Granollers**

(Grup GRIHO- Universitat de Lleida, Spain  
[tonig@diei.udl.cat](mailto:tonig@diei.udl.cat))

**Afra Pascual**

(Grup GRIHO- Universitat de Lleida, Spain  
[afra@griho.net](mailto:afra@griho.net))

**Jesus Lorés** (in memoriam)

**Abstract:** Over the last few years, usability is a software attribute defined as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use”. Although usability has been assessed for different contexts of use, no comparative study about usability has been performed for academic websites in the Spanish-speaking context of use. Considering that this context involves hundreds of millions of potential users, the AIPO Association is running the UsabAIPO Project in order to provide a general diagnosis of this particular environment. The ultimate goal is to promote an adequate translation of international standards, methods and ideal values related to usability in order to adapt them to diverse Spanish-related contexts of use. This paper presents the main statistical results coming from the Second and Third Stages of the UsabAIPO Project, where an Heuristic Evaluation and seven Cognitive Walkthroughs were performed over 69 university websites. The planning and execution of the Heuristic Evaluation and the Cognitive Walkthroughs, the definition of two usability metrics, as well as the outline of the *UsabAIPO Heuristic Management System* prototype are also sketched.

**Keywords:** Usability Engineering, Usability Testing, Heuristic Evaluation, Cognitive Walkthrough, Spanish-speaking Context of Use.

**Categories:** H.3.1, H.3.2, H.3.3, H.3.7, H.5.1

## 1 Introduction

*Usability* is a software attribute defined as “*the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*” [ISO:98]. In this setting, a *context of use* is a description of the actual conditions under which the product (an interactive system) will be used in a normal situation, including cultural issues as those proposed in [Marcus:00, Shen:06]. Due to the growing relevance of usability as a key factor of

software quality [Hornbae:06, Ivory:01], *usability testing* has proven to be a crucial part of the Evaluation Stage under Usability Engineering (UE) [Dumas:00, Holzinger:05, Nie:93, Mayhew:99, Rosson:02]. Indeed, identifying the most general usability problems of a context of use can help not only to evaluate the interfaces belonging to it, but also to prevent usability errors when a novel interactive system in it is being developed. In the last years many efforts have been made towards testing usability of different context of use, specially those involving webpages. However, no comparative study about the usability of academic websites has been performed for the Spanish-speaking context of use,<sup>1</sup> (in what follows C) which involves hundreds of millions of potential users (persons who speak Spanish either natively or by adoption) whose cultural background is primarily associated with the Spanish language and culture, regardless of ethnic and geographical differences.

Given the above situation, the *Asociación Interacción Persona Ordenador* (AIPO)<sup>2</sup> is running the *UsabAIPO Project* (UP), a project centered on usability research which involves the participation of 15 multidisciplinary university research groups specialized in Human-Computer Interaction, which are formed not only by Spanish scientists but also by researchers coming from Spanish-speaking American countries. The UP was started in the second semester of 2004 (for a presentation of results corresponding to the First Stage see [Lores:05]) and it is still underway with the aim of analyzing the usability of official webs in C. Indeed, the UP ultimate goal is to promote an adequate translation of international standards, and ideal values related to usability in order to adapt them to C, aiming at a later extension of results to other contexts of use related to web design in Spanish language. Currently, within the UP the context C is represented by a sample group of systems formed by the web sites of the 69 universities listed in the *Universia* portal,<sup>3</sup> a widely used portal about universities available for Spanish-speaking countries. Besides, within all UP Stages four different categories were considered, namely *Design*, *Content*, *Navigation* and *Search*.

This article presents some statistical results associated with the Second and Third Stages of the UP, involving a Heuristic Evaluation (HE) [Nie:93] and seven Cognitive Walkthroughs (CWs) [Wharton:94] respectively. A sketch of some results presented here were previously published on [González et al., 2006]. Note also that for space reasons, the qualitative processing of usability results is not considered here, although we are aware of its relevance as a key factor under UE [Dumas:00, Ivory:01]. The qualitative treatment of the results coming from CWs is shown in [Gonzalez:07]. Additionally, an article analyzing the qualitative processing of results obtained through HE has also been submitted and is currently under evaluation.

This paper is structured as follows. First, next section discusses some related work describing usability testing of different contexts of use carried out during the last years. Then, the Second Stage of the UP is presented, sketching the planning and execution of the HE performed, as well as a discussion of some statistical results. A description of the metric *UsabAIPO-H* and an outline of the *UsabAIPO Heuristic Management System* prototype are also included. Next, the Third Stage of the UP is

---

<sup>1</sup> The Spanish-speaking context of use is formed by Spain and all Spanish-speaking countries in America, including almost 35 million people living in USA with Spanish-related cultural origins.

<sup>2</sup> See [www.aipo.es](http://www.aipo.es) (webpage of the Asociación Interacción Persona Ordenador AIPO).

<sup>3</sup> See [www.universia.es](http://www.universia.es)

presented by discussing seven CWs defined for three different user profiles. The metric *UsabAIPO-CW* is sketched, as well as some statistical results. Finally, the paper concludes and discusses future work.

## 2 Related Work

In the last years many efforts were made towards testing usability in different contexts of use, focusing on premises coming from cross-cultural usability [Marcus:00]. Indeed, the notion of context of use was defined to bring together people with similar social and cultural behavior, even considering educational level and unemployment rates [Blazic:07]. In this setting, although Spanish is one of the most widely spoken languages in the world, only in the last years the definition of context of use for Spanish-speaking scenarios has emerged as a crucial field. To the best of the UP partners' knowledge, there is no similar approach to particularizing the existing international usability standards with a focus on C as presented in this paper.

