

A Framework for Integrating Usability Evaluations Methods: The Mawhiba Web Portal Case Study

Areej Al-Wabil and Hend Al-Khalifa

College of Computer and Information Sciences, King Saud University

areej@acm.org, hend@ccis.edu.sa

Abstract

Different usability evaluation methods (UEMs) are often combined to assess the compliance of interactive systems with usability standards and guidelines. In practice, combinations of UEMs have been applied in a rather ad hoc manner resulting in inadequate analyses of rich usability data sets. While the potential offered by recent advances in technologies such as eye tracking for usability is recognized, and the challenges of evaluating complex web systems such as portals which involve different interaction styles are highlighted, the issue of effectively integrating multiple UEMs to make sense of observations has emerged. This paper presents a framework developed for effectively integrating UEMs in usability evaluations by matching the methods' capabilities and limitations with a classification of usability problems. The framework is applied to evaluating the usability of the Mawhiba web portal with target user populations which demonstrates how better coverage of usability issues can be achieved.

1. Introduction

For more than two decades, conducting usability evaluations has been considered a key quality assessment technique in website evaluations. Increasing expectations of easy access to the web has led to the proliferation of research examining and comparing usability inspection methods in Human-Computer Interaction (HCI) literature [4]. Different Usability Evaluation Methods (UEM) have emerged and evolved in research and practice in the field of usability. Nowadays, a wide array of UEMs is available for practitioners to employ depending on the product, context of use, and project constraints.

Competing UEMs have been compared in the usability literature to assess their efficacy in uncovering usability problems in a system. Many authors have

argued for employing different UEMs in usability evaluations so that these methods supplement each other rather than compete (e.g. [6],[8]). The variability of UEM features and limitations poses a challenge for practitioners and researchers in determining how to effectively integrate these tools in usability evaluations. While strengths and limitations of each UEM can guide the planning process to determine how these methods complement each other, they are often applied in an ad hoc manner. Thus, much of the rich usability data sets are inadequately analyzed, the underlying usability issues causing these problems are not understood, and consequently the reported problems lack the quality descriptions needed for effective downstream implementation by developers and designers for improving the usability of the system.

Issues of cultural usability have also emerged in the past decade [2],[14]. Usability inspections often involve standardized forms and surveys that have been developed and validated in predominately Western cultures. Recent studies have shown that usability is perceived differently across cultures and subjective satisfaction ratings may be influenced by the cultural background of users. Therefore, evaluating usability of bi-directional interfaces (e.g. combine Right-to-Left flow of Arabic script and Left-to-Right flow of numbers and Latin script) with end-users of cultures which have rarely been examined further complicates the process. The challenge can be attributed to the lack of validated usability tools specific to the cultural context and limited support from design guidelines that inadequately translate to bi-directional interfaces. Specifically, direct interpretation of findings is not feasible in observations related to content and User Interface (UI) element designs. Thus, combinations of UEMs are essential to obtain better coverage of usability issues and an improved understanding of the underlying causes of these problems, and provide supplemental data to substantiate conclusions drawn from evaluations.

The general lack of understanding of the capabilities and limitations of different UEMs has led to an intensified need for usability practitioners and researchers to determine the most appropriate combination of UEMs and understand how to effectively integrate these methods. A systematic approach is needed for guiding the evaluations, making sense of the findings, and reporting the results. In this paper, we propose a framework to effectively integrate different UEMs by considering the usability problem profile of a system, and apply this approach to a case study of evaluating the usability of an Arabic web portal.

2. A framework for integrating UEMs

The framework for integrating UEMs is depicted in Figure 1. Usability evaluations would be based on a classification of usability problems conducted after an initial expert review (e.g. Heuristic Evaluations) or user based UEM (e.g. lab-based testing) in what we refer to

as defect-tracking inspections. We adopted the usability problem profile of Chattratichart & Lindgaard [1] as a classification scheme to guide the selection of UEMs in usability evaluation projects. For example, problems related to visibility of interface elements would be classified as Graphics & UI objects (#3) and problems related to labeling of links would be classified as Wording & Language problems (#6).

Further examinations of observations that require supplemental data and supporting evidence are determined based on the severity of the problem (impact, persistence) and project constraints. A complete understanding of the underlying causes of usability problems is essential for cost-effective analysis and establishing re-design recommendations. For this reason, further examination in UEMs would provide the practitioner with evidence needed to substantiate recommendations and avoid drawing conclusions that would go beyond what the usability data could support.

