On the Double Interpretation of the Description of the "Consistency and Standards" Heuristic in the Heuristic Evaluation Method

Bruno Thomé Júnior, Lucas de Souza Timóteo,
Wagner Ferreira da Silva Júnior and Celmar Guimarães da Silva[a]
School of Technology, University of Campinas, Limeira, Brazil

Keywords: Heuristic Evaluation, Consistency.

Abstract: Heuristic Evaluation (HE) is a well-known method for assessing usability in Human-Computer Interaction. Among Nielsen's usability heuristics for HE, "Consistency and standards" heuristic (H4) is related to consistency, a relevant concept for usability. However, the description of this heuristic does not deal with some relevant aspects of the concept of consistency. Therefore, the literal interpretation of this description can force some evaluators to search for another heuristic to associate with some inconsistency problems – a search that may or may not be successful. This paper is part of a research that questions if the description of this heuristic must be improved. In particular, this paper aims to understand how evaluators may behave when they do this literal interpretation during the execution of a HE. Our first step towards these goals was to analyze Nielsen's heuristics in order to search if other heuristics than H4 also have consistency aspects. We also hypothesized how some evaluators may associate these heuristics with specific inconsistency problems. We noted that at least 5 heuristics (in addition to H4) have some aspect of consistency. Besides, we observed that it is possible to construct a logical reasoning that may associate inconsistency problems to one of these heuristics. We suggest that it is worth doing a discussion about extending the current description of the consistency heuristic to include more aspects of consistency, instead of allowing this literal interpretation to deviate evaluators from the expected interpretation of H4 heuristic.

1 INTRODUCTION

Heuristic Evaluation (HE) is a well-known user interface evaluation method in Human-Computer Interaction (HCI) area. It is an inspection method introduced by Nielsen and Molich (1990) that aims to reveal usability problems without the need of user participation in the evaluation process. Its authors based their research on an extensive set of guidelines. They summarized them into nine heuristics, which were further refined to ten heuristics (Nielsen and Molich, 1990; Nielsen, 1994) (which we will refer to as "Nielsen's heuristics", for short).

Despite its widespread acceptation and use for more than 25 years by the HCI community, we observed that the Nielsen's heuristics hide some subtleties that may confuse novice evaluators (as we observed in HCI classes about HE), and that may be invisible to many of the expert ones. In this paper we focus on the fourth heuristic ("Consistency and standards", or H4 for short), which is related to the concept of consistency. It is known that designers must strive for consistency (Shneiderman and Plaisant, 2004) in many aspects, such as consistency among screens of the same software, or between the designers' and the users' mental models. However, the H4 heuristic states only that "Users should not have to wonder whether different words, situations, or actions mean the same thing" and that they must "Follow platform conventions" (Nielsen, 1994). In other words, this description is only related to parts of the concept of consistency.

We hypothesize that this conciseness may be a source of ambiguity for evaluators. On the one hand, we believe that some evaluators may link every inconsistency problem to H4, even when this problem is not related to the description of this heuristic. This may happen due to two situations at least. First, it may be an oversimplification of the problem by the novice evaluator, to which every inconsistency undoubtedly matches the heuristic "Consistency and standards" (regardless its description), as we observed in some classes about HE. Second, the expert evaluator con-

[a] https://orcid.org/0000-0001-6112-892X

Thomé Júnior, B., Timóteo, L., Silva Júnior, W. and Guimarães da Silva, C.
On the Double Interpretation of the Description of the "Consistency and Standards" Heuristic in the Heuristic Evaluation Method.
DOI: 10.5220/0010313002220228
In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2021) - Volume 2: HUCAPP*, pages 222-228
ISBN: 978-989-758-488-6

sciously extends the meaning of the H4's description to cover complementary aspects of consistency due to his experience with HCI (as in the detailed version of H4's description presented by Abulfaraj and Steele (2020)).

On the other hand, we think that part of the evaluators will try to follow literally the description of the heuristic. This may happen when the novice evaluator is executing a heuristic evaluation by the first time and follows exactly the heuristics' descriptions (as we also observed in HCI classes), or because the evaluators simply believe that the description of H4 defines all possibilities covered by the heuristic. When considering this literal meaning of the H4's description, they may discover that a given inconsistency problem does not fit this description. In this case, they will probably try to find another heuristic that can capture some aspects of the inconsistency problem under evaluation.

