

WDP vs. Nielsen's Heuristics: A Comparison

Oliver Petter¹, Svenja Polst², Achim Ebert¹, Denis Feth²

Technische Universität Kaiserslautern, Deutschland¹

Fraunhofer-Institut für Experimentelles Software Engineering, Kaiserslautern, Deutschland²

{o_petter14, ebert}@informatik.uni-kl.de,
{svenja.polst, denis.feth}@iese.fraunhofer.de

Abstract

Most usability evaluation methods (UEM) and heuristics, such as Nielsen's heuristics, are tailored to usability professionals. However, these experts are rare and expensive. Thus, we empirically evaluated whether the web usability inspection technique based on design perspectives (WDP) can achieve equally valuable results while being less dependent on expert knowledge. The results indicate that WDP applied by developers can be a good alternative for usability evaluations by experts.

1 Introduction and Related Work

Software usage in modern workflows is increasingly characterized by the use of web applications. When evaluating the usability of web applications, it is important to consider their characteristics, e.g., responsiveness, cross-browser support and data protection. There is a wide range of usability evaluation methods (UEM) that can be used to evaluate web applications during their development process. Each of these methods requires different resources and performs differently in a given context. Most studies have identified heuristic evaluation as a suitable UEM (Andre et al., 1999). However, Nielsen's heuristic evaluation works best for usability experts (Nielsen, 1992), especially as the heuristics are abstract and inconclusive (Quinones et al., 2014). From an economic point of view this is problematic because the time of experts is very rare and expensive. In addition, there is no clear guidance for the application of heuristics.

Perspective-based UEMs partially compensate this deficiency (Zhang et al., 1998). This approach is taken up by WDP (Conte et al., 2009). It extends the widely used heuristics of Nielsen (Molich & Nielsen, 1990) by three perspectives that are particularly tailored for the characteristics of web applications:

- Conceptual perspective focuses on elements characterizing the application domain
- Presentation perspective focuses on the layout and structure of the UI
- Navigation perspective focuses on the access to and alignment of navigation functions

Each heuristic is viewed through at least one of these perspectives. Additional hints support non-experts in the evaluation. In our study, we carried out a comparative experiment on a real-world web application developed by Fraunhofer IESE, called IND²UCE¹. Our goal was to evaluate whether developers of a particular application can achieve similar results with WDP than usability experts achieve by applying Nielsen's heuristics.

2 Methodology and Results

In order to evaluate the effectiveness and efficiency of WDP, we follow the process shown in Figure 1. Three groups were formed to evaluate a predefined subset of IND²UCE. The first group consisted of six software developers with low knowledge in the field of usability. The second group consisted of four students with almost no usability experience. The third group consisted of three usability experts with high experience.