General recommendations on the basis of a particular cultural environment have been developed, especially those related to the question of which dimensions should be taken into account to localize a web page [Brandon:01, Bean:00]. An important reference can be found in [Nie:02], where ideal values for screen space distribution are proposed for Corporative Associations websites written in English. Results are based on an exhaustive study of the usability of 50 homepages belonging to the context of use considered. Another interesting results are shown in [Shen:06]. This paper discusses the redesign of a computer interface for a Chinese user target group, incorporating a consistent and culturally rooted metaphor (a Chinese garden metaphor). As part of the experimentation a desktop interface (which is normally based on an American--English metaphor) and a web browser design are presented and ideal values for the context of use of the Chinese garden are outlined.

A case involving Spanish-speaking culture is presented in [Sturm:02], where an interesting comparison between homepages of three universities (the *Hogeschool van Amsterdam* in The Netherlands, the *Universidad de la República* in Uruguay and the *Instituto Tecnológico de Monterrey* in Mexico) is discussed. The screenshots of the three universities are compared in order to visualize localization related to the cultural level of internationalization. However, only a high-level analysis is presented, and no discussion about the ideal values for the corresponding contexts of use (i.e. homepages in Dutch- and Spanish-speaking academia) is included.

## 3 Heuristic Evaluation under the UsabAIPO Project

Heuristic Evaluation (HE) is a usability evaluation method which consists in the “*systematic examination (inspection) of a finished product, design or prototype from the point of view of its usability by intended end users*” [Nie:94]. As explained in Section 1, the Second Stage of the UP consisted on the usability evaluation of 69 websites of the Universia portal (representing the context of use C) through an HE. The evaluation team was formed by three usability experts and two CS advanced students with solid knowledge about UE. All members were also frequent users of C.

<i>Cat</i>	<i>Heuristic</i>	<i>Related questions</i>
D	Graphic Design	<i>Friendly Interface</i> : Does the website have a friendly interface, with uniform colors in most pages, matching with the university corporate image?
		<i>Clean Interface</i> : Does the website offer a clean interface, without visual noise and with a correct usage of space?
		<i>Text Design</i> : Does the text have a simple design, with enough contrast between background and text, limiting the font type and other text att.?
		<i>Liquid Design</i> : Is a liquid design being used?
	Images	<i>Labelled Images</i> : Are images labelled? Does their title appear when the mouse is moved over them?
		<i>Animated elements</i> : do animated elements exist?
<i>Image Resolution</i> : Has image resolution been taken into account, so that images are not pixeled and have an appropriate size for their correct visualization?		
N	Navigation Area	<i>Number of elements in menus</i> : Has the number of elements and terms per element been checked in order not to produce memory overload?
		<i>Visibility</i> : Is the totality of the elements in the navigation area visible without requiring any interaction from the user?
		<i>Existence of site map</i> : Is there a site map on the website?
		<i>Affordance from links to applications</i> : If a link leads to an application, is that clearly indicated?
	Orientation	<i>Access to homepage</i> : can the homepage always be reached from any navigation level?
		<i>Orientation elements</i> : are there elements that allow the user to know exactly where he/she is in the website and how to move back? (breadcrumbs)
<i>Identifiable links</i> : do links clearly indicate where they lead to, using an appropriate title so that the user can predict the system response to his/her action?		
C	Information	<i>News update</i> : Are academic news properly updated? Do they have publication date?
		<i>Contact data</i> : Is it easy to access to the information in the different areas of the university (secretary, departments, etc.)? Is there clear information about the data needed to contact some particular area (phone, email, etc)?
		<i>News clarity</i> : Are news published on a salient place on the website, with a link to the news associated with the news headline, and with a clear abstract of the contents of the news?
	Internationalization	<i>Language</i> : Does the website offer a multilanguage option?
		<i>Scope</i> : The academic information available for the different languages is one page or is it the most part of the website?
B	Search Area	<i>Visibility and Simplicity</i> : Is it easy to start a search? Is the textbox used to search terms on the website to be found on the homepage? Is it easily accesible from any place of the website?
		<i>Size</i> : Is the text entry box for search wide enough? (it must contain between 15 and 30 visible characters)
		<i>Complexity</i> : Is there an advanced search option? Are there enough options to carry out a properly bounded search ?

		<i>Engines</i> : Are links to Internet search engines included?
	Search Result	<i>Comprehensibility</i> : Are search results shown in a clear form which is comprehensible enough for the user?
		<i>Assistance</i> : is the user assisted when no results can be found?

