www.irbas.academyirmbr.com March 2016
International Review of Basic and Applied Sciences Vol. 4 Issue.3

The Influence of End-users on Non-Expert Evaluator Output within the Heuristic Usability Method

JEHAD AlQURNI
School of Mathematical & Computer Sciences
Heriot-Watt University; United Kingdom
Email: jsa2040@hotmail.com

ROBERT POOLEY
School of Mathematical & Computer Sciences
Heriot-Watt University; United Kingdom
Email: r.j.pooley@hw.ac.uk

Abstract

Usability is one of the main and core concepts for fulfilling the quality requirements of any software product, particularly web-based systems. A widely used method to measure the level of usability of software applications is heuristic evaluation (HE). In this method, experts commonly use Nielsen's heuristics to assess the usability of a software product. However, it can also be used by non-experts. Although experts yield better outcomes than non-experts, the former are difficult to obtain and recruit, unlike the latter. In order to improve the effectiveness of heuristic evaluation, several studies have sought to develop its heuristics and procedures. However, there are few studies that involve the end-user within this method, and to the best of our knowledge there are no studies involving end-users with non-expert evaluators within heuristic evaluation method procedures. Therefore, the aim of this study is to investigate the impact of end-users on non-expert evaluator results within the heuristic evaluation process, and through that, to explore the number of usability problems and their severity. The research proposed two sessions: a user exploration session (UES-HE) and a user review session (URS-HE). The former session aims to help the non-expert evaluator to understand the end-user, while the latter session tries to help the non-expert evaluator to review his results through communication with the end-user. One group took part in each session and the outcomes were compared with the traditional HE method as a benchmark. The outcomes show that the end-user has a significant impact on non-expert evaluator results in both sessions through the number and type of usability problems discovered. In the UES-HE method, the results outperformed both URS-HE and HE in terms of usability problems discovered, and it tended to find more major, minor and cosmetic problems than the URS-HE and HE methods. In the URS-HE method, the results outperformed HE but not UES-HE, and tended to find more major problem problems than the HE method.

Key Words: *Usability, Inspection Usability, Heuristic Evaluation, Non-expert Evaluator, User Session.*

Introduction

A revelation in technologies has led to a significant spread of system products and thus increased the demand for system product development. One of the most popular types of products is web-based systems, which play an important role in providing information and services for users. Thus, improving the quality of the user interface of web-based systems affects the interaction between the user and the website, where usability is a critical part of the website quality and it plays a significant role in fulfilling users' satisfaction.

Chen (2012) states that the success or failure of a product is significantly affected by usability. Nielsen (2001) also reports that the reason behind the abandonment of 50% of sales websites is poor website usability. In addition, commercial websites may suffer from difficulties in the competitive environment due to poor usability (Osterbauer, Kφhle, Grechenig, & Tscheligi, 2000). With this in mind, it is imperative that attention be paid to the design of websites, as any websites found wanting when it comes to ease of use may be adversely affected. For example, nearly 39% of online buyers have failed to accomplish their purchases online due to difficulty of use (S. Y. Chen & Macredie, 2005). Consequently, it is clear that bad or good usability plays a pivotal role in the success or failure of website products.

The broadly used definition of usability is based on a definition by the International Organization for Standardization (ISO) "the extent to which the product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" (ISO, 1998). The goal of using the usability evaluation is to ensure the quality requirements of any software product. Thus, usability evaluation methods can be classified into three categories depending on the role of the evaluators: testing, inspection, inquiry, and metrics-based (Zhijun, 2007). Among usability evaluation methods, HE and usability testing are considered to be the most widely employed usability methods, according to a survey conducted by (Sherman, 2009). The main difference between them is that HE depends on evaluator expertise to predict potential usability problems, while usability testing depends on observing end-users to find usability problems.

However, there is still no agreement about which is the best technique, as none of them can find all the usability problems and each of them has its advantages and disadvantages. With regard to usability testing methods, the primary advantage involves recruiting potential end-users and their primary disadvantage is that they are expensive and time-consuming in terms of a laboratory room, testing resources and analysing the results. As for the heuristic evaluation (HE) method, the primary advantages are that it is less expensive, and it requires fewer resources than other methods, and is therefore less time-consuming to conduct than usability testing. However, the primary disadvantage is that it does not involve end-users and it is a subjective assessment (different evaluators can produce different results) and it depends on the evaluators' experience. Conducting both methods will offer better results, but will be costly. Therefore, it is important to improve the effectiveness of one of these methods and practically this will be HE. As Baker, Greenberg, and Gutwin (2001) stated, "Heuristic evaluation (HE) is a widely accepted discount evaluation method for diagnosing potential usability problems in user interfaces" and "HE is popular with both researchers and industry".