This research focus on this double interpretation of H4: the generalized interpretation of the H4's description versus its literal meaning. In a broad sense, given this duplicity, we want to discover how evaluators perceive the meaning of H4 and of its description, and if it needs changes to be better understood and used.

As a preliminary step towards this direction, this paper aims to explore situations in which an evaluator, focused on the literal meaning of H4's description, faces an inconsistency problem that does not fit this description. First of all, we argue that some consistency aspects are also present in other of the Nielsen's heuristics. Considering this argument, we hypothesize that this kind of evaluator will try to find another heuristic with complementary consistency aspects to associate with the inconsistency problem under analysis. We suggest that both interpretations enable evaluators to detect inconsistency problems through HE. However, we argue that a better description of H4 would help evaluators to group all the consistency-related problems they found, instead of distributing them over a set of heuristics. Besides, this improved description may reduce the learning curve of HE by novice evaluators and also may enable a consistent use of the "Consistency and standards" heuristic.

This paper is organized as follows. Section 2 presents some works related to the concept of consistency, including mismatches related to the classification of inconsistency problems in a HE. Section 3 points out some consistency aspects we found in Nielsen's heuristics other than H4. It also illustrates how inconsistency problems may be associated with these heuristics. Besides, this section argues about the need to change the H4's description to embrace more aspects of consistency. Section 4 concludes the paper and points out future work.

2 RELATED WORK

This section briefly discusses some aspects of the concept of consistency and presents some classifications of these aspects. Besides, it presents some works that sought to broaden the understanding or coverage of Nielsen's heuristics. It also makes some considerations about the execution of HE by novice evaluators and their expected failures during this evaluation.

The term "consistency" has many embedded aspects. Some of them are registered in the works of Nielsen and Molich (Molich and Nielsen, 1990; Nielsen and Molich, 1990; Nielsen, 1994) and Shneiderman and Plaisant (2004). When Molich and Nielsen published their first nine heuristics (Molich and Nielsen, 1990), they called evaluators' attention for consistency in their HE method, with a heuristic called "Be consistent". According to its description, "Users should not have to wonder whether different words, situations, or actions mean the same thing. A particular system action – when appropriate – should always be achievable by one particular user action. Consistency also means coordination between subsystems and between major independent systems with common user populations".

Four years after the publication of these heuristics, Nielsen (1994) changed the set of heuristics and replaced this heuristic by a new one called "Consistency and standards" (which is the focus of this paper, and was already presented at the Introduction section). Its description became more concise than the previous one, and part of the description related to the comparison among systems was lost.

The eight golden rules of interface design from Shneiderman and Plaisant (2004) also include a rule related to consistency: "Strive for consistency". It is related to some forms of consistency, such as using consistent sequences of actions for similar situations, and using identical terminology and formatting (colors, layout, fonts etc.) throughout the system.

Rocha and Baranauskas (2003) goes even further by indicating the relevance of achieving consistency throughout all the media that composes the interface: screens, help, training material and so on.

Consistency is also related to accessibility. When the Nielsen's heuristics are compared with the WCAG (Web-Content Accessibility Guidelines) (W3C, 2008), consistency is partially related to the principles "Operable" and "Understandable" (Casare et al., 2019). In other words, a consistent interface

may be easier to understand and operate.

Some authors tried to classify some aspects of the concept of consistency. Tanaka et al. (1991) (apud AlTaboli and Abou-Zeid (2007)) present two possible classifications of consistency. In the first one, *internal consistency* refers to consistency within a task, and *external consistency* is related to consistency among distinct tasks (such as in the heuristic "Be consistent"). A second classification defines *cognitive consistency* as "the consistency in what the user knows", and *display consistency* as "the consistency in the layout of screen displays" (AlTaboli and Abou-Zeid, 2007).

Other authors (Kellogg, 1989; Schlatter and Levinson, 2013; Moran, 2020) define the terms "internal" and "external" in a different way. They call *internal consistency* the consistency between screens of an application or a family of products, and *external consistency* the consistency between applications. We will use these definitions along this paper.