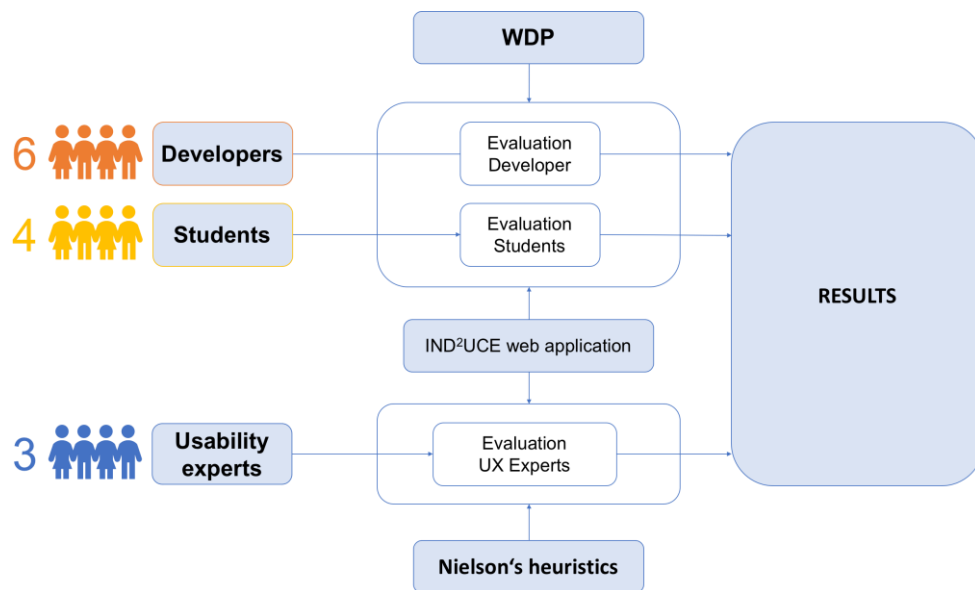


Figure 1- Evaluation Process

Half of the developers and half of the students had prior knowledge about the development of IND²UCE—however, this showed no impact on the results. The developers and students

¹ <https://www.ind2uce.de>

applied WDP, while the usability experts applied Nielsen's heuristics. As Nielsen's heuristics are not meant for UX-laics, we did not analyze how developers and students would perform using these heuristics. All participants documented potential usability issues with the help of a browser extension². Afterwards, these findings were assessed for their severity by a usability expert and a developer of the IND²UCE team (Nielsen, 1995).

To compare the findings of the three groups objectively, we analyzed six UEM performance measures as suggested by (Hartson et al., 2003):

First, *thoroughness* (Sears, 1997) was calculated by dividing the number of problems found by each participant by number of distinct usability problems found by all participants. The mean of the results for each group showed a value of 20% for the developer group, one of 10% for the student group and one of 13% for the usability expert group.

Second, *validity* (Sears, 1997) was calculated by dividing each participant's number of potential issues with a severity rating greater than zero by the number of all problems found by this participant. The higher the validity, the more problems were rated as real usability problems. All three groups achieved values around 90%.

Third, *reliability* (Sears, 1997) was calculated as: "1 - coefficient of variation". Thus, the higher the reliability, the lower is the level of dispersion around the mean of real problems found by a group. Developers and usability experts achieved values slightly above 40%. Students only achieved a value of 24%.

Fourth, we calculated the *number of usability issues found by each participant per minute*.

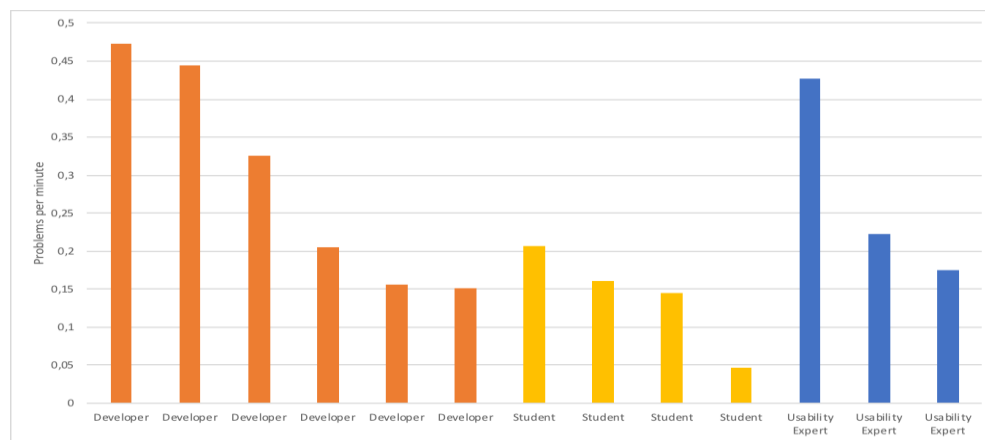


Figure 2 – Number of Usability Issues found by each Participant per Minute

² <https://addons.mozilla.org/de/firefox/addon/wdp-tool/>

The mean of the results shown in Figure 2 has a value of 0.29 issues per minute for the developers, 0.14 issues per minute for the students and 0.28 issues per minute for usability experts.

