



**HAL**  
open science

# Towards a Framework for the Classification of Usability Issues

Martin Böckle, Jens Rühmkorf

► **To cite this version:**

Martin Böckle, Jens Rühmkorf. Towards a Framework for the Classification of Usability Issues. 17th IFIP Conference on Human-Computer Interaction (INTERACT), Sep 2019, Paphos, Cyprus. pp.610-614, 10.1007/978-3-030-29390-1\_47. hal-02878655

**HAL Id: hal-02878655**

**<https://inria.hal.science/hal-02878655>**

Submitted on 23 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Towards a Framework for the Classification of Usability Issues

Martin Böckle<sup>1</sup> and Jens Rühmkorf<sup>2</sup>

<sup>1</sup> BCG Platinion, Design and Engineering, Berlin, Germany

<sup>2</sup> BCG Platinion, Cologne, Germany

{boeckle.martin, ruehmkorf.jens}@bcgplatinion.com

**Abstract.** This case study proposes a novel framework, which aims to support UX designers and practitioners in the classification of identified usability issues. Existing inspection methods often lack in providing further information and steps about the effort to fix those issues. Therefore, the developed framework provides four categories for classifying usability issues in order to generate a score, which represents the present level of UX maturity. Furthermore, the score enables the comparison of UX maturity levels between systems with similar customer journeys, based on the identified usability issues. Results support system designers in their decision-making process of replacing, maintaining or modernizing the system in question. First results reveal that the proposed framework could be used as a meaningful extension of existing usability methods within different application domains, for instance the heuristic evaluation.

**Keywords:** Usability Issues; Classification; User Experience; Usability Method

## 1 Introduction

The rising level of end-user expectations for seamless user-experience (UX) including the development of novel user interfaces (UI), which aim to follow current usability standards, reveals the importance for the assessment of existing and probably outdated information technology applications. These type of systems (e.g., legacy systems) share common characteristics that certain usability issues have been identified, which conflict with some of the best practices of user-centered design (UCD) and therefore provide an overall negative user-experience (UX). Although, the present body of literature provides several metrics for measuring usability from different perspectives [5, 1], a classification for the identified issues, which aim to support the decision-making process about the overall the re-design or fixing a number of issues is currently missing. Therefore, this case study provides a framework, which clusters usability issues by their severity ratings [7] and the assumed effort to fix from a technological perspective.

We believe that the proposed framework provides a valuable contribution to UX researchers and practitioners, by highlighting the impact of usability issues on the user-journey and the relationship to the expected effort to fix those.

## 2 Usability Metrics and Methods

The term usability has its roots in human computer interaction (HCI) and is defined as the capability to be used by humans easily and effectively” [2, p. 340] or the *effectiveness, efficiency* and *satisfaction* with which specified users achieve defined goals in particular environments [3]. First approaches provided techniques (e.g., thinking aloud) for usability-testing, followed by guidelines and discussions on how to actually measure usability [4]. Further metrics are being discussed, for instance the system usability scale, which provides a rough indication about the systems overall usability, developed by [8] or how to summarize several metrics into one single score [5].

Moreover, the literature provides a range of methods, which can be used at different stages for the assessment of the usability, for instance a heuristic evaluation [6, 9] or the PURE methodology [10]. Although, after applying these methods, the following steps like how to deal with the identified issues has often been neglected. Specifically, for legacy systems, a decision for the overall redesign or how much effort is needed to solve each individual issue needs to be addressed as well and will be covered in this case study.

## 3 Framework development

For the development of the proposed framework, two usability experts conducted a heuristic evaluation of two applications within the finance domain by following the guidelines of [9]. The evaluation was based on the fundamental customer-journey, which comprised of similar characteristics in their outcomes and goals. After the identification of several usability issues (first system = 10; second system = 24), the severity ratings [7], which indicate a rough estimation for usability problems, have been applied. Within this process, key concepts were uncovered including a cross-check between both raters. Furthermore, the identified usability-issues have been structured and clustered according to their characteristics and degree of conflict with the overall user journey. This iterative process has been performed with several software developers in order to come up with four generic categories, which represent the transition between the identified usability issues and the underlying system architecture, described in Table (1):

**Table 1.** Framework categories

---

**I) UI - Optimizations** represent minor improvements on the UI level like: color codes, inconsistent buttons, menus, warnings and feedback. The effort to fix those issues is relatively low.

