Usability Evaluation with Heuristics, Beyond Nielsen's List

Toni Granollers
Human-Computer Interaction and Data Integration Research Group (GRIHO)
Polytechnic Institute of Research and Innovation in Sustainability (INSPIRES)
University of Lleida
Lleida, Spain
email: antoni.granollers@udl.cat

***Abstract*— The work presented here is born of our extensive experience evaluating the usability of user interfaces and observing that some traditional methods need to be updated and improved. Here, we put the focus on the Heuristic Evaluation (HE) technique. It is one of the important topics in Human-Computer Interaction (HCI) when talking about usability evaluation. Different research works have been discussing the effectiveness of the current HE, but it is important to improve its effectiveness. A substantial improvement is presented, consisting of: (i) a new list of principles for evaluating any interface, (ii) a set of specific questions to be answered when analysing every principle, (iii) an easy rating scale for each question, and finally, (iv) a method to obtain a quantitative value, called the Usability Percentage. It gives a numeric idea about how usable the evaluated interface is. An experiment by a group of experts helped to validate the implications of the proposed solution.**

Keywords- usability; evaluation; heuristics; principles; user interface; human-computer interaction.

I. INTRODUCTION

Heuristic Evaluation (HE) is one of the most widely used methods for evaluating the usability of an interactive system. This technique became the most popular user-centred design approach in the 1990s but has become less prominent with the move away from desktop applications [3]. It is an effective method to assess user interfaces (UI) by taking the recommendations based on User Centred Design (UCD) principles. These recommendations come in terms of design principles, heuristics, guidelines, best practices or user interface design patterns and standards [1] that are supposed to serve interface designers and evaluators.

Nevertheless, this technique has some aspects that, if improved, will increase its efficiency. For example, having to be concerned with the need for adapting the heuristic set to the specific features of each interactive system represents a drawback. Thus, evaluators must combine different recommendation sources to review the specific application domain. This involves either choosing J. Nielsen's list [10] (the most well-known and widely used), or a long reviewing process of guideline collections that inevitably causes conflicts between various resources [6]. In addition, each set uses different scoring systems to score the recommendations, so it is necessary to adjust it in the resultant set.

The process of combining different heuristic sets usually finishes up with an extensive list of duplicated recommendations, similar statements using different terms and potentially conflicting guidelines. A clean-up and selection process is then required to provide a reliable, consistent and easy to use heuristic set [1].

Furthermore, our experience reveals that, when someone uses Nielsen's list [10], it is often due to a lack of deep knowledge about it and, possibly, about the technique itself. This assertion is based on Nielsen's statement: "*these principles are broad rules of thumb and not specific usability guidelines*" [10], making the list itself impossible to be used to evaluate.

Another important aspect to be improved resides in the qualification method of the principles, providing very subjective and sometimes confusing final reports. HE combines the qualitative and quantitative results [12], but the qualitative results usually take on greater relevance. Attempts to improve quantitative outputs have been done in [1][2][9][14][20], among others. Other more ambitious works, such as work done by Masip et al. [6][7], offered a full framework that enables a semi-automatic process to provide the most adequate set of heuristics for each specific situation. It also classifies the heuristics (defining a User eXperience degree, UX-degree) in terms of different constraints and enables an automatic classification of the problems found (removing the post-evaluation meeting held by the evaluators) for a better, fuller process. Nevertheless, the rigorousness of this process makes it so complex that it is not widely used. Therefore, in this paper, we propose a substantial improvement in evaluating the usability of user interfaces, in the form of a new set of heuristics with a new evaluation method.

The article is organized as follows: Section 2 explains the sources of information consulted and the revision methodology. In Section 3, the new set of heuristics proposal is explained as well the new methodology for evaluating using it. Section 4 explains the experimentation carried out for validating the proposals. Section 5 presents our discussion and Section 6 concludes the article and provides future work plans.

II. COMBINING COMMON HEURISTIC SETS

A. Sources consulted

Since B. Schneiderman, in 1987, established his well-known *Eight Golden Rules of Interface Design* [18], and going via the no less well known Nielsen's *Ten general principles for interaction design* [10] or Tognazzini's *First Principles of Interaction Design* [21], several authors have designed new sets (usually modifying Nielsen's list and/or adding new principles to evaluate specific aspects not

covered) to help user interface designers and/or experts in their goal of enhancing usability. A complete review of several sets of usability heuristics created for specific domains by different authors can be found in [14] (specifically in the appendix A of this reference).