Some expert evaluators recommend to verify these last two aspects of consistency during a HE. Salgado et al. (2016) surveyed a set of 38 tactics suggested by experts to improve the performance of novice evaluators when they are executing a Heuristic Evaluation. Three of these tactics are directly related to H4. First, the tactic T12 is related to the analysis of whether there is a consistent use of terms (or other interface elements) when referring to a same content. The tactic T13 states that evaluators should verify the consistency about causes and effects, i.e., if similar actions cause similar effects in the interface. Finally, the tactic T14 suggests that evaluators must analyze if the application is consistent with other similar applications already known by the users. In other words, T14 is a tactic related to external consistency, and T12 and T13 refer to internal consistency.

Some works focus on enhancing the comprehension (Abulfaraj and Steele, 2020) or the coverage (Granollers, 2018a,b; Pribeanu, 2017) of the Nielsen's heuristics. Abulfaraj and Steele (2020) provided a detailed version of these heuristics, based on descriptions of each heuristic provided by experts. They split H4 into two subheuristics: "Consistency" and "Standards". "Consistency" is related to internal consistency and is divided into consistency of meaning, function, effort, organization, and feeling. In its turn, "Standards" is related to external consistency.

Granollers (2018a,b) proposed a combination between Nielsen's heuristics and Tognazzini's principles of interaction design. This combination also has a "Consistency and standards" heuristic that is detailed in the following questions to be answered by the evaluator:

- 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, ...)

Pribeanu (2017) merged Nielsen's heuristics and ergonomic criteria. His new list of heuristics has 14 heuristics that were grouped into the following general ergonomic criteria: user guidance, user effort, user control and freedom, and user support. He also provided detailed versions of each heuristic. In particular, his consistency heuristic considered the following items:

- Provide similar phrasing, text justification, color, and punctuation.
- Display similar objects (windows, menus, exit buttons, etc.) in the same way and at the same location.
- Provide similar procedures for similar functions and tasks.
- Follow platform conventions.

It is worth noting that the works of Abulfaraj and Steele (2020), Granollers (2018b,a), and Pribeanu (2017) did not use the Nielsen's definition of H4. They provided their own definitions with more details about their consistency heuristics, aiming to guide evaluators' comprehension about them.

Regarding classification mismatches related to H4, a relevant work from Salgado and Fortes (2016) surveyed the number of expert evaluators (among 13 ones) that believed that novice evaluators would associate with a heuristic *b* a problem that should be associated with heuristic *a*. For the problems in which *a* is H4, 10 of the experts supposed that novice evaluators could mistakenly associate a given consistency problem with the heuristic "Match between system and the real world". Besides, 5, 4, and 4 experts, respectively, supposed that novice evaluators can erroneously relate the problem to the heuristics "Recognition rather than recall", "Aesthetic and minimalist design", and "Help users recognize, diagnose, and recover from errors". For each one of the other heuristics, 2 or 3 experts believe that the problem is related to that particular heuristic and not to H4. This research indicates that expert evaluators expect some kind of confusion regarding the classification of problems related to the

H4 heuristic. However, the paper does not discuss why this mismatching may happen.

In summary, consistency is a relevant concept of usability and has some facets that should be considered by novice and expert evaluators. Particularly, it is possible that novice and expert evaluators link inconsistency problems to different heuristics (which may not be an illogical choice, as we will explain later). Regarding the consistency heuristic, the suggested tactics for novice evaluators to avoid wrong classifications are to pay attention to both internal and external consistency concepts during the execution of the HE. This suggests that these aspects could be added to the description of H4.

3 METHODOLOGY AND RESULTS

In the current work, our first step was to identify consistency concepts that we observed inside the Nielsen's heuristics (except H4). Next, we analyzed examples of inconsistencies related to these concepts and that are not related to the literal description of H4. Each example was associated with a heuristic that could capture the consistency aspect related to the usability problem under analysis.

3.1 Consistency Concepts inside Heuristics

Despite the existence of the heuristic "Consistency and standards", which is clearly related to consistency, we noted that some of the other Nielsen's heuristics also have aspects related to consistency, and that these aspects are not covered by the literal description of H4. We list some of these heuristics in this section.

First of all, the heuristic "Match between system and the real world" (H2) has the following description: "The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order" (Nielsen, 1994). However, if we rephrase this description, its consistency-related aspects become evident: "The system should speak the users' language: it should use words, phrases and concepts that are *consistent* with those known by users, rather than system-oriented terms. Be *consistent* with real-world conventions, making information appear in a natural and logical order". In other words, this heuristic could be renamed to "*Consistency* between system and the real world".

