Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240596

Usability Checklists for Health Technology: Case Study and Experts' Opinions

Romaric MARCILLY[a,b,1], Helen MONKMAN[c,d], Paul QUINDROIT[a], Annaëlle DAVID[a] and Blake LESSELROTH[c,d]
[a]*Univ. Lille, CHU Lille, ULR 2694 - METRICS: Évaluation des technologies de santé et des pratiques médicales, F-59000 Lille, France*
[b]*Inserm, CIC-IT 1403, F-59000 Lille, France*
[c]*School of Health Information Science, University of Victoria, Victoria, BC, Canada*
[d]*University of Oklahoma, School of Community Medicine, Tulsa, Oklahoma, USA*
ORCiD ID: Romaric Marcilly https://orcid.org/0000-0002-7077-7267, Helen Monkman https://orcid.org/0000-0003-0772-9075, Blake Lesselroth https://orcid.org/0000-0001-6170-7964

Abstract. Application of usability evaluations throughout the health technology lifecycle is necessary to improve the efficiency, safety, and effectiveness of health service delivery. Unfortunately, technology vendors and healthcare organizations may not have funding, time or expertise to conduct usability studies. In this paper, we describe how usability checklists can potentially fill this gap. First, we introduce a case study using a checklist to identify usability issues with a primary care dashboard. Then we provide an expert summary of the strengths and limitations of usability checklists. Findings suggest that checklists are efficient to identify important usability issues. They can be used effectively by project team members – including clinicians – without formal usability training. However, checklists should complement rather than replace usability evaluations with representative users.

Keywords. Usability; heuristics; checklists; health technology; dashboard; user-centered design

1. Introduction

Usability evaluation of healthcare technologies (HT) improves the detection of potentially serious usability, safety, and performance problems associated with HT and health services delivery. Several methods are available for formative evaluation (*i.e.*, aiming to improve usability during system development). Evaluations can be conducted with end-users (*e.g.,* user tests) or without end-users (*e.g.,* heuristic evaluations, cognitive walkthrough). These two broad approaches are complementary because they tend to identify different types of problems [1]. Methods involving end-users, however, are often time-consuming and costly to conduct and analyze. By contrast, evaluations without end-users can be completed quickly in resource-constrained settings and are well suited to the iterative processes used in (re)design-evaluation cycles. Researchers using

[1] Corresponding Author: Romaric Marcilly; E-mail: romaric.marcilly@univ-lille.fr.

evaluation methods often need specialized expertise, a strong foundation in interaction design concepts, and some knowledge of the medical domain. For example, the cognitive walkthrough, which evaluates an interface by analyzing the cognitive processes involved in the interaction, requires knowledge of how clinicians think [2]. Inspection methods wherein evaluators use a standardized rubric of heuristics or a checklist, are somewhat unique in that they can potentially be used by non-experts in usability [3].

Organizations with fixed resources or tight timelines often favor inspection methods to quickly generate usability data and identify opportunities for improvement. Common inspection approaches include heuristic evaluation and checklists. In the first approach, evaluators determine whether the interface design adheres to predetermined principles (*i.e.,* heuristics). Heuristics are general statements that the evaluator may adapt to the technology and use case. Deviations are identified as usability problems. In the second approach, the evaluator uses a checklist of more specific usability criteria to evaluate the system's interface. Checklist items are generally formulated as closed questions to which the evaluator can answer yes/no/not applicable. Studies that compared checklists with heuristics found that checklists may detect additional low-risk or less severe problems overlooked with heuristic evaluation [4,5]. The authors of these studies tend to agree that checklists are better suited to evaluators with less experience [4,5]. For this reason, checklists for non-expert evaluators have been developed to assess HT, including mobile device collection forms [6], medication alerting systems [7], auto-injector pens [8], data visualization dashboards [9,10], and health literacy screening instruments [11].

In this paper, we share our experiences and lessons learned using a checklist to evaluate the usability of a primary care electronic dashboard. We supplement these findings with the opinions and recommendations from usability evaluation experts. Through this process, we hope to (1) highlight the role of checklists in user-centered design, (2) address perceived barriers to the use of checklists, and (3) describe best practices for implementing checklists in operational settings.

2. Case study

We evaluated a clinical dashboard prototype that provides general practitioners (GPs) with an overview of their patient panel activity (Figure **1**), including number of patients seen and consultations completed. The prototype was fully functional, interactive, and populated with real, anonymized patient data. The dashboard was intended to help GPs quickly audit their productivity and workload (e.g., reimbursement rates) or identify practice patterns (e.g., rates of drug prescriptions). We conducted our evaluation during the first evaluation and (re)design cycle. Our project team included usability specialists and clinical personnel. GPs did not take part in this evaluation to save their time for the user tests in which they are essential. Considering the various levels of expertise within the project team, we decided to use a checklist rather than a heuristic evaluation. Checklists are more suitable for evaluators without or with limited prior training in human factors [4,5].