Table 1. Heuristic-related questions for the UsabAIPO Project (D, N, C and S corresponds to Design, Navigation, Content and Search Categories)

In order to specialize Heuristic Evaluation to evaluate C, two general heuristic principles for each of the UP categories were defined on the basis of 300 answers corresponding to a poll carried among users belonging to C and results coming from the UP First Stage [Lores:05] (in which the usability of the homepages in C were assessed by means of a HE specialized for homepages). Each general heuristic principle was decomposed in several heuristic-related questions, all of them related to features to be tested. As a result, twenty-five questions were defined, five corresponding to the *Content* Category, seven corresponding to the *Design* Category, seven corresponding to the *Navigation* Category and six corresponding to the *Search* Category (see Table 1). Ranges and ideal values were defined for each possible heuristic-related answer. Then a specialized software tool called *UsabAIPO Heuristic Management System* (UHMS) was developed to support the usability evaluation of C on the basis of the twenty-five heuristic-related questions that were defined (see screenshot in Figure 1). Indeed, the UHMS provides a simultaneous visualization of the current website that is being evaluated, the current heuristic-related question that needs to be answered (including possible answers), its significance within the website

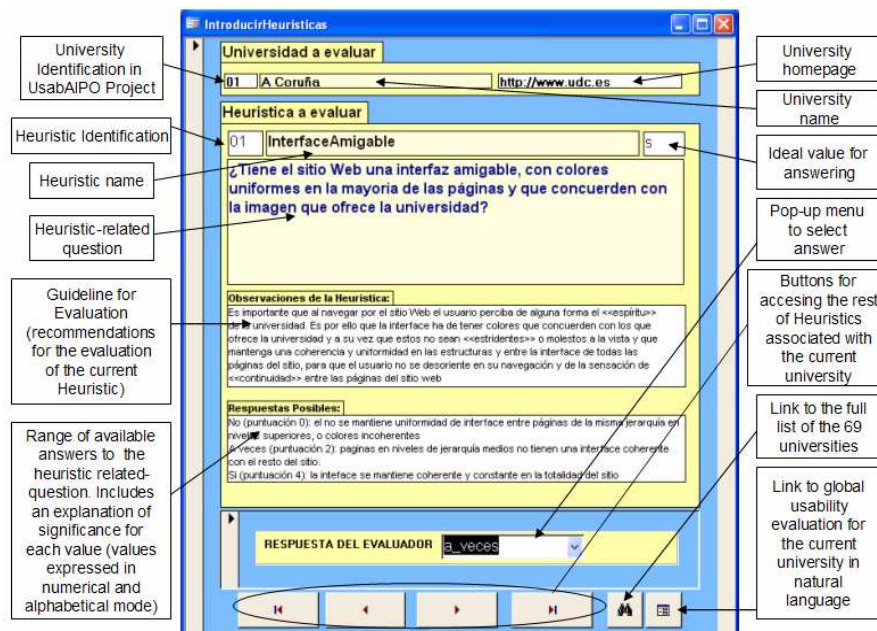


Figure 1. Interface of the UsabAIPO Heuristic Management System.

being considered and a pull-down menu to carry out the evaluation. To do this, every heuristic-related question was linked with an attribute in an output spreadsheet which stores the result of the evaluation. Besides, the tool provides a special form that allows the evaluator to add comments expressed in natural language, which will be automatically stored as part of the answer.

Besides, the UsabAIPO-H metric was defined to condense in a single value a general result for each system considered (each website belonging to C). Note that values coming for each heuristic-related question must be normalized in order to compute UsabAIPO-H, as not all possible heuristic-related answers are to be found between the same ranges. Moreover, each UP category was weighed according to its percentage of heuristic-related questions (the percentage of the total heuristic-related question performed, i.e. 25 questions), as follows:

$$UsabAIPO-H(w) = D*0,28 + N*0,28 + C*0,20 + S*0,24 \quad (1)$$

In this equation  $w$  is a website belonging to C,  $D$  represents the sum of the values obtained for the Design Category (28% of the total, i.e.  $7*100/25=28\%$ ),  $N$  represents the sum of the values obtained for the Navigation Category (28% of the total, i.e.  $7*100/25=28\%$ ),  $C$  represents the sum of the values obtained for the Content Category (20% of the total, i.e.  $5*100/25=20\%$ ), and  $B$  represents the sum of the values obtained for the Search Category (24% of the total, i.e.  $6*100/25=24\%$ ). Next, the evaluation team carried out the usability evaluation of the 69 websites considered by applying the specialized HE described before. The browsers MS Internet Explorer 6.0, Mozilla Firefox 5.0 and Netscape 8.0.1 were used to visualize the websites. The strategy used for each HE execution included a first overview of the website under evaluation about 10 minutes, followed by an in-depth evaluation carried out by means of the UHMS. As a result, 1,725 data records were obtained (25 heuristic-related questions applied to assess 69 websites). The information was stored in a spreadsheet called *Usab\_HE*, which was automatically generated by the UHMS. Then the UsabAIPO-H function was calculated on the basis of the data stored in *Usab\_HE* (see Figure 2). According to UsabAIPO-HE, only 2% of the websites in C presented a usability value above 70 (ideal usability value: 100). Most websites (54%) scored around 60 (between 51 and 70), while a 13% of them reached values between 26 and 50. No website in \cont was assigned a usability value below 26 or over 77.

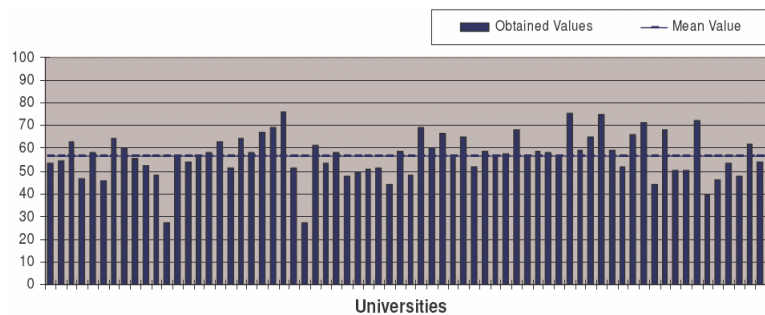


Figure 2. Results corresponding to the UsabAIPO-H calculation.