Nevertheless, the reality (mainly in private companies) is that when almost everybody refers to evaluate usability or UX with heuristics they refer to Nielsen's list.

In our case, after more than twenty years of experience involved in evaluation of interfaces, we decided to take Nielsen's and Tognazzini's lists to do this present work. We selected only these two because they are well-known references because its unquestionable quality. So, there is no need to spend much time refining other sets or providing specific new ones. Moreover, as we said before (and confirmed by our own experience), Nielsen's list needed something more specific to be useful. Hence, we decided to complete it by combining both.

B. Methodology

Inspired by the recommendations found in Rusu et al. [14] and Quiñones's [13] works, the process followed for deciding our list of proposal principles was as follows:

1) **Revision** of the two chosen lists. The first step is to read carefully all the principles of Nielsen's and Tognazzini's lists. The revision has been done in terms of understanding the deep meaning of each principle.
2) **Compare** similarities. As the intention is to get a compact and complete solution for evaluating all kinds of user interfaces, we compared all the similarities in order to merge principles as much as possible.
3) **Integrate** these similarities. Previous comparison showed that some principles from both lists are identical while others are technically the same. The reader can identify these cases in marked cells (with the symbol "⇔") in the 2nd column in TABLE I.
   Also, in this step we identified some cases from Tognazzini's list that can also be joined. The intention is not to question his proposal, only to have the shortest list possible without losing its efficiency. These cases are marked with the symbol "+" in the 4th column in TABLE I. and the following paragraphs explain the reasoning:

- "Use of Metaphors" and "Human Interface Objects" can be considered as a single principle because both have the objective of creating a mental connection between the user and an object. The user uses this object in his/her daily life that is associated with a functionality of the system.
- "Learnability" and "Anticipation". In Tognazzini's own words, anticipation tries to "bring to the user all the information and tools needed for each step of the process", a characteristic that is directly related with the learning curve, which defines learnability.
  Additionally, both concepts can be understood with the capacity for minimizing the user's memory load by making objects, actions, and options visible. Hence, our decision of proposing the 5th principle "Recognition rather than memory, learning and anticipation".
- Protecting the work and saving the user's status can be grouped together as the purpose is for the user to continue working from where he was previously without losing his job. Regardless of whether there has been an unexpected failure of the system or the user who has closed the application. The main difference lies in how to prevent these errors, depending on whether an application is installed on the device or is consulted online. The 11th principle "Save the state and protect the work" serves to group these concepts together.
- Finally, readability and colour principles have been grouped together because both deal with design features to be easily seen and understood by every user. In fact, readability includes not only colour but fonts, typography or the text contrast with the background.

As a result, a final list with 15 general principles (or heuristics) resulting from mixing Nielsen's and Tognazzini's lists was created (see the far-right column of TABLE I.). All this part is explained and expanded in [4].

TABLE I. RESULTING LIST OF PRINCIPLES. THE READER CAN OBSERVE THE CORRESPONDING PREDECESSORS.

Nielsen		Tognazzini			**Resulting Principles**
Visibility of system status	⇔	Visible Navigation	+	Discoverability	**1.- Visibility and system state**
Match between system and the real world	⇔	Human Interface Objects	+	Metaphors, Use of	**2.- Connection between the system and the real world, metaphor usage and human objects**
User control and freedom	⇔	Explorable Interfaces			**3.- User control and freedom**
Consistency and standards	⇔	Consistency			**4.- Consistency and standards**
Recognition rather than recall	⇔	Anticipation	+	Learnability	**5.- Recognition rather than memory, learning and anticipation**
Flexibility and efficiency of use	⇔	Efficiency of the User	+	Efficiency of the User	**6.- Flexibility and efficiency of use**
Help users recognize, diagnose, and recover from errors					**7.- Help users recognize, diagnose and recover from errors**
Error prevention					**8.- Preventing errors**
Aesthetic and minimalist design	⇔	Aesthetics	=	Simplicity	**9.- Aesthetic and minimalist design**
Help and documentation					**10.- Help and documentation**
		Protect Users' Work	+	State	**11.- Save the state and protect the work**
		Colour	+	Readability	**12.- Colour and readability**
		Autonomy			**13.- Autonomy**
		Defaults			**14.- Defaults**
		Latency Reduction			**15.- Latency reduction**