Heuristic evaluation is a usability inspection method and it requires a small set of evaluators to use a set of usability principles (called "Heuristics"), as presented in Table 1, to examine the user interface in order to predict potential usability problems.

Table 1: Nielsen Heuristics (Nielsen, 1994a)

1. Visibility of system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and standards
5. Error prevention
6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Help users recognise, diagnose, and recover from errors
10. Help and documentation

Evaluator expertise is the most important factor among various factors affecting the HE method and its results. In an early study, Nielsen (1992) classified evaluators according to their expertise as follows:

- Novice/non-expert evaluators, who lack experience both in usability and in the system domain but who have a solid background in computer science.
- Regular usability specialists, who have expertise in usability but not in the domain of the interface.
- Double usability specialists, who have experience in both usability and the domain of the interface.

In addition, Nielsen stated that regular usability specialists are 1.8 times better than non-expert evaluators, whereas double usability specialists are 2.7 times better than non-expert evaluators, and 1.5 times better than regular usability specialists. Although expert evaluators give better results, they are difficult to find and may not be available. Therefore, some studies have investigated the effectiveness of non-experts (Äijö & Mantere, 2001; Botella, Alarcon, & Peñalver, 2013; Fernandes, Conte, & Bonif'cio, 2012; Howarth, Smith-Jackson, & Hartson, 2009; Koutsabasis, Spyrou, Darzentas, & Darzentas, 2007; Slavkovic & Cross, 1999).

In order to improve HE effectiveness, several studies have examined the major factors of HE such as lists of heuristics or evaluator expertise. Expertise is one of the most important factors contributing to the improvement of the HE (Hwang & Salvendy, 2007). Heuristic evaluation can be performed by expert or by non-expert evaluators. Although the former yield better results (Nielsen, 1992), several studies indicate that expert evaluators are difficult to find (Äijö & Mantere, 2001; Desurvire & Thomas, 1993; Nielsen, 1999; Paz, Paz, Villanueva, & Pow-Sang, 2015). With regard to the role of the user in the HE method, the majority of studies on the HE method depend on Nielsen's (1993) argument that HE is not performed by end-users but by evaluator experts or non-experts. In contrast to Nielsen's view, Muller, Matheson, Page, and Gallup (1998) developed HE by involving users as 'web-domain experts' to be a part of the evaluation team, and the technique then becomes a Participatory Heuristic Evaluation (PHE), by combining experts with users. In addition, this technique extended the heuristics of HE by adding new heuristics to well-known Nielsen heuristics. The authors claim that if users are easy to recruit, then PHE can be as cost-effective as traditional HE. However, users in this method are described as 'web-domain experts' and that is different from a real end-user. In addition, this method still requires expert evaluators on the PHE team.

In light of this, it is necessary to continue research in order to study and address the usability topic, with emphasis on the exploration of the key factors for usability. Few studies have investigated the role of end-users in such a method. Until now, to the best of our knowledge, there have been no studies with a focus on the influence of users on non-expert evaluator output within the HE method. Thus, the main aim of this research is to investigate the impact of end-users on non-expert evaluator output in terms of the number of usability problems and their severity, in cases where users are easy to recruit and expert evaluators are difficult to find.

Methods

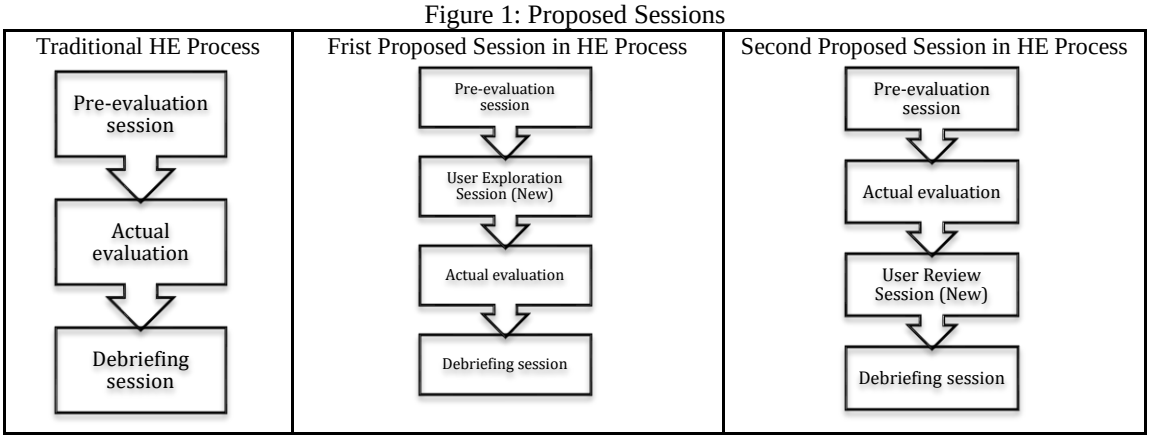
The main aim of this study is to investigate the impact of end-users on non-expert evaluator results through the HE method and this impact on the results will be clearly seen through the number of usability problems discovered. To achieve this aim, two user sessions were proposed to examine the influence of end-users on non-expert evaluator results, as illustrated in Figure 1:

User Exploratory Session (UES-HE): The non-expert evaluators used a "question-asking protocol" with the end-users in order to understand whether they had difficulties with understanding and using the system. In this session the users performed tasks and the evaluator tried to encourage them to complete the tasks with direct questions, while noting any issues. In addition, the evaluator tried to ask questions about the task in more depth in order to understand any issues that the task did not cover.

User Review Session (URS-HE): In order to minimise potential problems that the user does not currently face, or to modify existing issues, non-expert evaluators reviewed the potential usability problems list with the end-user by using a "free discussion protocol". The evaluator was not restricted to gathering the users'

impressions about every usability issue, but they were able to expand the review to discuss other usability issues that may have emerged during the review, in order to add new issues.

Figure 1: Proposed Sessions



Three groups were recruited; each of them consisted of eight non-expert evaluators, with two groups representing the two proposed user sessions (UES & URS) and the third group representing the traditional HE method. In order to examine end-user impact, eight end-users were involved in each of the first two groups, and all eight evaluators in the last group were non-expert evaluators.

There is no specific procedure for performing HE. This research adopts Nielsen's (1994b) approach recommendation for conducting HE. These processes can be summarised as three phases. The first phase is a pre-evaluation (training) session, in order to familiarise the evaluator with the HE process. Then in the actual evaluation session, each evaluator inspects individually the interface and identifies the predicted usability problems, according to a short list of heuristic guidelines as shown in Table 1. Finally, the individual results of each evaluator are integrated into a single list during a debriefing session. Each usability problem is classified depending on its severity rating. Nielsen (1994b) recommended using severity ratings in order to estimate the severity of each problem as follows:

1: Cosmetic problem only – need not be fixed unless extra time is available.
2: Minor usability problem – fixing this should be given low priority.
3: Major usability problem – important to fix, so should be given high priority.
4: Usability catastrophe – imperative to fix this before product can be released.

Results

The results derived from the UES-HE, URS-HE and HE methods were compared in terms of the numbers of problems and their types. Table 2 shows that the UES-HE method discovered 50 usability problems, which is equivalent to 77% of the total usability problems discovered by all three methods, thus outperforming the other two methods. The URS-HE method uncovered 35 usability problems, which is equivalent to 54% of the total usability problems discovered. However, the HE method discovered only 23 usability problems, with equates to 35% of the total problems.

Table 2: Usability methods and discovered usability problems

	UES-HE method	URS-HE method	HE method
No. of discovered problems	50	35	23
Percentage of discovered problems	77%	54%	35%

In terms of the severity rating for each usability problem, Table 3 illustrates the number of cosmetic, minor, major and catastrophic usability problems for all three methods. The UES-HE method tended to find more minor and cosmetic problems compared to the other two methods, and it was also able to find the same number of major problems as those found by the URS-HE and HE methods together. The URS-HE method is the second best method, and the HE method delivered the worst result.

Table 3: Usability methods and severity rating of usability problems

	UES-HE method	URS-HE method	HE method
1:Cosmetic	15	10	8
2 Minor	13	9	7
3:Major	17	12	5
4:Catastrophic	5	4	3
Total	50	35	23

Conclusion

Heuristic evaluation is a widely accepted inspection method for diagnosing potential usability problems, and it can be conducted by expert and non-expert evaluators. Although the former type of evaluator yields better results than the second type of evaluator, they are hard to find. In order to improve the results of non-expert evaluators, this research has proposed two user sessions within the HE process. The outcome of this paper shows that non-expert evaluators within the HE method should be encouraged to include end-users during their evaluation activity, because it has been proven to be more effective than the traditional method. Involving end-users can play a vital role in the results in both proposed methods (UES-HE & URS-HE). This preliminary result will be followed by further research in order to compare the results of the methods in this research and the user testing method. In addition, these results suggest that further experiments should be conducted in the future by combining both user sessions in one method for more investigation.

