Journal of Biomedical Informatics 44 (2011) 700–701

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Commentary

Methodological concerns in usability evaluation of software prototypes

R. Khajouei [a,b,*], L. Ahmadian [a,b], M.W.M. Jaspers [a]

[a] Department of Medical Informatics, Academic Medical Center - University of Amsterdam, The Netherlands
[b] Kerman University of Medical Sciences, Kerman, Iran

Karahoca et al. [1] addressed an important topic in their paper concerning the use of usability evaluation methods to choose an appropriate software prototype for tablet personal computers. They applied a combination of usability evaluation methods to evaluate the usability of two software prototypes with different graphical user interfaces (GUIs); iconic and non-iconic. These GUIs were designed to replace the paper-based forms at an emergency department of a Turkish hospital. The whole healthcare staff of the department, consisting of 6 physicians and 32 nurses, participated as evaluators in the study. The findings of the comprehensive evaluation showed that the iconic GUI prototype had better usability than the non-iconic GUI prototype. This study contributes to the body of knowledge concerning the usability of GUIs. Such studies are important from a practical perspective because in the competitive market of clinical software they help health care organizations in selecting systems that best suits their users' needs. Moreover, the results of these studies provide practical input to software developers concerning the design of software that is easy to use and that fits the workflow of healthcare providers.

Although Karahoca and his colleagues applied a comprehensive method and obtained interesting findings, we would like to draw attention to some methodological issues concerning the way the usability evaluation methods were employed in this study.

The authors mention that Hom [2] identifies three types of usability evaluation methods, which include testing, inspection and inquiry and state that they applied heuristic evaluation and cognitive walkthrough (CW), both expert inspection methods, to evaluate the usability of the two prototypes.

1. Recruitment of evaluators for heuristic evaluation

The authors recruited potential users of the prototypes as heuristic evaluators. Based on a computer literacy test, half of these users were classified as novice users. In a study identifying the factors affecting heuristic expertise and levels of expertise permissible to conduct a heuristic evaluation, according to Kirmani [3] three factors (usability experience, experience with the heuristic evaluation, and heuristic training) significantly affect the outcomes of heuristic evaluation. This study showed that domain expertise does not have a large impact on the outcomes. Nielsen [4] likewise showed that the performance of novice heuristic evaluators, having general computer knowledge but no special usability expertise, was fairly poor compared to the performance of evaluators with usability expertise. Novice evaluators must first become knowledgeable of and proficient in applying heuristics [5]. Therefore, the validity of the heuristic evaluation results of Karahoca et al. using "usability novices" of which half was computer illiterate can be disputed.

2. Recruitment of evaluators for cognitive walkthrough

CW is a usability inspection method that evaluates the ease with which a typical new user can successfully learn to perform a task using a given interface design. As also stressed by Hom, in CW either usability specialists or software developers should examine the behavior of the interface [6]. In the study performed by the authors CW was again carried out by end users and not by usability experts. In usability inspection methods such as CW, experts evaluate a user interface without involving users. This is in contrast to usability testing where evaluators let users work with the system while recording the user sessions for later analysis of usability problems.

The approach that the authors followed to assess the learnability of the prototypes by real users is not in agreement with the CW method. In CW, inspectors should know the interface before applying the method and then speculate about the ease with which a novice user can learn how to use the system taking user background knowledge such as computer literacy into account.

3. Usability problems

This usability evaluation study lacks the detection of usability problems, which is the main goal of every usability evaluation study [4,7]. This lack can affect the results of the study in several ways. (A) Comparing the effectiveness of the prototypes based on scenario completion rates and completion time, without a careful review and analysis of the main usability problems that potentially can affect user interaction and task outcomes, does not seem valid. Users, for example, could have completed a scenario in a shorter time by skipping some none mandatory steps hindering them during the

* Corresponding author at: J1B-115.2, Department of Medical Informatics, Academic Medical Center, PO Box 22700, 1105 AZ Amsterdam, The Netherlands. Fax: +31 20 691 9840.

E-mail addresses: r.khajouei@amc.uva.nl, rkhajouei@kmu.ac.ir, rkhajouei@yahoo.com (R. Khajouei), l.ahmadian@amc.uva.nl (L. Ahmadian), m.w.jaspers@amc.uva.nl (M.W.M. Jaspers).

1532-0464/$ - see front matter
doi:10.1016/j.jbi.2010.09.003

interaction because of certain usability problems. (B) Because of the low number of physicians (6 out of 38 participants) the authors did not carry out an adequate statistical analysis, they only reported the usability evaluation results of the nurses. Not reporting the results of the usability evaluations of the physicians is a limitation of the study. Although the physicians' data could not be statistically analyzed, qualitative analysis of the usability problems encountered by both groups of users in terms of types, severity and number of usability problems would have benefited the comparison of the two prototypes. The physicians might have encountered different types of usability problems during usage of the two prototypes than the nurses. (C) Heuristic evaluation is meant to reveal usability problems due to violations of the heuristic principles, for example, the principles defined by Nielsen [4]. Perceptual evaluation of the two interfaces by users by solely giving scores based on Nielsen's 10 criteria may be more an assessment of the users' satisfaction than performing a heuristic evaluation.

4. Conclusion

In summary, a variety of inspection and testing methods are at the disposal of usability evaluators. The credibility of the results of usability evaluation studies depends on the methods used in a certain context, the kind of expertise used to apply each method, and the proper implementation of these methods.

Conflict of interest statement

None.

Acknowledgment

The authors would like to thank Arie Hasman for the revision of the manuscript.

References

[1] Karahoca A, Bayraktar E, Tatoglu E, Karahoca D. Information system design for a hospital emergency department: a usability analysis of software prototypes. J Biomed Inform 2010;43(2):224–32.
[2] Hom J. The Usability Methods Toolbox., 1998 [cited 2010 May 25]; Available from: URL: http://jthom.best.vwh.net/usability/.
[3] Kirmani S. Heuristic evaluation quality score (HEQS): defining heuristic expertise. J Usability Stud 2008;4(1):49–59.
[4] Nielsen J. Usability engineering. Boston: Academic Press; 1993.
[5] Jaspers MW. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009;78(5):340–53.
[6] Wharton K, Bradford J, Jeffries R, Franzke M. Applying cognitive walkthroughs to more complex user interfaces: experiences, issues, and recommendations. CHI '92 1992:381–8.
[7] Liljegren E. Usability in a medical technology context assessment of methods for usability evaluation of medical equipment. Int J Ind Ergon 2006;36:345–52.