III. NEW SET AND EVALUATION PROPOSAL

A. Resulting set

As we have seen in the introductory Section, to effectively use them for UI evaluations we need "something more specific". This is also what every experienced usability/UX evaluator agrees. This "something more specific" are precise questions to be answered by the evaluators to assess every principle. Then, beyond the list of principles, a set of questions to evaluate every heuristic was created. In this case, 60 questions cover all the principles (TABLE II. shows all of them: principles and their corresponding questions).

At this point it is important to know two aspects: first, the list of questions arises directly from the analysis of our referents, mainly Tognazzini's, which is more expressive in this aspect. And, second, the final list presented here is not the initial one. As it is presented later in Section IV, the experiment also served for testing the principles and their corresponding questions. The final list has the enhancements provided by the evaluator's feedback

TABLE II. HEURISTIC LIST PROPOSED WITH THEIR CORRESPONDING EVALUATION QUESTIONS

1.- Visibility and system state
- Does the application include a visible title page, section or site?
- Does the user always know where it is located?
- Does the user always know what the system or application is doing?
- Are the links clearly defined?
- Can all actions be visualized directly? (No other actions are required)
2.- Connection between the system and the real world, metaphor usage and human objects
- Does information appear in a logical order for the user?
- Does the design of the icons correspond to everyday objects?
- Does every icon do the action that you expect?
- Does the system use phrases and concepts familiar to the user?
3.- User control and freedom
- Is there a link to come back to initial state or homepage?
- Are the functions "undo" and "re-do" implemented?
- Is it easy to come back to an earlier state of the application?
4.- Consistency and standards
- Do link labels have the same names as their destinations?
- Do the same actions always have the same results?
- Do the icons have the same meaning everywhere?
- Is the information displayed consistently on every page?
- Are the colours of the links standard? If not, are they suitable for its use?
- Do navigation elements follow the standards? (Buttons, check box, ...)
5.- Recognition rather than memory, learning and anticipation
- Is it easy to use the system for the first time?
- Is it easy to locate information that has already been searched for before?
- Can you use the system at all times without remembering previous screens?
- Is all content needed for navigation or task found in the "current screen"?
- Is the information organized according to logic familiar to the end user?
6.- Flexibility and efficiency of use
- Are there keyboard shortcuts for common actions?
- If there are, is it clear how to use them?
- Is it possible to easily perform an action done earlier?
- Does the design adapt to the changes of screen resolution?
- Is the use of accelerators visible to the normal user?
- Does it always keep the user busy? (without unnecessary delays)
7.- Help users recognize, diagnose and recover from errors
- Does it display a message before taking irreversible actions?
- Are errors shown in real time?
- Is the error message that appears easily interpretable?
- Is some code also used to reference the error?
8.- Preventing errors
- Does a confirmation message appear before taking the action?
- Is it clear what information needs to be entered in each box on a form?
- Does the search engine tolerate typos and spelling errors?
9.- Aesthetic and minimalist design
- Is used a design without redundancy of information?
- Is the information short, concise and accurate?
- Is each item of information different from the rest and not confused?
- Is the text well organized, with short sentences and quick to interpret?
10.- Help and documentation
- Is there the "help" option?
- If so, is it visible and easy to access?
- Is the help section aimed at solving problems?
- Is there a section of frequently asked questions (FAQ)?
- Is the help documentation clear, with examples?
11.- Save the state and protect the work
- Can users continue from a previous state (where they had previously been or from another device)?
- Is "Autosave" implemented?
- Does the system have a good response to external failures? (Power cut, internet not working, ...)
12.- Colour and readability
- Do the fonts have an adequate size?
- Do the fonts use colours with sufficient contrast with the background?
- Do background images or patterns allow the content to be read?
- Does it consider people with reduced vision?
13.- Autonomy
- Does it keep the user informed of system status?
- Moreover, is the system status visible and updated?
- Can the user take their own decisions? (Personalization)
14.- Defaults
- Does the system or device give the option to return to factory settings?
- If so, does it clearly indicate the consequences of the action?
- Is the term "Default" used?
15.- Latency reduction
- Is the execution of heavy work transparent to the user?
- While running heavy tasks, is remaining time or some animation shown?