---

**II) Weak System Support** - The system does not support the end-users in their customer journey. For instance, the accomplishment of tasks (e.g., complete booking process) is con-

---

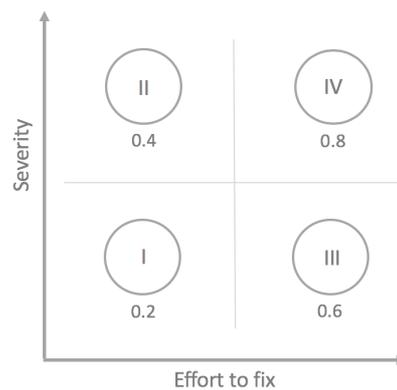
---

nected with considerable effort.

**III) Structural adjustments within the system architecture** are connected to a higher effort to fix (e.g., changes in the business logic)

**IV) High usability conflict with intended customer journey** - These issues prevent the end-user to accomplish their goals and need to be fixed immediately. While end-users will probably drop out of the process, the effort to fix as well as the severity rating is relatively high.

---



**Fig. 1.** Conceptual framework

The first quarter represents usability issues, which are easy to resolve. Generally, those issues don't touch the business logic of a system and only violate UI best practices, like consistency of colors or other elements like menus. In contrast to the second quarter, issues share a higher severity, but the effort to fix is still very low. Moreover, this classification summarizes issues which are connected to weak system support. Examples include the interaction with UI elements, like bad form design, weak error messages or bad general design of the customer journey, which lead to unknown end-user behaviors. The third quarter has a low severity rating but a relatively high effort to fix. For instance, if changes in the database design are necessary, like adding new fields to a form or migrations towards new technologies in order to improve usability aspects (e.g., loading times or adding personalization to an existing system). Furthermore, the redesign of complete processes, which require large changes in the frontend are part of this quarter as well. The fourth quarter represents the area with the highest severity and effort to fix. Those issues should be handled with care, because a high number informs the evaluators about serious usability problems, which are also connected to technological progress or accessibility. Generally, these issues prevent the user from the accomplishing of the intended user-journey like completing a buying process or a weak system architecture design, which makes the system unusable. Moreover, in order to compare similar applications and their effort to fix the identified usability issues, we came up with a simple formula, which aims to highlight the significance of each issue from a practical perspective. Firstly, we considered weights (e.g.,  $wI$ ) with a difference of 0.2 between each quarter to express an increasing relevance for the overall effort. Secondly, we added increasing exponents to the

number of identified usability issues ( $x$ ) in order to explicitly express the degree of severity and effort to fix, which will clearly lead to different scores. The overall score represents the maturity level of the assessed application:

$$y_u = w_1 * x^1 + w_2 * x^2 + w_3 * x^3 + w_4 * x^4$$

Finally, we calculated the proposed usability factor for both systems with following results, described in Table (2):

**Table 2.** Results Usability Inspection

| Usability Issues - System 1 |                  |       | Usability Issues - System 2 |                  |       |
|-----------------------------|------------------|-------|-----------------------------|------------------|-------|
| Quarter                     | Number of issues | Score | Quarter                     | Number of issues | Score |
| Q1                          | 11               | 2.2   | Q1                          | 4                | 0,8   |
| Q2                          | 5                | 10    | Q2                          | 1                | 0,4   |
| Q3                          | 4                | 38,4  | Q3                          | 5                | 75    |
| Q4                          | 4                | 204   | Q4                          | 0                | 0     |
|                             | 24               |       |                             | 10               |       |
|                             |                  | 254,6 |                             |                  | 76,2  |

First results reveal that system one contains a higher level of usability issues than system two. Especially, due the high exponent (4) in the last quarter, the four issues had a huge impact on these results. Thus, only through a high amount of usability problems in the third and specifically in the fourth quarter, the proposed factor will increase.