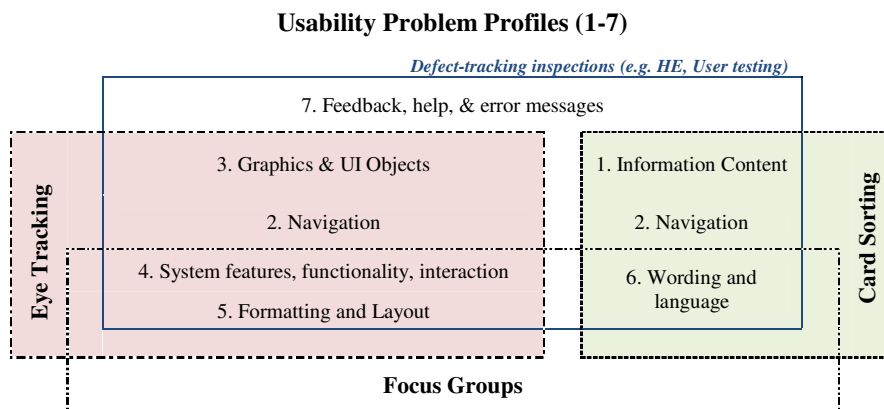


Figure 1 Framework for integrating UEMs using usability problem profiles

3. Case study: Applying the framework in evaluating a web portal

We demonstrate how this framework can be applied to usability evaluations of web portals by presenting a case study that applied this framework to evaluating a web portal with an Arabic interface. The portal evaluated is that of Mawhiba (<http://www.mawhiba.org/>), a non-profit organization aimed at fostering creativity and supporting gifted individuals. The web portal is designed as a gateway to services, information and interactive applications for those involved; namely students, parents, inventors (patent holders), teachers, and specialists in creativity and gifted education.

3.1. Methodology

We first conducted a Heuristic Evaluation (HE) of the portal based on Nielsen's heuristics [11], [12]. The objective was to assess the portal's compliance with usability standards and design guidelines. This was a precursor to launching usability testing with end-users. This approach has been shown to be effective in prior usability research [6], [8]. The usability issues were classified according to the usability problem profile and given severity ratings on a 3-point scale ranging from severe, moderate to minor issues. Following that, scenarios were developed to expose severe usability

issues that were uncovered by our HE and as many of the moderate and minor heuristic violations as possible.

After taking into account considerations related to complexities of the tasks, context of use and session timing, a final list of 20 tasks was developed for use in sessions with end-users. Examples of tasks include registering as a new member on the portal, subscribing to Mawhiba’s newsletter, voting on the opinion polls, and looking up information related to patents.

Figure 2 depicts our approach in applying the framework to the Mawhiba web portal evaluation. First, the user testing was effective in uncovering a wide range of usability problems, which were then classified according to the 7-segment Usability Problem Profile of the framework. Problems of specific classifications were assigned to a specific UEM and further evaluations were considered based on severity of problems and project constraints. For example, problems related to visibility of UI elements were examined further in the eye tracking UEM; problems related to disorientation and understanding how content is organized in some portal segments were examined further in the card sorting UEM; variations in subjective satisfaction ratings were examined further in focus group discussions.

This was needed because we were evaluating an Arabic portal with native users. In this context, the

defect-tracking phase in user testing uncovered key issues but failed to provide substantial data to explain the findings because the usability literature was scarce to substantiate explanations for difficulties that were related to the language, formatting and design of UI elements. The think-aloud protocol incorporated in user testing was helpful in providing an insight to some of the difficulties that users were experiencing but this approach was limited in providing an understanding of why users were experiencing difficulties with some aspects of the interface. The limitations of the think-aloud protocol were evident in two aspects. First, the concepts of usability testing in general and thinkaloud in particular were new to participants in this context. Despite efforts by facilitators to continuously elicit feedback from participants, the cultural background of participants influenced their ability to freely express their opinions or eloquently describe their thoughts and feelings during interaction. Second, some problems were related to disorientation and users were in a confused or frustrated state and were not capable of describing what they were experiencing, seeing, feeling etc. Taken together, these challenges made it essential to combine multiple UEMs to obtain better coverage of usability issues and a deeper understanding of the observed difficulties.