In order to study the UsabAIPO-H metric, a statistics analysis was performed over UsabAIPO-H using the *Minitab* Platform.<sup>4</sup> The study included calculation of central tendency measures (mode, mean, median and quartiles), measures of dispersion (standard deviation, variance, obliquity and kurtosis), and an univariate analysis (considering distribution, the symmetry and presence of atypical values). Figures 3 and 4 depict some results corresponding to the calculation of central tendency and dispersion measures. In addition, a multivariate analysis was included in order to study the relationship among the results corresponding to the different UP categories. In that respect, Figure 5 shows the general values obtained for each category (left) and particular values corresponding to the Content Category (right). Besides, the degree of independence of the categories according to the HE results was tested. In this case the statistical function  $\chi^2$  was used, calculated with 9 degrees of freedom (See Figure 6). Concerning the  $\chi^2$  function, note that the independence of the four categories was proven, as it holds that  $(p\text{-value}) < \alpha$  with  $\alpha = 0.05$ .

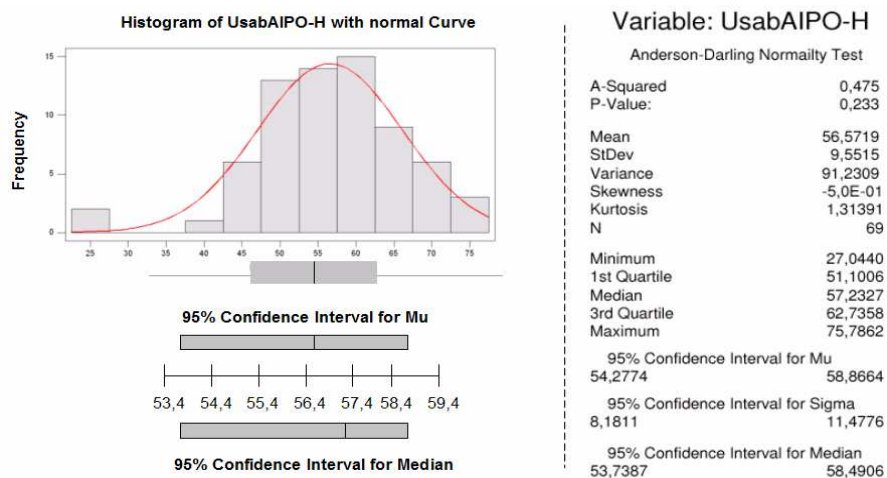


Figure 3. Statistical analysis of UsabAIPO-H. Left: Histogram and central tendency measures. Right: general results obtained.

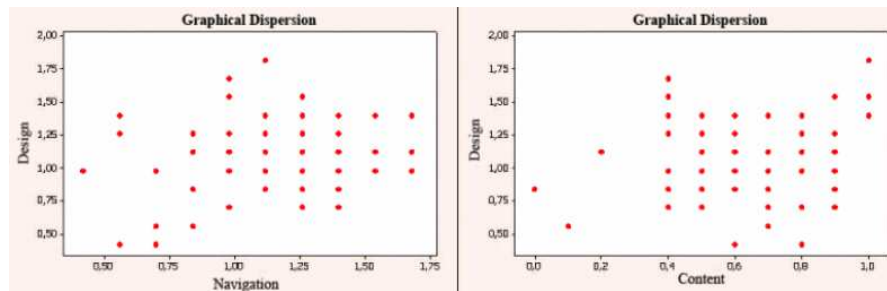


Figure 4. Statistical analysis of UsabAIPO-H. Some measures of dispersion.

<sup>4</sup> Licence No. 254052000082695, Universitat de Lleida.

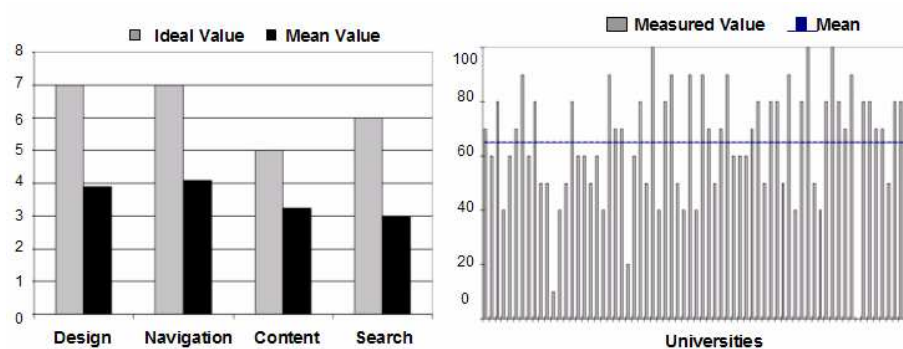


Figure 5. Multivariate Analysis of UsabAIPO-H. Left: general values for each Category. Right: particular values scored by the Content Category

### Test Chi Square

Expected counts are printed below observed counts

	Design	Navigation	Content	Search	Total
1	4 8,00	4 8,00	11 8,00	13 8,00	32
2	14 18,00	9 18,00	22 18,00	27 18,00	72
3	25 23,75	23 23,75	25 23,75	22 23,75	95
4	26 19,25	33 19,25	11 19,25	7 19,25	77
Total	69	69	69	69	276

$$\text{Chi-Sq} = 2,000 + 2,000 + 1,125 + 3,125 + 0,889 + 4,500 + 0,889 + 4,500 + 0,066 + 0,024 + 0,066 + 0,129 + 2,367 + 9,821 + 3,536 + 7,795 = 42,831$$

$$\text{DF} = 9, \text{ P-Value} = 0,000$$

Figure 6. Test  $\chi^2$  for statistical analysis of UsabAIPO-H.