Generally, the framework represents an extension of the heuristic evaluation by classifying the ranked severity ratings into the proposed framework categories (Table 1) in order to generate an overview of the overall usability issues and their impact on the development process. The classification process should involve system designers, which are able to provide realistic effort assumptions about the identified issue and their relationship to the proposed categories in Table 1 (e.g., “Does the identified issue cover simple UI optimizations or structural adjustments as well?”). Furthermore, the proposed framework should also support the decision-making process of replacing, maintaining or modernizing the system in question (e.g., complete redesign or which issue should be solved first). Although the approach of comparing two systems on the basis of their usability and the effort to solve those issues seems promising, the technological perspective needs to be addressed as well. For instance, systems with different technology stacks cannot directly be compared. Thus, we analyzed the effort to fix from a practical perspective by considering technological dependencies, which may influence the evaluation process and need to be taken into consideration:

**Table 3.** Technological dependencies

|  |
|--|
| “Component based architectures foster reusable components and therefore aim to handle changes more quickly.” |
| “UI libraries of modern web-based architectures can be updated much more flexible than                       |

|   |
|---|
| older systems with less capabilities.”  |
| “Modern user interfaces solve problems in different ways and therefore a comparison of identified issues between two system could be unrealistic.”  |
| “Generally, current tech stacks have a higher focus on maintainability.”  |
| “Especially for legacy systems, the effort to fix evolves around the replacement, maintenance or the modernization of functionalities or the entire system. Consequently, this requires extensive expertise, preparation and planning as well. Therefore, the decision-making process is also more time consuming.” |

## 4 Conclusion

The presented framework aims to support the process of classifying identified usability issues regarding their effort to fix by coming up with a score, which highlights the present UX maturity level. The evaluation of two applications demonstrates the potential of this approach and reveals that the identified issues with a high severity rating and effort to fix should be considered as highly problematic. Furthermore, we discussed technological dependencies when comparing two similar applications within the same application domain. While this framework highlights a meaningful extension for a range of usability methods (e.g., heuristic evaluation, cognitive walkthrough etc.), we plan to further develop the scoring design and range of clusters to come up with more accurate results.

## References

1. Hornbaek, K.: Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, 64, 2005, pp. 79-102
2. Shackel, B.: Usability - Context, framework, definition, design and evaluation. *Interacting with Computers*, 21, 2009, pp. 339-346
3. ISO 9241.: Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts, 2018
4. Nielsen, J., Levy, J.: Measuring usability: preference vs. performance. *Communication of the ACM*. 37 (4), 1994, pp. 66-75
5. Sauro, J., Kindlund, E.: A Method to Standardize Usability Metrics Into a Single Score. In *Proceedings of the CHI 2005*, Portland, USA
6. Nielsen, J., Molich, R.: Heuristic evaluation of user interfaces. In *Proceedings of ACM CHI’90*, Seattle, WA, 1990, pp. 249-256
7. Nielsen, J.: Reliability of severity estimates for usability problems found by heuristic evaluation, In *Proceedings of CHI’92*, 1992, pp. 129-130
8. Brooke, J. (1996). "SUS: a "quick and dirty" usability scale". In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland. *Usability Evaluation in Industry*. London: Taylor and Francis
9. Nielsen, J. - “Heuristic Evaluation: How to conduct a heuristic evaluation?”, <https://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/>, last accessed 2019/6/3
10. Rohrer, C.P., Wendt, J., Sauro, J., Boyle, F., Cole, S.: Practical Usability Rating by Experts (PURE): A Pragmatic Approach for Scoring Product Usability. In *Proceedings CHI 2016*, pp. 786-795, San Jose, USA