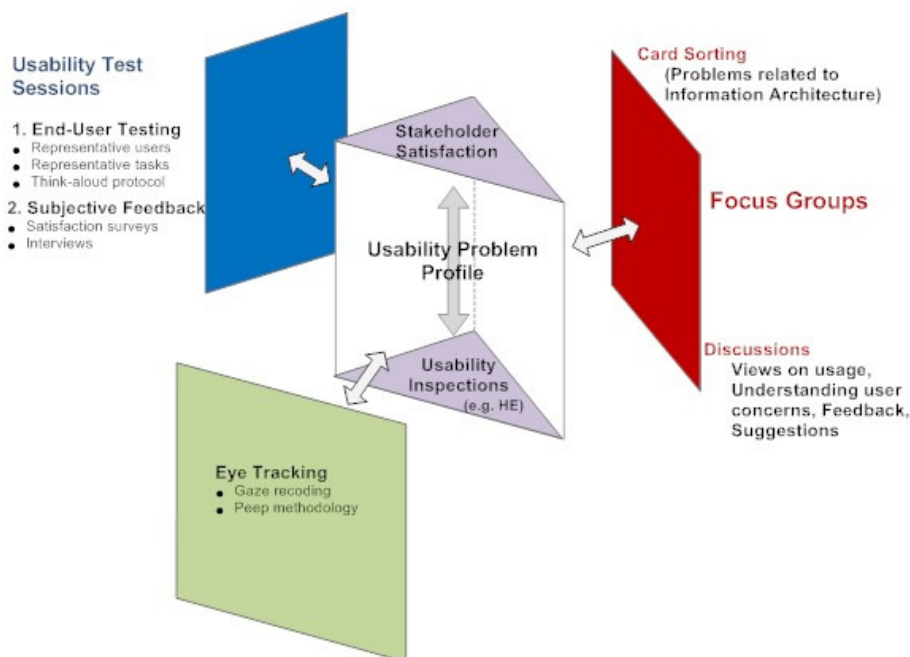


Figure 2 A structured approach for integrating UEMs in evaluating a web portal

3.2. User testing

In the first phase of usability evaluations, systematic testing of the web portal was conducted with representative users sampled from target populations. Interactions with the portal were recorded and analyzed using Morae software. It has been asserted that conducting usability sessions with five users is sufficient to uncover 80-85% of usability problems in an interactive system [13], [16]. While this number is often considered sufficient for each unique segment of end-users to yield acceptable levels of return on investment, recent HCI research has questioned this notion of 5 users and suggested striking the balance between task and user selection [9]. Therefore, in our evaluations, we administered a thorough set of 20 tasks requiring approximately one hour for each user and aimed to sample 10 users in each target group. Table 1. shows the number of users that took part in our usability tests along with the mean age in each group. All but one group were sampled by >5 users (the inventors' group was distributed geographically and only four were able to participate.)

Table 1. User demographics

Target user population		Users	Age
Gifted Students	Middle school level	6	14
	High school level	5	16
	University level	4	19
Inventors		4	19
Parents		10	40
Specialists in gifted education		10	31

This UEM was sufficient to obtain a comprehensive set of usability problems. We then classified these problems according to the Usability Problem Profile. Issues that ranked high on the severity scale or significantly impacted interaction were considered for further examination if the underlying reason for the observation were not clear and recommendations were not directly attainable due to lack of evidence.

Culture, as mentioned in the introduction, has been shown to influence subjective satisfaction ratings in usability evaluations [2]. In the absence of standardized surveys specific to our target populations, we adapted the System Usability Scale (SUS) to the web portal under investigation by translating the statements and scales and asking users to rate the usability of the portal after the session. Ratings, shown in Figure 3, varied considerably amongst participants in the (1) adult group between parent and specialist groups; while this could have been attributed to the variation in technical skills and Internet experience that was measured in the

demographic survey it warranted further examination and (2) the student group between students of varying age groups. We conducted two separate focus groups to examine the different opinions of individuals in these groups to further understand the variation and to substantiate the reporting of user satisfaction with explanations.

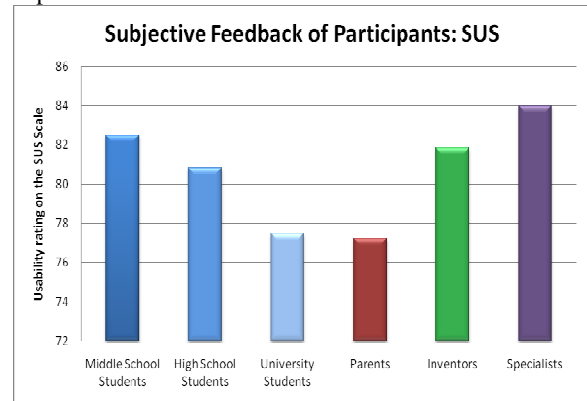


Figure 3 Subjective satisfaction ratings from target user groups

3.3. Focus groups

Focus group sessions were conducted with two groups sampling different demographics, specifically students and adults, to obtain a sense of how opinions related to the portal's design vary. The first session was conducted with 15 students and the second focus group session was conducted with 10 adult participants, 5 parents and 5 specialists. Topics for discussion were related to the design and the envisaged usage of interactive segments of the portal. Findings revealed contrasting views of users in different target populations.