Other heuristics cannot be easily rephrased to include the concept of consistency, but are also related to it. As a first example, the heuristic "Error prevention" (H5) states that designers must prevent a problem from occurring instead of presenting good error messages. A careful design that prevents an error to occur is *consistent* with designer's and user's mental models.

"Recognition rather than recall" (H6) is another example of a heuristic that is not easily related to consistency. However, suppose a situation in which a system has two parts, and a same button in both. The button has distinct functions in each part. Users must remember that the button in part 1 has a distinct meaning from the button in part 2. This may also be seen as an inconsistency problem to be solved.

As a third case, the heuristic "Aesthetic and minimalist design" (H8) states that "Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility". We understand that inconsistencies such as wrong formatting, unnecessary repetition of objects, blank screens or screens without information are problems that compete with the content of a site, and then they may be associated with this heuristic.

Last but not least, the heuristic "Help and documentation" (H10) defines that the system and its help should be *consistent* to each other.

All of these situations focused on internal consistency. They are not an exhaustive list, but they point out that some inconsistency problems may be associated with H4 (in the generalized interpretation of H4's description) or to one of the other five heuristics discussed in this section (if the literal description of H4 is considered). That is, there is a chance of association between an inconsistency problem unrelated to H4's description and another heuristic. However, it is not possible to ensure that the heuristics other than H4 can be associated with every inconsistency problem not related to H4's description. This is an argument in favour of improving the description of H4.

It is worth noting that 3 of our situations (H2, H6, and H8) are also cited by Salgado and Fortes (2016) as part of the "top 4" problems regarding classification of consistency problems. Given that our analysis was made without prior knowledge about their work, this intersection signalizes that the literal interpretation of H4's description may be a cause for some of the problems they cite.

Figure 1: Distinct menus in Portuguese (top) and English (bottom) versions of the same site.

3.2 HE Considering H4's Literal Meaning

In this section we exemplify some real-world usability problems that are related to consistency and that do not fit the H4's description. We associate each of these problems with another heuristic than H4, and present explanations about the choice we made.

- *Distinct versions of the same website, for distinct languages, do not have the same content.* As an example, a website in Portuguese has a simplified English version of its content for foreigners, with less content and a menu with less options. In Figure 1 (top), the blue menu has the options "Apresentação" (Presentation), "Coordenação de Curso" (Course Coordination), Comissões ("Commissions"), "Cursos" (Courses), and "Disciplinas e horários" (Subjects and schedules). In Figure 1 (bottom), the options have names of courses. These versions are therefore inconsistent with each other. If we consider that in the real world a translation does not remove parts of the original text or change its structure, we can classify this problem as disregarding the heuristic "Match between system and the real world" (H2).
- *Distinct links are so close that they seem to be a single one.* Figure 2 illustrates this situation. The figure shows a list of five bullets. In each bullet we have two distinct links, but in the first two these links are concatenated. Probably this situation happened due to an incorrect link reuse. The use of a "2-in-1 link" is not consistent with the link usage in web pages, and this may lead users to select a wrong link. Therefore, the heuristic "Error prevention" fits this situation.

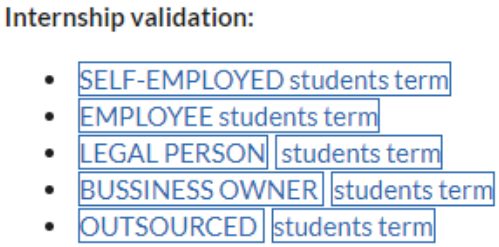

Figure 2: Each of the last three bullets have two links. Each of the two first bullets also have two links, but they are concatenated and seems to be just one link.