## 4 Cognitive Walkthroughs under the UsabAIPO Project

Cognitive Walkthroughs (CWs) [Wharton:94] involve evaluators inspecting a user interface by going through a set of tasks, assessing its understandability and ease of learning. Evaluators construct tasks and then play the part of a user working with that interface, “walking through” it. During the Third Stage of the UP, seven CWs were performed to assess the usability of C. Again, the evaluation team was formed by three usability experts and two CS advanced students, all of them frequent users of C. Three user profiles were considered to adequate the definition of different CWs with respect to C: *student*, *professor* and *administrative* profiles. Besides, seven CWs



involving 37 tasks were defined on the basis of an exhaustive poll carried out during four months among 400 different users of C. Eight questions were linked to each task (two questions for each category in UP). The definition of the seven CWs and the tasks related to the CW #1 are:

- CW#1 (student profile): locate and visualize the study plan corresponding to a given undergraduate degree offered by the university under study./ CW#2 (student profile): locate and visualize information concerning academic regulations (enrollment information, etc.)/ CW#3 (administrative profile): locate and visualize information about a training course oriented towards the administrative staff and offered by the university/ CW#4 (administrative profile): locate and visualize data for contacting a person belonging to the University administration using the option of People Search/ CW#5 (professor profile): locate and visualize information about a posgraduate course or seminar offered by the university/ CW#6 (all profiles): locate and visualize email access facilities/ CW#7 (all profiles): locate and visualize a particular news made by the university.
- CW#1 (task description): *Task 1*: Visualization of types of degrees offered by the University (starting from the homepage); *Task 2*: Visualization of undergraduate degrees offered by the University (starting from the visualization of type of degree offered); *Task 3*: Visualization of study plan (starting from visualizing undergraduate degrees); *Task 4*: Short walkthrough using the study plan (courses, etc.).

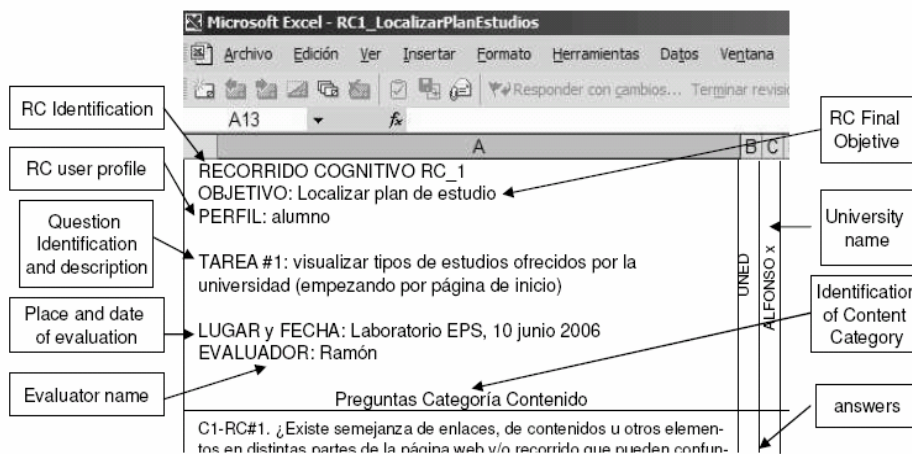


Figure 7. Zoom-in over the spreadsheet used to perform the Task 1 associated with CW#1.

Different ranked values for assessing users' cognitive effort were selected, namely (from bottom to top) *non\_mesurable*, *insignificant*, *low*, *normal*, *high* and *very\_high* (corresponding numerical values form 0 to 5). Besides, ideal values between *insignificant* and *normal* were related to every task. Then two auxiliary spreadsheets were designed to support the usability evaluation of the 69 selected websites on the basis of the 37 defined tasks. As an example, Figure 7 shows a zoom-in on the first questions of the Task 1 related to the CW#1. As in the UHMS, every task-related question was linked with an attribute in these spreadsheets. Besides, the *UsabAIPO-*

*CW* metric was defined to condense in a single value a general result for each website in *C*. Given a website *w*, the metric UsabAIPO-CW is simply the average of the seven values  $E_1 \dots E_7$ , each of them representing the cognitive effort  $E_i$  assessed for a given CW#*i* when evaluating *w*. The cognitive effort  $E_i$  is calculated based on the values  $T_1^i \dots T_{\text{LastTask}(i)}^i$ , where  $\text{LastTask}(i)$  is a function that returns the number of tasks related to CW#*i*, and the values  $T_n^i$  are obtained when measuring each task *n* (with  $1 \leq n \leq \text{LastTask}(i)$ ) during the performance of the CW#*i*. Note that each  $T_n^i$  is also weighted wrt a numeric value  $P_n^i$  to distinguish three different situations detected in the behavior of many users within *C*: at the beginning of the CW#*i*, the users' cognitive effort increases, which seems to be due to their lack of familiarity with *w*; in the intermediate tasks of CW#*i*, this effort decreases as the users get accustomed to the interface of *w*; and in the last tasks of CW#*i*, the users' cognitive effort increases again (in this case maybe due to their mental tiredness associated with performing the task). Thus, the UsabAIPO-CW is defined as follows:

$$\begin{aligned} \text{UsabAIPO-CW}(w) &= (\sum_{i=1}^7 E_i(w))/7 \\ \text{with } E_i(w) &= T_1^i * P_1^i + \dots + T_{\text{LastTask}(i)}^i * P_{\text{LastTask}(i)}^i \end{aligned} \quad (2)$$

where  $P_1^i > P_j^i < P_{\text{LastTask}(i)}^i$  for  $j \leq \text{LastTask}(i)/2$ , and  $(\sum_{j=1}^{\text{LastTask}(i)} P_j^i) = 1$ . Next, the evaluation team was divided in two sub-teams in order to carry out two independent usability evaluations of the 69 websites considered by applying the redefined CW method described before. Due to the number of measures to be assessed, the execution of the seven CWs took about seven months. Again, the browsers MS Internet Explorer 6.0, Mozilla Firefox 5.0 and Netscape 8.0.1 were used to visualize the websites. The strategy used for carrying out each CW consisted in a first overview of the website under evaluation for about 10 minutes, followed by the performing of the CW itself. As a result, each sub-team collected 10,212 data records that were stored in temporary spreadsheets.

After processing all this data, each evaluation sub-team controlled the results produced by the other sub-team, and 24 data records were rejected as they had missing values. These missing values were related to unfinished CWs, which could not be completely assessed due to circumstances out of the scope of the evaluators (e.g. unknown password, nonexistent link, etc.). Afterwards both spreadsheets were condensed in a final spreadsheet called *Usab\_RC*, containing 20,400 data records. The data were normalized in a scale from 0 to 100 to enhance the comparison of results corresponding to the Second and the Third Stage of the UP. Following this idea, also the function  $E_i$  had to be normalized by changing the condition  $(\sum_{j=1}^{\text{LastTask}(i)} P_j^i) = 1$  for  $(\sum_{j=1}^{\text{LastTask}(i)} P_j^i) = 100$ . Then, the UsabAIPO-CW function was calculated on the basis of the normalized data stored in *Usab\_RC* (see Figure 8). For the CWs involving three tasks, values to weigh the corresponding  $E_i$  function were selected as follows: 0.4 for weighting the first and third tasks, and 0.2 for weighting the second task. With respect to the CWs including four tasks, values to weigh the associated  $E_i$  function were chosen as follows: 0.375 to weighing the first and last tasks, and 0.125 for weighting the two intermediate tasks.

According to UsabAIPO-CW, only one university in *C* presented a usability value above 70. Besides, 42% of the websites in *C* scored between 60 and 70, while 50.56% of them reached values between 50 and 60. Besides, 6% of the cases had a value

lower than 50; and no website was assigned a usability value below 35 or over 71. Note that for the UP Project, the ideal usability value associated with CWs after normalizing the data stored in Usab\_RC should be between 30 and 40, as each of the seven CWs involved three or four tasks, each task ideally associated with a minimal cognitive effort (i.e. all  $T_n^i=1$  for  $1 \leq i \leq 7$  and  $1 \leq n \leq \text{LastTask}(i)$ , representing that all data related to the seven CWs were assessed with value=*insignificant*).

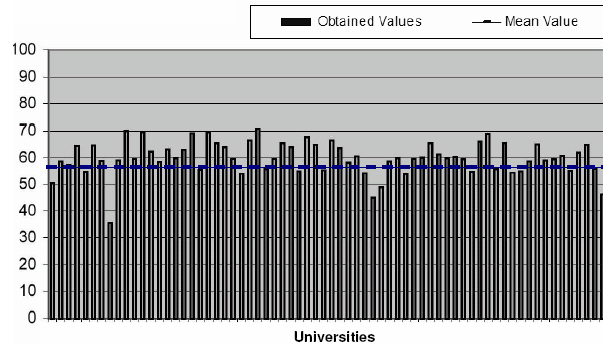


Figure 8. Results corresponding to the  $\backslash$ metricarec calculation.

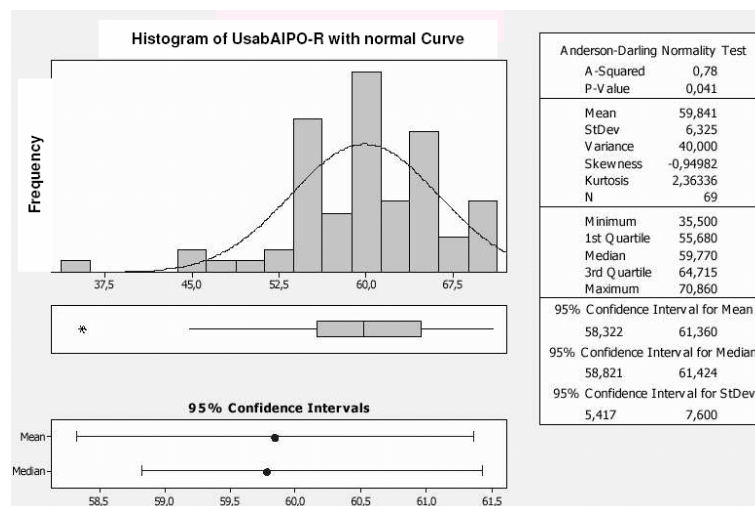


Figure 9. Statistical analysis of UsabAIPO-CW. Left: Histogram and central tendency measures. Right: general results obtained.