3.4. Card sorting

Card sorting was conducted to examine disorientation problems that suggested that users were not able to understand the structure of the portal, or confused with labelling of portal sections, and the organization of information within the portal. Findings showed low agreement of how users structured the portal's content in some segments with how designers structured the content. This in turn confirmed that disorientation problems, uncovered in user-testing, were due to confusion in wording and structuring of information content. Taken together, recommendations were substantiated with evidence of the problem and its underlying causes.

3.5. Eye tracking

With recent advances in eye tracking technologies, usability practitioners are increasingly having access to eye tracking tools that involve effective (though often expensive in terms of time, effort and cost) methods of gathering observational data of users interacting with systems. However, the added value of eye tracking in usability evaluations has been debated in the HCI domain when considering returns on investment in effort and cost. Our approach in incorporating eye tracking involved focusing our effort on the relevant sub-set of usability issues (that were identified from the Usability Problems Profile to be related to visibility and affordances of UI elements). These problems were selected for further observations with eye tracking, thus reducing the time, cost and effort involved with eye tracking and focusing the analysis on understanding the underlying problem for these usability issues. Four eye tracking sessions were conducted with native Arabic web users using a Tobii x50 eye tracker and Tobii Studio for analysis.



Heatmap (Aggregate viewing patterns)



Clusters of fixations on UI elements

Figure 4 Visibility problems with UI elements

Examples demonstrating how eye tracking supplemented the findings of user testing with evidence essential for understanding the underlying problem and suggesting design solutions accordingly are shown in Figure 4. Eye tracking provided an insight into human

cognition while users interacted with the portal in search of specific information. Prior to eye tracking, our user-testing revealed a possible visibility problem with a key element in the UI (eOasis, highlighted in the rectangular area in Figure 4). Eye tracking provided the evidence demonstrating that this element failed to attract users' attention. Furthermore, in cases in which users fixated on this element, eyetracking revealed that it failed to sustain their attention long enough for users to process and understand its relationship to the target page. The clusters revealed that users expectations was to find this element in the main right-hand-side menu and systematically scanned it from right to left until the decorative image of the navigation button was reached which indicated that options have been exhausted in this region and users consequently glazed over the label and missed the target. The focused analysis of this UEM data enabled us to pinpoint the underlying problem and provide recommendations related to increasing the visibility of this element to attract and sustain attention as well as adjust the design of the navigation element (i.e. placing the textual label on the right in-line with user expectations).

Another example was with competing areas of interest in an interface in which we suspected problems related to distraction while users were searching for the hotline number of the organization; with this problem, the think-aloud protocol was not capable of providing sufficient explanations as to why users could not find the element on the UI. In contrast, eye tracking revealed this behaviour with only sampling ten percent of users.



Figure 5 Visual attention on competing AOIs

4. Related work

Triangulation has been reported in studies of usability evaluations in which multiple UEMs are used to provide supplementary data (e.g. [15]). Previous work has assessed the capabilities and limitations of different UEMs (e.g. [5], [10], [6]); Of particular interest for integrating UEMs is the work of Koutsabasis, Spyrou and Darzentas [7] who conducted

an evaluation of an academic web site using four UEMs, namely heuristic evaluations, cognitive walkthroughs, think-aloud protocols and co-discovery learning. However, the focus in prior work was mainly on comparing the efficacy of the methods rather than examining how combining these methods can complement each other to effectively uncover usability issues in a given system.

The framework presented in this paper contributes to related work by synthesizing the capabilities and limitations of different UEM and usability problem profiles into a structured approach for conducting usability evaluations. This study also presented a case study demonstrating the use of the framework in the evaluation of a web portal which provided insights into the analysis and reporting of usability findings from different UEMs.

