Usability evaluation: an investigation on combination of analytical and empirical methods

Setia Hermawati, Glyn Lawson

Human Factors Research Group, University of Nottingham

SUMMARY

This study investigated the effect of combining analytical (heuristics and cognitive walkthrough) and empirical methods in usability evaluation. Data from two usability studies were used to simulate the outcomes of different combinations of usability evaluation methods. The findings show that the combined analytical methods significantly reduce the number of participants required in the empirical method without compromising the results of the usability evaluation.

KEYWORDS

Usability evaluation, Heuristics evaluation, Cognitive walkthrough, User testing

While empirical evaluation (user testing) is desired as part of usability evaluation, it is often costly and cumbersome to conduct (Nielsen, 1993) because they require the need to recruit test participants that match the target user for a system. Analytical methods, also called discounted methods, were developed to reduce the costs of usability evaluation associated with empirical methods and involve the participation of experts in human factors. Heuristic evaluation is cheap, fast and able to predict major usability problems that could potentially occur during usability testing (Jeffries et al., 1991; Tang et al., 2006; Hwang and Salvendy, 2010). However, it is also reported to often discover low-priority usability problems, and its output is largely dependent on the quality of the evaluators involved (Jeffries and Desurvire, 1992; Hwang and Salvendy, 2010). Another analytical method of interest is cognitive walkthrough. It is costlier than heuristic evaluation and requires extensive knowledge in cognitive psychology (Bias and Mayhew, 2005; Hwang and Salvendy, 2010). Although it is more effective in finding severe problems (Sears, 1997), it can reveal only about a third of the usability problems detected by a heuristic evaluation (Jeffries et al., 1991).

Several studies have compared the performance of analytical and empirical methods (see, e.g., Karat et al., 1992; Ahmed, 2005; Tan et al., 2009; Thyvalikakath et al., 2009; Petri and Power, 2012). Although general findings suggest that analytical methods can identify usability problems that severely affect interaction with a system, most studies found that it is very unlikely that analytical methods alone can identify all severe usability problems. There is an argument that the analytical methods should not be used to justify the omission of the empirical methods as part of usability evaluation (Jeffries and Desurvire (1992) and that a combination of empirical methods and analytical methods should be adopted (Ahmed, 2005); Tan et al., 2009).

Unfortunately, there has not been a study exploring how combinations of analytical and empirical methods affect the discovery rate of usability problems. This study was aimed to fill this gap and sought to provide evidence of the benefit of combining usability evaluation methods. To achieve this aim, we conducted a simulation study that was based on data from actual usability studies to

fully investigate the interaction among the different usability evaluation methods with respect to the number of unique usability issues that could be identified and the risk of missing severe usability issues.

We conducted two independent usability evaluation studies on two different software. The software was designed for educational/training purposes with the first software aimed at trainers of assembly line operators in a manufacturing setting, and the other one was aimed at students in a higher education setting. Two different and yet similar themed software were intentionally used as it would allow, to some extent, generalisation of the outcome of this study. In each study, three types of usability evaluation methods (heuristic, cognitive, and empirical) were conducted, and five participants were assigned in each evaluation method. The heuristics and cognitive walkthrough involved participants who completed the graduate or post-graduate level of coursework's in Human Factors. In the empirical evaluation, participants consisted of trainers of assembly line operators and students at the Nottingham University.

For each study, the severity of each usability issue was identified. Next, the usability issues were grouped and coded to remove redundancies. The coded usability issues were then assigned to each participant. After this step was completed, a simulation of hypothetical groups that represented variations of all participants and number of participants in each method was performed. For each study, a total of 150 groups were created, combining different numbers of participants in each evaluation method. In each group, subgroups were then created to reflect different combinations of participants. The creation of subgroups ensured that the simulation considered differences in the performance of the participants in identifying the usability problem. Between 5 to 1000 subgroups were created in each group. The simulation was achieved by creating a programming code in MATLAB.

The findings of this study showed that, in comparison to heuristic method, cognitive walkthrough identified more unique usability problems and resulted in a better prediction of usability problems that would be encountered by end-users in empirical method. Furthermore, we also found that the combination of cognitive and heuristic methods identified 98.2% of known usability problems, compared to 44.8% and 68.9% by heuristic and cognitive methods, respectively. Although this finding is promising and suggests the potential of combined analytical methods, this study also found that there was still a risk of missing severe usability issues when usability evaluation relied solely on analytical evaluation methods. Regarding the question 'how many participants are required in the empirical method to compensate for combined analytical methods', our study showed that the participation of just one participant in the empirical method complemented combined analytical methods and successfully reduced the risk of missing severe usability problems to less than 75%.

The results of this study also revealed that, at least nine participants (4 and 5 usability experts to conduct heuristics and cognitive walkthrough, respectively) were required in combined analytical methods to identify 85% of unique usability problems. This finding disputed assertions that about 85% of known usability problems could be identified by 5 participants. This study also found that the diminishing return relationship between the number of participants in empirical method and the number of usability issues identified was also applicable to combined analytical methods. This suggests that Nielsen and Landauer's (1993) equation that illustrates the relationship between the discovery of usability problems and the number of participants could also be used to estimate the total number of participants required in combined analytical methods.

References

Ahmed, S. M. Z. (2008) A comparison of usability techniques for evaluating information retrieval system interfaces. Performance Measurement and Metrics. 9(1), 48-58.

Bias, R. G., & Mayhew, D. J. (2005) Cost-justifying usability: an update for the internet age. San Francisco: Morgan Kaufman.

Hwang, W., & Salvendy, G. (2010) Number of people required for usability evaluation: the 10±2 rule. Communications of the ACM, 53(5), 130-133.

Jeffries, R., & Desurvire, H. (1992) Usability testing vs. heuristic evaluation: was there a contest. ACM SIGCHI Bulletin, 24(4), 39-41.

Jeffries, R., Miller, J.R., Wharton, C. & Uyeda, K. (1991) User interface evaluation in the real world: a comparison of four techniques. In Proceedings of the SIGCHI conference on Human Factors in computing systems: Reaching through technology (pp. 119-124).

Karat, C-M., Campbell, R., & Fiegel, T. (1992) Comparison of empiric testing and walkthrough methods in user interface evaluation. In Proceedings of CHI 1992 (pp. 397-404).

Nielsen, J. (1993) Usability Engineering. Cambridge, MA: Academic Press.

Petri, H. & Power, C. (2012) What do users really care about? A comparison of usability problems found by users and expert on highly interactive websites. In Proceedings of CHI 2012 (pp. 2107-2116).

Sears, A. (1997) Heuristic walkthroughs: finding the problems without the noise. International Journal of Human-Computer Studies Interaction, 9(3), 213–234.

Tan, W-s., Liu, D., & Bishu, R. (2009) Web evaluation: heuristic evaluation vs. user testing. International Journal of Industrial Ergonomics, 39, 621-627.

Tang, Z., Johnson, T.R., Tindall, R.D., & Zhang, J. (2006) Applying heuristic evaluation to improve the usability of a telemedicine system. Telemed J E Health, 12(1), 24–34.

Thyvalikakath, T. P., Monaco, V. M., Thambuganipalle, H., & Schleyer, T. (2009) Comparative study of heuristic evaluation and usability testing methods. Studies in Health Technology and Informatics, 143, 322-327.