- *Using two languages in a same screen without a clear reason.* Some websites whose content is written in a given language present contents in other languages. This may be due to the use of interface components that are not correctly configured or translated. As an example, web pages from laboratories of an institute have content in Portuguese, except one that is in English. It is a consistency problem, but it is not related to H4's description. If a user wants to navigate in this page and has no knowledge about English language, he may need to rely on his previous experience regarding the interaction with the pages of other laboratories in Portuguese. Therefore, the heuristic "Recognition rather than recall" may be associated with this problem.
- *A text has a format that is inconsistent with the format of other texts with similar function in the screen, without a clear reason.* For example, in Figure 3, the description of a course is in bold, and the description of other courses is not. Given that the description of H4's does not cover this case, evaluators may consider this situation as an aesthetic problem. In this case, the bold font is an extra unit of information that may diminish the relative visibility of the other course descriptions. Therefore, it violates the heuristic "Aesthetic and minimalist design".
- *Showing contentless or incomplete web pages.* A typical example of this situation is a page that has only the message "coming soon" or "under construction" (Figure 4). When a web page points to this kind of problematic pages, an evaluator may define this case as an inconsistency problem.

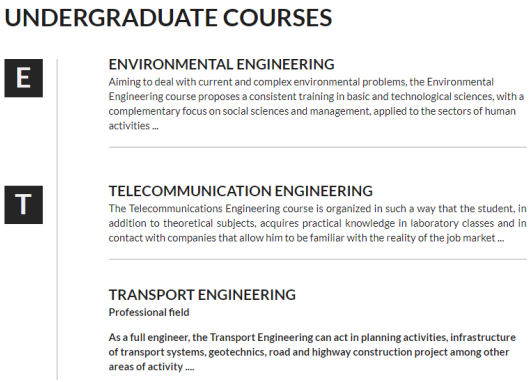

Figure 3: A web site using bold for the description of one course (Transport Engineering) and normal font for the other descriptions.

Figure 4: A web site with a menu, a title "TEAM" and a message "Coming soon".

In other words, it would be consistent to provide links that point to pages with useful information. Again, H4's description does not cover this case. We may consider that this problem disrespect the heuristic "Aesthetic and minimalist design", given that the interface provides a useless link that could be removed.

- *Using symbols with unknown meaning for the user.* The use of a symbol whose meaning is not clear for the target users may be explained as an inconsistency between the symbols that the system uses and the ones that the users know. For example, a button with the symbol "@" that sends an e-mail to a person in his profile page is useless if the user does not understand the "@" as a kind of abbreviation to the concept of e-mail (Figure 5). The heuristic "Match between system and the real world" (H2) is directly related to this kind of usability problem. Note that, as we already argued, the proposed match is a kind of consistency, but with a particular heuristic that should be used instead of H4.

Figure 5: The char "@" as a link to a personal e-mail.

It is worth reinforcing that in all these cases the concept of consistency is present, but the description of the consistency heuristic is not related to them.

3.3 Discussion

The last two subsections presented some situations that violate the concept of consistency but that are not related to the description of H4. They revealed how restrict is this description, and why some evaluators, with focus on the description, may try to select other heuristics than H4 for an inconsistency problem.

We presented in Section 3.2 a reasoning that elected other heuristics instead of the consistency heuristic when its description is not applicable to a given inconsistency problem. However, considering that the heuristics should be easy to learn, we believe that it would be relevant to the HCI community to think about expanding the description of H4 to include other inconsistency situations. As a matter of fact, we believe that the inclusion of the concepts of internal and external consistency – even that quickly explained inside the heuristic's description – would be of great help for evaluators, and mainly for novice ones.

4 CONCLUSION

In this paper, we argued that a literal interpretation of the description of "Consistency and standards" heuristic may force evaluators to find another Nielsen's heuristic to associate with some types of inconsistency problems. We observed that some heuristics other that H4 also have aspects related to consistency, and presented how some evaluators may resort to these heuristics to reveal inconsistency problems. Supposing that the original intention of the heuristics' authors was that the fourth heuristic should be associated with most of the inconsistency problems, we argue that HCI community can expend some effort on refining the description of this heuristic. The efforts

of some works on providing detailed versions of their heuristic sets may help to avoid some of the ambiguities regarding H4 that were presented in this paper.

In any way we want to propose the creation of a very detailed description of the fourth heuristic. However, we believe that generalizing and enhancing this description would avoid some negative effects of its literal interpretation, e.g. the detaching of inconsistency problems in the HE report, or some situation in which no other heuristic could be associated with the inconsistency problem at hand.