Fifth, we analyzed the *distribution* of severity ratings. The groups of the developers and usability experts had rather similar distributions. Cosmetic problems made up the largest share for both of them with nearly 40%. The student group achieved slightly better results with a higher proportion of detected minor usability issues.

Sixth, we analyzed the *feedback of the participants* regarding their intensity of usage of the heuristics and perspectives during the evaluation sessions. Around 80% of the developers and students perceived the perspectives as helpful. However, both groups perceived the heuristics as slightly more helpful than the perspectives.

3 Conclusion

The experiment was designed to investigate the performance of WDP in comparison to Nielsen's heuristics by using recognized and comparable measurements.

We were able to show that the results of the developers using WDP were at least equal in all metrics to those of the usability experts using Nielsen's heuristics and their theoretical knowledge about human perception and cognition. The thoroughness value of the developer group was even higher than the one of the experts, i.e., the developers found more usability issues. However, the developers also spend on average ten minutes more time on the evaluation. Therefore, the average number of usability issues found per minute was similar for both groups. Moreover, the developers achieved results equivalent to those of the experts, which is reflected in an almost equal distribution of severity ratings and a low number of false positive classifications (validity).

During an interview after the experiment, usability experts mentioned concerns regarding the tool, as it limits their freedom in evaluation and reporting. This could have led to less motivation, which poses a threat to validity and has to be considered in future evaluations. Developers did not mention any concerns regarding the tool, but positively accentuate its structure.

In contrast to the developers, students using the WDP method generally achieved worse results in terms of quantity than usability experts. However, the variance in the student group was quite high. Our relatively small sample size strongly weighted the negative deviation of one student. This might partially explain this observation. As part of our future work, we will repeat the experiment with a larger sample size in order to mitigate this threat.

We deduce that WDP applied by developers (and potentially also similar groups) can be a good alternative for heuristic evaluations by experts. This can be a major benefit from an economic point of view, as access to usability experts is usually more limited and cost-intensive than access to developers.

References

- Andre, T. S., Williges, R. C., & Hartson, H. R. (1999). The effectiveness of usability evaluation methods: Determining the appropriate criteria. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 43, pp. 1090–1094). SAGE Publications Sage CA: Los Angeles, CA. Retrieved from <http://journals.sagepub.com/doi/abs/10.1177/154193129904302007>
- Conte, T., Massolar, J., Mendes, E., & Travassos, G. H. (2009). Web usability inspection technique based on design perspectives. *IET Software*, 3(2), 106. <https://doi.org/10.1049/iet-sen.2008.0021>
- Hartson, H. R., Andre, T. S., & Williges, R. C. (2003). Criteria for evaluating usability evaluation methods. *International Journal of Human-Computer Interaction*, 15(1), 145–181.
- Molich, R., & Nielsen, J. (1990). Improving a Human-Computer Dialogue. *Communications of the ACM*, 33(3), 11.
- Nielsen, J. (1992). Finding usability problems through heuristic evaluation. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 373–380). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=142834>
- Nielsen, J. (1995, January 1). Severity Ratings for Usability Problems: Article by Jakob Nielsen. Retrieved May 25, 2017, from <https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/>
- Quinones, D., Rusu, C., & Roncagliolo, S. (2014). Redefining Usability Heuristics for Transactional Web Applications (pp. 260–265). IEEE. <https://doi.org/10.1109/ITNG.2014.46>
- Sears, A. (1997). Heuristic Walkthroughs: Finding the Problems Without the Noise. *International Journal of Human-Computer Interaction*, 9(3), 213–234. https://doi.org/10.1207/s15327590ijhc0903_2
- Zhang, Z., Basili, V., & Shneiderman, B. (1998). An empirical study of perspective-based usability inspection. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 42, pp. 1346–1350). SAGE Publications Sage CA: Los Angeles, CA. Retrieved from <http://journals.sagepub.com/doi/abs/10.1177/154193129804201904>