B. Evaluation method

Formal, the HE method where evaluators rate every question to deliver a final report. The intention is to combine the qualitative and quantitative findings, but usually the reports are mainly qualitative. So, some efforts have been made to improve the qualitative answer to the client.

For example, Nielsen [11] proposed a combination of three factors (frequency, impact and persistence) each of which must be rated on a scale of 5 values to assess the market impact of the problem according to its severity. Or, Rubin and Chisnell [10] whose problem severity scale in a single parameter had 4 possible values (unusable, severe, moderate and irritant).

The previous examples, together with others that are not mentioned and our own experience, support the idea that existing quantitative proposals are not good enough. Nielsen himself declares that his proposal is a good measure but "*It is difficult to get good severity estimates from the evaluators during a heuristic evaluation session*" [11]. An idea also shared by UX professionals and researchers such as J. Sauro who, in [18], argues that it is usually difficult to distinguish easily between too many rating levels.

In this research, we propose a simple but effective rating method. Our proposal has three main characteristics:

- A **4-option rating scale**: "Yes", "Neither", "No", and "Not applicable"

  Our experience demonstrates that evaluators find it exhausting when they must choose from a long list of values. This feature worsens as the evaluation time progresses, giving unreliable answers as the evaluators reach the last principles. We believe that by reducing the answers, as is proposed, this characteristic is minimized.

  Obviously, for each question, the evaluator can write as many qualitative comments as needed. They will reinforce the chosen selection and will provide hints to argue with other evaluators and for the programmers responsible to solve them.
- **Questions for the principles are interrogative questions written in a way that is favourable to usability**

  Literature and examples present design principles written in several forms: affirmative, negative, a mix of both or with and without question form, making (sometimes) confusion in the evaluator's work. A study presented by Masip et al. [7] revealed that evaluators prefer interrogative sentences because they are more intuitive, direct and easy for the evaluation job. Adopting this recommendation, all the principles in this proposal are written in question form.

  Moreover, all these questions are formulated in the way that when the answer is "Yes" it means that this feature represents good usability. Consequently, "No" represents the opposite, and "Neither" represents some value in between. The last possible answer, "Not applicable" is needed for those cases where this question has no meaning to the case.

  In fact, this is because we "*do not have to obsess too much over whether higher severity problems should have higher numbers or lower ones. It's the order that has meaning*" [18], what is really important is to find the mistakes.
- A final usability value, called "**Usability percentage**", UP.

  Apart from the qualitative insights, as we have seen in the introduction, some works [1][2][9][14][20] have been looking to complement them with a number to quantify how usable the evaluated interface is. This responds to one characteristic of human nature, always looking to quantify everything.

  In our case, it is important to know that there is no intention of giving a deterministic value with a hard meaning. Our aim is to give a kind of orientation about the level of usability of this interface.

  For this purpose, we will make use of the evaluators' answers to provide a final number. This is too simple, and we assign a numeric value to each possible answer. Thus, "Yes" is weighted with 1 point, "Neither" with 0.5 points and "No" with 0 points. "Not applicable" answers will not be taken into consideration, as if they do not exist.

  TABLE III. shows an example corresponding with the evaluation of the "Consistency and standards" principle. Here we can observe how the values are taken in each case.

TABLE III. PUNCTUATION EXAMPLE WITH THE PRESENTED PROPOSAL

4 - Consistency and standards		
Do link labels have the same names as their destinations?	Neither	0.5
Do the same actions always have the same results?	Yes	1
Do the icons have the same meaning everywhere?	Not applicable	-
Is the information displayed consistently on every page?	No	0
Are the colours of the links standard? If not, are they suitable for its use?	Yes	1
Do navigation elements follow the standards? (Buttons, check box, ..)	Yes	1

Finally, we add up all the values to obtain the final number.