In order to study the above results, a statistics analysis was carried out over UsabAIPO-CW. This analysis included the same statistical study as the one calculated during the Second Stage of UP. As before, the *Minitab* Platform was used. Figure 9 depicts relevant results corresponding to the calculation of central tendency measures, including some numerical results associated with the measures of dispersion. Concerning the multivariate analysis of the obtained results, five functions were calculated for each of the seven CWs: a function to show general values obtained for

each category, and four functions associated with values scored by each particular UP category (one category per function). As an example, Figure 10 shows general values obtained for each category for the CW#1 (left), and particular values scored by the Content Category for the same CW#1 (right).

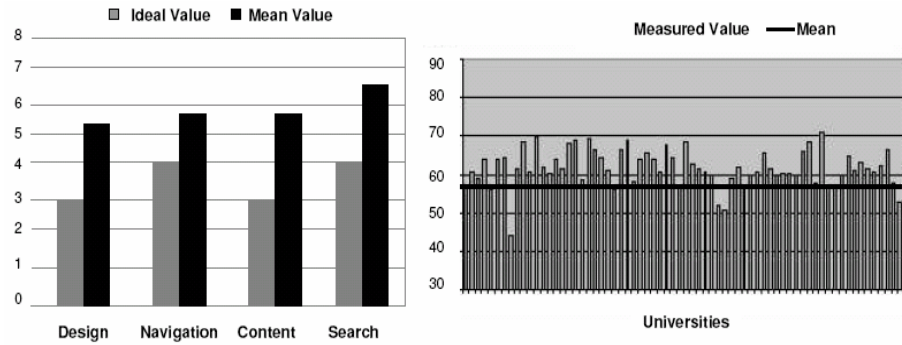


Figure 10. Multivariate Analysis of CW#1. Left: general values for each Category. Right: particular values scored by the Content Category.

**Chi-Square Test**

Expected counts are printed below observed counts  
Chi-Square contributions are printed below expected counts

	Content	Design	Nav.	Search	Total
1	14 19,00 1,316	25 19,00 1,895	10 19,00 4,263	27 19,00 3,368	76
2	14 18,00 0,889	22 18,00 0,889	19 18,00 0,056	17 18,00 0,056	72
3	28 19,50 3,705	15 19,50 1,038	28 19,50 3,705	7 19,50 8,013	78
4	13 12,50 0,020	7 12,50 2,420	12 12,50 0,020	18 12,50 2,420	50
<b>Total</b>	<b>69</b>	<b>69</b>	<b>69</b>	<b>69</b>	<b>276</b>

Chi-Sq = 34,073; DF = 9; P-Value = 0,000

Figure 11. Test  $\chi^2$  for statistical analysis of UsabAIPO-H.

Moreover, Figure 11 shows results corresponding to the calculation of the  $\chi^2$  function with 9 degrees of freedom. Again, note that the independence of the four categories was proven for UsabAIPO-CW, as it holds that  $(p\text{-value}) < \alpha$  with  $\alpha=0.05$ .

## 5 Conclusion and Future Work

Usability is a software attribute related to the easy of use and learning of a given interactive system at any development stage (from a prototype up to a fully developed one). Measuring usability of a given context of use as a whole is a challenging problem nowadays, and providing an appropriate solution for such a problem can be extremely helpful for software development and Usability Engineering in several respects. On the one hand, usability evaluators can rely on the usability features detected for a particular context of use as a whole when assessing a new interface belonging to it. On the other hand, the contextual knowledge about usability features of these context of use can help software developers to prevent possible usability problems when novel interactive system belonging to it is under development. Indeed, different contexts of use have been evaluated as a whole by specializing general usability methods, adapting international standards and ideal values to each particular environment under consideration. However, although the academic Spanish-speaking community includes hundreds of millions of potential users, to the best of UP partners' knowledge no comparative study about usability has been performed for academical websites in this context of use.

In this paper we have presented relevant quantitative results coming from an Heuristic Evaluation (HE) and seven Cognitive Walkthroughs (CWs) that were performed to assess the usability of C, a context of use formed by academic websites written in Spanish language and represented by the 69 websites of the Universia Portal. The HE and the CWs were performed as part of the 2<sup>nd</sup> and 3<sup>rd</sup> Stages of the UsabAIPO Project (UP), a project still underway on which four categories (namely *Design*, *Content*, *Navigation* and *Search*) are considered. Indeed, within the UP the usability of official webs in C is currently being studied with the final goal of promoting an adequate translation of international standards and ideal values related to usability to adapt them to this particular context of use.

Concerning the HE, we started by defining twenty-five usability-related questions to focus on C, linking each question with one category in UP. Appropriate ranges and ideal values were established for possible answers. The specialized tool *UsabAIPO Heuristic Management System* was created to facilitate the HE execution, and the *UsabAIPO-H* metric was defined to condense in a single value the results assessed for each website in C. Then, the specialized HE was performed and 1,725 data records were collected and the *UsabAIPO-H* metric was calculated. A statistical study was performed over *UsabAIPO-H*, including the calculation of central tendency and dispersion measures, as well as a  $\chi^2$  test performed over *UsabAIPO-H* to prove the independence of the four categories considered. In the same way, 37 different task were defined to specialize the seven CWs. As before, usability related questions were associated with the UP categories, and an adequate range and ideal values were defined for each possible answer. Besides, the *UsabAIPO-CW* metric was outlined to summarize the results assessed for each CW, considering a final score for each evaluated website. Note that *UsabAIPO-CW* distinguished three situations related to different cognitive efforts that were detected in the behavior of many users within C. After performing the seven CWs, the assessed values were condensed in a final spreadsheet called *Usab\_CW*, containing 20,400 data records. Based on them, the *UsabAIPO-CW* metric was calculated. To study the obtained *UsabAIPO-CW*, a

similar statistical study as the one performed during the 2<sup>nd</sup> Stage of UP was carried out. As before, the independence of the UP categories was proven. Results coming from both 2<sup>nd</sup> and 3<sup>rd</sup> Stages of the UP showed that, according with the HE and the CWs performed, webpages belonging to C suffer from some lack of usability, especially when a task must be performed. However, more testing is needed to verify not only the obtained results, but also the correctness of the proposed specialized methods, metrics and ideal values.