Acknowledgements

This work is supported by the University of Dammam, the Ministry of Education-Higher Education, Kingdom of Saudi Arabia.

References

Äijö, R., & Mantere, J. (2001). *Are non-expert usability evaluations valuable?* Paper presented at the 18th International Symposium on Human Factors in Telecommunications (HfT 2001), Bergen, Norway.

Baker, K., Greenberg, S., & Gutwin, C. (2001). Heuristic Evaluation of Groupware Based on the Mechanics of Collaboration. In M. Little & L. Nigay (Eds.), *Engineering for Human-Computer Interaction* (Vol. 2254, pp. 123-139): Springer Berlin Heidelberg.

Botella, F., Alarcon, E., & Peñalver, A. (2013). *A new proposal for improving heuristic evaluation reports performed by novice evaluators*. Paper presented at the Proceedings of the 2013 Chilean Conference on Human - Computer Interaction, Temuco, Chile.

Chen, C. (2012). A usability inspection expert system based on HE, GRY and GST. *International Journal of Intelligent Information Processing, 3*(1), 1-15.

Chen, S. Y., & Macredie, R. D. (2005). The assessment of usability of electronic shopping: A heuristic evaluation. *International Journal of Information Management, 25*(6), 516-532. doi:10.1016/j.ijinfomgt.2005.08.008

Desurvire, H., & Thomas, J. C. (1993). *Enhancing the performance of interface evaluators using non-empirical usability methods.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Fernandes, P., Conte, T., & Bonif'cio, B. (2012). WE-QT: A Web Usability Inspection Technique to Support Novice Inspectors. 11-20. doi:10.1109/sbes.2012.30

Howarth, J., Smith-Jackson, T., & Hartson, R. (2009). Supporting novice usability practitioners with usability engineering tools. *International Journal of Human-Computer Studies, 67*(6), 533-549. doi:http://dx.doi.org/10.1016/j.ijhcs.2009.02.003

Hwang, W., & Salvendy, G. (2007). What makes evaluators to find more usability problems?: A meta-analysis for individual detection rates. *Human-Computer Interaction, Pt 1, Proceedings, 4550*, 499-507.

ISO. (1998). ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability: ISO.

Koutsabasis, P., Spyrou, T., Darzentas, J. S., & Darzentas, J. (2007). On the performance of novice evaluators in usability evaluations. *Proc. PCI, Patra*.

Muller, M. J., Matheson, L., Page, C., & Gallup, R. (1998). Methods & tools: participatory heuristic evaluation. *interactions, 5*(5), 13-18.

Nielsen, J. (1992). *Finding usability problems through heuristic evaluation*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Monterey, California, USA.

Nielsen, J. (1993). *Usability Engineering*: Morgan Kaufmann Publishers Inc.

Nielsen, J. (1994a). *Enhancing the explanatory power of usability heuristics*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, Massachusetts, USA.

Nielsen, J. (1994b). Heuristic evaluation. In N. Jakob & L. M. Robert (Eds.), *Usability inspection methods* (pp. 25-62): John Wiley & Sons, Inc.

Nielsen, J. (1999). User interface directions for the Web. *Commun. ACM, 42*(1), 65-72. doi:10.1145/291469.291470

NIELSEN, J. (2001). Did Poor Usability Kill E-Commerce? Retrieved from http://www.nngroup.com/articles/did-poor-usability-kill-e-commerce/

Osterbauer, C., Kφhle, M., Grechenig, T., & Tscheligi, M. (2000). Web Usability Testing: A case study of usability testing of chosen sites (banks, daily newspapers, insurances). *the Sixth Australian World Wide Web Conference*.

Paz, F., Paz, F. A., Villanueva, D., & Pow-Sang, J. A. (2015, 13-15 April 2015). *Heuristic Evaluation as a Complement to Usability Testing: A Case Study in Web Domain.* Paper presented at the Information Technology - New Generations (ITNG), 2015 12th International Conference on.

Sherman, P. (2009). UPA Salary Survey. *Usability Professionals' Association, Illinois, US*.

Slavkovic, A., & Cross, K. (1999). *Novice heuristic evaluations of a complex interface*. Paper presented at the CHI '99 Extended Abstracts on Human Factors in Computing Systems, Pittsburgh, Pennsylvania.

Zhijun, Z. (2007). Usability Evalutation. In Z. Panayiotis & K. Sri (Eds.), *Human Computer Interaction Research in Web Design and Evaluation* (pp. 209-228). Hershey, PA, USA: IGI Global.