With this, the maximum value is 60, corresponding with the case that all the questions have been answered with a "Yes". Nevertheless, as the "Not applicable" answer makes it impossible to know the maximum number for every evaluation, we translate this value to a percentage value, taking into consideration only the "Yes", "Neither" and "No" answers. It provides a number (a percentage) that enables the comparison among the assessments done by the other evaluators.

The value for the case of TABLE III. is 3.5 that will be added to the final evaluation number. Bear in mind that, when translating to the percentage value, only 5 of the 6 questions will be taken into account (the 3rd answer is "Not applicable").

IV. EXPERIMENTATION

To validate our proposal, we designed an experiment that has a multipurpose goal. On the one hand, we want to obtain feedback from experienced evaluators to validate the heuristic list, the corresponding questions (see TABLE I), the rating scale and the Usability Percentage (UP) final value. Secondly, we also wanted feedback about the questions to be answered by the evaluators when assessing every principle. And, finally, we wanted to have a real evaluation scenario that tests all the proposals. All together it should provide us with helpful comments and critiques.

For this we planned a real evaluation. By that time the faculty where we work had launched a completely redesigned website (http://www.eps.udl.cat) and the dean asked us about its usability. So, it was a perfect exercise to do. We could meet the request of our dean while being able to test the heuristic validation proposal.

We have recruited 7 evaluators, all them with valuable experience doing evaluations with heuristics. Three with more than 10 years of experience, two with more than 5 years, one of whom developed her PhD completely in this specific field, and a PhD student that follows the previous one. The last student has less experienced but is supervised by one of the seniors.

The evaluation process was very simple. Every evaluator received a spreadsheet file containing: a sheet for the project description and the evaluator data, 15 sheets (one for each principle) containing the principle and its questions, and a last sheet with the global calculation (blocked for the evaluator).

The questions on the principles sheets had to be answered with the score (one of the four available options) and with an additional cell for comments to describe the evaluator's feedback.

TABLE IV shows the final scores of all the evaluators for every principle (this is what is shown in the last sheet). In this table, the important values are those along the bottom row, corresponding with the Usability Percentages.

Analysing these results, we can observe that, in general terms, the website evaluated has a good usability level. Figure 1 summarizes all the evaluations, the mean of all the evaluations is 77.5%, that means (or, could mean) a reasonably good usability level.

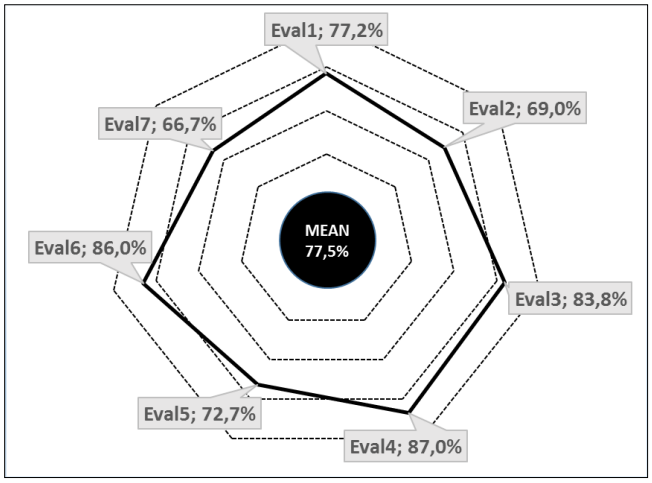

Figure 1. Usability percentage.

Moreover, if we want to be more precise, we can analyse the evaluation by each principle. Then, principles "*10 - Help and documentation*", "*11 - Save the state and protect the work*", "*14 – Defaults*" and "*15 - Latency reduction*" demonstrate the lowest values (in TABLE IV values are coloured -from red for worst values to green for the best-, easing this analysis), then it is very easy to see which are the aspects to be enhanced.

V. DISCUSSION

Once this work is presented, the discussion will be focused on what we learned from the experiment to validate the proposal presented here.

The first aspect to comment refers to the principles list. We are convinced that the new list is certainly an improvement on its predecessors. This convincement relies on (i) the basis of it is the result of shaking the most well used list, Nielsen's, and another not so well used but more precise and complete list, Tognazzini's, and, (ii) on the comments provided by the evaluators who took part in the experiment. For example, the sentence "*this new list is a bit larger than Nielsen's, but it is much more complete*" had a unanimous consensus.