Indeed, part of our future work is focused on validating the performance of the defined HE and CWs, as well as the ideal values and usability-related questions here proposed. To achieve this, the 4<sup>th</sup> UP Stage is under consideration, involving the evaluation of academic websites of Spanish-speaking universities outside Spain under the same conditions established during the 2<sup>nd</sup> and 3<sup>rd</sup> Stages of UP. Another research line currently explored is the deployment of a novel version of the *UsabAIPO Heuristic Management System* to be integrated in an *Usability Evaluation Management System*. Taking as starting point the existing interface, our goal is to develop an HE management tool oriented towards Spanish language which will include UP features by default and will allow to easily redefine particular sets of heuristics, HE-related questions and answers, as well as range for answering and alternative ideal values.

**Acknowledgments.** This work was partially supported by CONICET, and Projects TIN2004-08000-C03-03 and SGR-00881. We acknowledge the collaboration of the different UP partners, as well as polled users. Particularly, we acknowledge Ramon Gabandé for his participation.

## References

- [ISO, 1998] (1998). ISO 9241-11. Ergonomic Req. Part 11: Guidance on Usability.
- [Bean, 2000] Bean, J. (2000). A framework for Globalization. Enterprise Development.
- [Blažič et al., 2007] Blažič, B., Law, E., Arh, T. (2007). An assessment of the usability of an Internet-based education system in a cross-cultural environment: The case of the Interreg crossborder program in Central Europe. *J. Am. Soc. Inf. Sci. Technol.*, 58(1):66–75.
- [Brandon, 2001] Brandon, D. (2001). Localization of web content. *J. Comput. Small Coll.*, 17(2):345–358.
- [Dumas and Redish, 2000] Dumas, J. , Redish, J. C. (2000). A Practical Guide to Usability Testing. Intl. Specialized Book Service Inc.
- [González et al., 2007] González, M.P., Lorés, J., Granollers, T. (2007). Assessing usability problems in latin-american academic webpages with cognitive walkthroughs and datamining techniques. In *Lecture Notes in Computer Science Springer Series. Proc. of the Int HCI'2007* (in press). Springer Verlag.
- [González et al., 2006] González, M. P., Lorés, J. Pascual Almenara, A., Granollers, T. (2006). Evaluación Heurística de Sitios Web Académicos Latinoamericanos dentro de la Iniciativa UsabAIPO. *Proc. VII Int. Cong. Interacción Persona-Ordenador*, ISBN 84-690-1613, pp. 145-157.

- [Holzinger, 2005] Holzinger, A. (2005). Usability engineering for software developers. *Communications of the ACM*, 48(1):71–74.
- [Hornbæk, 2006] Hornbæk, K. (2006). Current practice in measuring usability: challenges to usability studies and research. *Int. Journal on Human Computer Studies*, 64(2):79–102.
- [Ivory and Hearst, 2001] Ivory, M. Y., Hearst, M. A. (2001). The state of the art in automating usability evaluation of user interfaces. *ACM Comput. Surv.*, 33:470–516.
- [Lorés et al., 2005] Lorés, J., González, M. P., Pascual Almenara, A. (2005). First stage of the UsabAIPO project. *Proc. VI Cong. Interacción Persona-Ordenador INTERACCION'05* (available in Spanish), pages 217–221.
- [Marcus and West, 2000] Marcus, A., West, E. (2000). Crosscurrents: Cultural Dimensions and Global Web-User Interface Design. *Interactions*, VII(4):32–46.
- [Mayhew, 1999] Mayhew, D. J. (1999). *The Usability Engineering Lifecycle. A practitioner's handbook for user interface design*. M. Kaufmann.
- [Nielsen, 1993] Nielsen, J. (1993). *Usability Engineering*. M. Kaufman.
- [Nielsen and Mack, 1994] Nielsen, J., Mack, R. (1994). *Usability Inspection Methods*. John Wiley & Sons.
- [Nielsen and Tahir, 2002] Nielsen, J., Tahir, M. (2002). *Homepage Usability: 50 Websites Deconstructed*. New Riders Publishing.
- [Rosson and Carroll, 2002] Rosson, M., Carroll, J. (2002). *Usability Engineering: Scenario-based development of HCI*. M. Kaufmann.
- [Shen et al., 2006] Shen, S., Woolley, M., Prior, S. (2006). Towards Culture-Centred Design. *Interacting with Computers* (in press), 19(2):1–33.
- [Sturm, 2002] Sturm, C. (2002). TLCC-Towards a framework for systematic and successful product internationalization. *Proc. Int. IWISP'02*.
- [Wharton et al., 1994] Wharton, C., Rieman, J., Lewis, C., Polson, P. (1994). *Usability Inspection Methods*, chapter *The Cognitive Walkthrough Method: A Practitioners Guide*, page 1994. John Wiley & Sons Inc.