5. Conclusion

A framework has been developed to guide usability evaluations of web applications, facilitate effective integration of multiple UEMs, and improve the quality of usability problem reporting. The framework has been successfully applied to evaluate the Mawhiba portal, a web portal with an Arabic interface. Selecting which UEMs to use should be based on considering the alternatives available for the evaluation, goal of the project, and understanding the capabilities and limitations of each method. It is envisaged that the framework adopted in this work can be applied to other web application evaluation studies and inform usability practitioners' selection of appropriate UEMs.

6. Acknowledgements

This study was supported by Mawhiba, the King Abdulaziz Foundation for Creativity and Giftedness in Saudi Arabia. We thank Ms. Reem Al-Zamil for recruiting participants and Mawhiba portal staff for their valuable contributions throughout the project. Special thanks to KSU facilitators; Eihal Al-Owaishiq, Atheer Al-Khalifa, Bashair Alkhwaiter and Dania Al-Omar.

7. References

- [1] Chattratichart, J., Lindgaard, G.: A Comparative Evaluation of Heuristic-Based Usability Inspection Methods. In the Conference on Human Factors in Computing Systems CHI'08, pp. 2213-2220. ACM (2008).
- [2] Frandsen-Thorlacius, O., Hornbaek, K., Hertzum, M., Clemmensen, T.: Non-Universal Usability? A Survey of How Usability is Understood by Chinese and Danish Users. In the Conference on Human Factors in Computing Systems CHI'09, pp. 41-50. ACM, (2009)
- [3] Hartson, H.R., Andre, T.S., Williges, R.C.: Criteria for evaluating usability evaluation methods. *International Journal of Human-Computer Interaction*, 15(1), pp. 145-181 (2003).
- [4] Hollingsed, T., Novick, D.: Usability Inspection Methods after 15 Years of Research and Practice. In SIGDOC'07, pp. 249-255. ACM (2007)
- [5] Jeffries, R., Miller, J., Wharton, C., Uyeda, K.: User Interface Evaluation in the Real World: A Comparison of Four Techniques. In the Conference on Human Factors in Computing Systems, pp. 119-124. ACM, New York (1991).
- [6] Karat, J.: Software evaluation methodologies. In: Helander, M. (Ed.) *The Handbook of Human Computer Interaction*. Elsevier, New York (1988).
- [7] Koutsabasis, P., Spyrou, T., Darzentas, J.: Evaluating Usability Evaluation Methods: Criteria, Method and a Case Study. In: *HCI: Interaction Design and Usability*. LNCS, vol. 4550, pp. 569-578. Springer, Heidelberg (2007)
- [8] Lindgaard, G.: Notions of Thoroughness, Efficiency, and Validity: Are they Valid in HCI practice? In proceedings of the 4th International Cyberspace Conference on Ergonomics. International Ergonomics Association Press, (2005).
- [9] Lindgaard & Chattratichart, J.: Usability Testing: What Have We Overlooked? In the Conference on Human Factors in Computing Systems CHI'07. ACM, (2007).
- [10] Mack, R., Nielsen, J.: Usability Inspection Methods. In: Baecker, R., Grudin, J., Buxton, W., Greenberg, S. (eds.) *Readings in Human-Computer Interaction: Towards the year 2000*. Morgan Kaufmann Publishers, San Francisco (1995).
- [11] Nielsen, J.: Heuristic Evaluation. (1993) available at: http://www.useit.com/papers/heuristic/heuristic_evaluation.html
- [12] Nielsen, J., Molich, R.: Heuristic evaluation of user interfaces. In the Conference on Human Factors in Computing Systems CHI'90, pp.249-256. ACM (1990)
- [13] Nielsen, J. Why You Only Need to Test with 5 Users. Jakob Nielsen's Alertbox, March 19, 2000. Available online at <http://www.useit.com/alertbox/20000319.html> (2000)
- [14] Smith, A., Yetim, F.: Global Human-Computer systems: Cultural determinants of Usability. *Interacting with Computers*, 16, 1 (2004)
- [15] Ssemugabi, S., de Villiers, R.: A Comparative Study of Two Usability Evaluation Methods Using a Web-Based E-Learning Application. In SAICSIT, pp. 132-142. (2007)
- [16] Virzi, R.: Refining the Test Phase of Usability Evaluation: How Many Subjects is Enough? *Human Factors*, 34, pp. 457-468 (1992)
- [17] Virzi, R.: Usability Inspection Methods. In *Handbook of Human-Computer Interaction*, Helander, M., Landauer, T. & Prabhu, P. (Eds), 2nd Edition, North-Holland: Elsevier Science. (1997).