The second aspect is about the questions. Here we must note that it is impossible to assess any single principle without the questions, hence their need. In our case, we delivered a list that also comes from the analysis of the two references used. We believe that these questions cover all the aspects to carry out a proper UI analysis. Certainly, they can be enhanced, but it is a reasoned and well-balanced list, with an appropriate number of questions for each principle.

The next characteristic refers to the rating method. Here, the first good decision was to write all the questions in question form and describe them with the same usability direction, the affirmative answer always representing good usability. This may seem unimportant, even trivial, but it is not usual to find it in the previous works. At the same time, it is a determinant for our score method. The second good decision was to simplify the rating scale. We know that larger scales allow evaluators to be more precise, but we also know that as the evaluation advances, the ratings for the last principles are less precise.

TABLE IV. RESULTS FORM ALL THE EVALUATORS

	Eval1	**Eval2**	**Eval3**	**Eval4**	**Eval5**	**Eval6**	**Eval7**
1- Visibility and system state	5	5	4	4.5	4	5	4
2 - Connection between the system and the real world, metaphor usage and human objects	3	2	4	4	3.5	4	2
3 - User control and freedom	3	2	3	2	2	1	1.5
4 - Consistency and standards	5	4	5	5.5	4	4	3.5
5 - Recognition rather than memory, learning and anticipation	5	4	4	5	5	5	4
6 - Flexibility and efficiency of use	5	3	3	6	3	5	3
7 - Help users recognize, diagnose and recover from errors	3	3	0	0	2	2	4
8 - Preventing errors	2	2	2	0	2	2	1
9 - Aesthetic and minimalist design	4	3	4	3	4	4	1
10 - Help and documentation	0	0	0	0	0.5	0	0
11 - Save the state and protect the work	2	0	0	1	1	0	0
12 - Colour and readability	4	3	2.5	4	4	2	2
13 - Autonomy	2	2.5	2	3	3	3	2
14 - Defaults	0	0	0	2	1	0	2
15 - Latency reduction	1	1	0	0	1	0	0
Usability Percentage (UP)	**77.2%**	**69.0%**	**83.8%**	**87.0%**	**72.7%**	**86.0%**	**66.7%**

Both previous characteristics, i.e., the form of the questions and the rating method, enable us to quantify every evaluation done by the evaluators. It gives a global evaluation point of view that we called the Usability Percentage (UP). This value must not be taken strictly as a full usability meaning, but as an orientation.

About this aspect, from the analysis of the experiment, beyond the result, we obtained the need to find the way to compare not only the UP but the results of every single principle. Let us explain, in TABLE IV we see a lot of numbers that at this moment are slightly insignificant. These are the values obtained at every principle by every evaluator. They are used to calculate the UP, so are significant. But we must find a way for comparing these values principle by principle. This will allow us to identify strengths and weaknesses and, then, to orientate the efforts to be made to improve the usability of the interface. Certainly, this will be the next step to enhance our methodology.

VI. Conclusion and Future work

A new set of design principles has been presented with the intention to be used in Heuristic Evaluations. Additionally, a new methodology for evaluating completes the proposal, providing a quantitative rate named Usability percentage which complements the usual qualitative data provided by this type of evaluations. An experiment demonstrated how useful HE is and how easy it was to put our proposal into practise.

This is a work in progress in the sense that finding the specific strengths and weaknesses will make the system better and more useful. The proposed evaluation system is a potential candidate to replace existing methods. It is the simplification of the evaluation and the final value that make our system attractive.

References

[1] K. Baker, S. Greenberg, and C. Gutwin, "Empirical development of a heuristic evaluation methodology for shared workspace groupware". Proceedings of Conference on Computer Supported Cooperative Work, pp. 96–105, 2002.

[2] L. Bonastre and T. Granollers, "A Set Of Heuristics for User Experience Evaluation in E-commerce Websites". Proceedings of the Seventh International Conference on Advances in Computer-Human Interactions (ACHI 2014). ISBN: 978-1-61208-325-4, pp. 27-34, 2014.

[3] C. Cockton, "Usability Evaluation". The Encyclopedia of Human-Computer Interaction, 2nd Ed, Chapter 15. [Online]. Available from: www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/usability-evaluation [retrieved: January, 2018].