This is an initial work regarding this theme. We know that Nielsen's heuristics were accepted and have been adopted by the HCI community for decades, and that they are general enough to be used to evaluate the usability of many kinds of interfaces. Therefore, our purpose here is to do a constructive criticism that may enhance the comprehension of these heuristics and reduce ambiguous interpretations.

We also believe that proposing changes to this consolidated set of heuristics is a work that should not be done according to the opinion of a small group of researchers. Therefore, future work includes researching how the HCI community classifies inconsistency problems in the Heuristic Evaluation, how often the literal interpretation of the "Consistency and standards" heuristic occurs, and promoting a discussion of the need to change that description to overcome interpretation problems.

ACKNOWLEDGEMENTS

We thank the grant #2015/21 from FAEPEX/PRP/Unicamp.

REFERENCES

Abulfaraj, A. and Steele, A. (2020). Detailed usability heuristics: A breakdown of usability heuristics to enhance comprehension for novice evaluators. In Stephanidis, C., Marcus, A., Rosenzweig, E., Rau, P.-L. P., Moallem, A., and Rauterberg, M., editors, *HCI International 2020 - Late Breaking Papers: User Experience Design and Case Studies*, pages 3–18, Cham. Springer International Publishing.

AlTaboli, A. and Abou-Zeid, M. R. (2007). Effect of physical consistency of web interface design on users' performance and satisfaction. In Jacko, J. A., editor, *Human-Computer Interaction. HCI Applications and Services*, pages 849–858, Berlin, Heidelberg. Springer Berlin Heidelberg.

Casare, A. R., Silva, C. G., Martins, P. S., and Moraes, R. (2019). Mapping WCAG guidelines to Nielsen's heuristics. In *Proceedings of the 16th International Conference on Applied Computing*, pages 79–88.

Granollers, T. (2018a). Usability Evaluation with Heuristics, Beyond Nielsen's List. In *Proceedings of the Eleventh International Conference on Advances in Computer-Human Interactions (ACHI)*, pages 60–65.

Granollers, T. (2018b). Usability Evaluation with Heuristics. New Proposal from Integrating Two Trusted Sources. In Marcus, A. and Wang, W., editors, *Design, User Experience, and Usability: Theory and Practice*, pages 396–405. Springer International Publishing.

Kellogg, W. A. (1989). The dimensions of consistency. In *Coordinating User Interfaces for Consistency*. Academic Press.

Molich, R. and Nielsen, J. (1990). Improving a human-computer dialogue. *Communications of the ACM*, 33(3):338–348.

Moran, K. (2020). *Usability Heuristic 4: Consistency and Standards*. Nielsen Norman Group. https://www.nngroup.com/videos/usability-heuristic-consistency-standards/ (video; accessed October 29, 2020).

Nielsen, J. (1994). Usability inspection methods. In *Conference Companion on Human Factors in Computing Systems*, CHI '94, page 413–414, New York, NY, USA. Association for Computing Machinery.

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.

Pribeanu, C. (2017). A revised set of usability heuristics for the evaluation of interactive systems. *Informatica Economica*, 21:31–38.

Rocha, H. V. and Baranauskas, M. C. C. (2003). *Design e Avaliação de Interfaces Humano-Computador*. NIED/UNICAMP.

Salgado, A. L. and Fortes, R. P. M. (2016). Heuristic evaluation for novice evaluators. In Marcus, A., editor, *Design, User Experience, and Usability: Design Thinking and Methods*, pages 387–398. Springer International Publishing.

Salgado, A. L., Lara, S., Freire, A., and Fortes, R. (2016). What is hidden in a heuristic evaluation: tactics from the experts. In *CONTECSI USP - International Conference on Information Systems and Technology Management - ISSN 2448-1041*, pages 2931 – 2946.

Schlatter, T. and Levinson, D. (2013). *Visual Usability Principles and Practices for Designing Digital Applications*. Elsevier.

Shneiderman, B. and Plaisant, C. (2004). *Designing the User Interface: Strategies for Effective Human-Computer Interaction (4th Edition)*. Pearson Addison Wesley.

Tanaka, T., Eberts, R. E., and Salvendy, G. (1991). Consistency of human-computer interface design: Quantification and validation. *Human Factors*, 33(6):653–676.

W3C (2008). *Web Content Accessibility World Wide Web Consortium (WCAG) 2.0*. https://www.w3.org/TR/WCAG20/ (accessed October 29, 2020).