[4] T. Granollers, "Usability evaluation with heuristics. New proposal from integrating two trusted sources". HCI International 2018. 20th International Conference on Human-Computer-Interaction.

[5] N. J. Lightner, "Evaluating e-commerce functionality with a focus on customer service".Communications of the ACM, 47 (1), pp. 88-92, 2004.

[6] L. Masip et al., "A design process for exhibition design choices and trade-offs in (potentially) conflicting user interface guidelines". Lecture Notes in Computer Science, Vol 7623, pp 53-71. IFIP International Federation for Information Processing, 2012.

[7] L. Masip, M. Oliva, and T. Granollers, "OPEN-HEREDEUX: open heuristic resource for designing and evaluating user experience". Proceedings of the 13th IFIP TC 13 international conference on Human-computer interaction (INTERACT'11), P. Campos, N. Nunes, N. Graham, J. Jorge, and P. Palanque (Eds.), Part IV. Springer-Verlag, Berlin, Heidelberg, pp. 418-421, 2011.

[8] L. Masip, T. Granollers, and M. Oliva, "A Heuristic Evaluation Experiment To Validate The New Set Of Usability Heuristics". Proceedings of the 8th International Conference on Information Technology: New Generations, (Washington, DC, USA, 2011), IEEE Computer Society, pp. 429-434, 2011.

[9] A.B. Martínez, A. A. Juan, D. Álvarez, and M. C. Suárez, "WAB*: A quantitative metric based on WAB". Proceedings of the 9th International Conference on Web Engineering (ICWE 2009), LNCS Vol. 5648, pp. 485-488, Springer-Verlag Berlin Heidelberg. 2009.

[10] J. Nielsen, "10 Usability Heuristics for User Interface Design" [Online]. Available from: https://www.nngroup.com/articles/ten-usability-heuristics, 1995. [retrieved: January, 2018].

[11] J. Nielsen, "Severity Ratings for Usability Problems" [Online]. Available from: www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems, 1995. [retrieved: January, 2018].

[12] J. Nielsen, "Finding usability problems through heuristic evaluation". Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 373-380). ACM, 1992.

[13] D. Quiñones, "A methodology to develop usability/user experience heuristics". Proceedings of the XVIII International Conference on Human Computer Interaction (Interacción '17). ACM, New York, NY, USA, Article 57, 2 pages. DOI: https://doi.org/10.1145/3123818.3133832, 2017.

[14] D. Quiñones and C. Rusu, "How to develop usability heuristics: A systematic literature review". Computer Standards & Interfaces, Volume 53, 2017, Pages 89-122, ISSN 0920-5489, https://doi.org/10.1016/j.csi.2017.03.009.

[15] P. C. Realpe, "Proceso de Diseño en Seguridad Usable y Autenticación mediante un enfoque Centrado en el Usuario". Doctoral Thesis, Universidad del Cauca (Colombia), 2017.

[16] J. Rubin and D. Chisnell, "Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests". Ed. Wiley (2 edition), 2008. ISBN-13: 978-0470185483

[17] C. Rusu, S. Roncagliolo, V. Rusu, and C. A. Collazos, "A methodology to establish usability heuristics". Proceedings ACHI2011: The Fourth International Conference on Advances in Computer-Human Interactions, 59-62, 2011.

[18] J. Sauro, "Rating the Severity of Usability Problems" [Online]. Available from: https://measuringu.com/rating-severity, 2013. [retrieved: January, 2018].

[19] B. Shneiderman, "Designing the User Interface: Strategies for Effective Human-Computer Interaction". Addison-Wesley Publ. Co., Reading, MA, 1987.

[20] M. C. Suarez, "SIRIUS: Sistema de Evaluación de la Usabilidad Web Orientado al Usuario y basado en la Determinación de Tareas Críticas". Doctoral Thesis. Oviedo (Spain), 2011.

[21] B. Tognazzini, "First Principles, HCI Design, Human Computer Interaction (HCI), Principles of HCI Design, Usability Testing". [Online]. Available from: http://www.asktog.com/basics/firstPrinciples.html, 2014 [retrieved: January, 2018].