3. Methods

We used Ansari and Martin's usability checklist because it was designed specifically for evaluating clinical dashboards [9]. This checklist includes 85 criteria grouped into 11

dimensions. Each finding is assigned a severity level (*i.e.,* major, minor). During two 90-minute sessions, three evaluators used the checklist independently to evaluate the dashboard: (1) a human factors trainee with no usability evaluation experience, (2) a primary care researcher with no usability evaluation experience, and (3) a human factors researcher with 15 years of experience evaluating HT. After completing our independent evaluations, we compared results and resolved discrepancies through discussion to reach a consensus. We calculated inter-rater agreement using Gwet's AC2 [12]. For each finding, we proposed interface improvement recommendations based on best practices for graphical user interface design [13].

Figure 1. Screenshot of the prototype primary care clinical dashboard under evaluation.

4. Results

Inter-rater agreement (AC2) was 0.65 (0.54, 0.76). Using conventions applied to rate Cohen's Kappa, this represented a "substantial" agreement [14]. Of the 85 criteria reviewed, 17 did not apply to the dashboard, 42 were satisfied, and 26 were usability issues (i.e., 20 major and 6 minor violations) (Figure 2). We identified several categories of problems. First, the visual representation of a tabbed layout was not intuitive to navigate (e.g., color difference between tab head and body). Second, the absence of titles and labels made charts and tables difficult to read. Third, without the ability to zoom or filter views, units and legends were difficult to interpret. Fourth, interactive buttons and icons were not easy to identify. Finally, users could not view multiple data series simultaneously to compare results.

5. Discussion and Expert Opinion

5.1. Discussion of case study results

The application of Ansari and Martin's checklist enabled evaluators without human factors expertise to efficiently identify major design violations without prior training. We believe this method offers an effective and efficient way for organizations to evaluate HT under time pressure and on a budget as soon as checklists specific to the technology

under evaluation exist. This fact notwithstanding, some items were difficult for evaluators to understand and required discussion to ensure inter-rater reliability. These technical challenges tended to be due to a lack of understanding of a dashboard-specific term (e.g., "thematic frame") rather than to a lack of knowledge of human factors. Importantly, our human factors expert identified additional usability problems that were not identified using the checklists. For example, he reported common interface problems such as small font size. It should be remembered that checklists may not identify all usability issues. It is best to combine them with complementary methods when the stakes are high, and completeness is crucial.

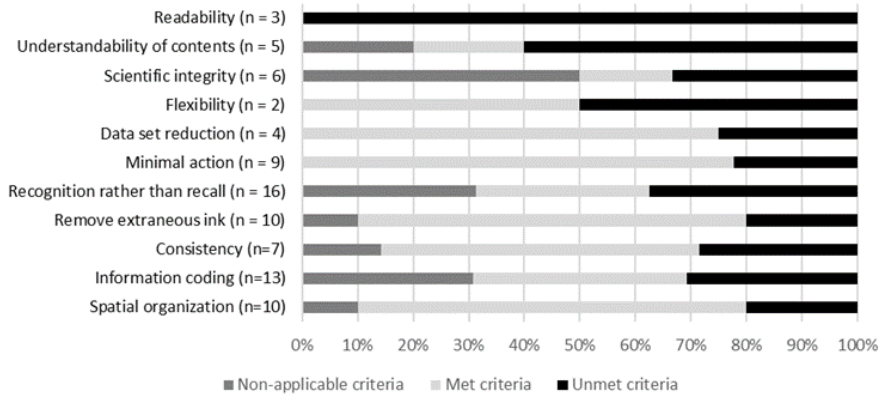

Figure 2. Usability findings for each dimension of Ansari and Martin's checklist.

5.2. Expert opinion on usability checklists

Three of the authors have a background in human factors and over 10 years of experience using mixed methods – including checklists – to evaluate HT. Here, they share perceived advantages, limitations, and opportunities using checklists.

Table 1. Summary of the Perceived Advantages and Limitations of Checklists

Perceived Advantages	Perceived Limitations
• Time: efficient evaluation methodology	• Comprehensiveness: may not identify all possible usability problems
• Cost: cost-effectiveness can increase adoption within organizations	• Fragmented Re-design: may promote fault-by-fault correction rather than large-scale reengineering
• Accessibility: can be used by practitioners without a human factors background	• False Assurance: may foster the mistaken belief that evaluators gathered all the information about the usability of the technology
• Awareness: can increase awareness of the importance, methods, and goals of usability evaluation	• Over Confidence: testers may overestimate their knowledge or ability in human factors evaluation

We have several recommendations to enhance the effective use of checklists for usability evaluations. First, checklists should be promoted to developers, vendors, and regulatory agencies as a method of usability evaluation. Second, checklists should encourage evaluators to report potential usability problems observed in addition to what is contained within the checklist. Third, checklist instructions should include statements about their validity and limitations (e.g., Are they evidence-based? What is their scope?). Fourth, checklists should be combined with other evaluation methods such as heuristic evaluations, user tests, naturalistic observations to ensure that the technology is assessed

comprehensively in different contexts and under different conditions. Fifth, ideally, at least two evaluators should use a checklist and compare their findings. Finally, software engineering and computer science curricula should include human factors topics, including usability evaluation and user-centered design frameworks to expand capacity in the industry.

6. Conclusions

Usability checklists are a feasible and effective method for quickly identifying usability issues and informing HT redesign. Project team members without usability expertise can use checklists to identify important issues. To increase the accuracy, completeness, and value of an evaluation using checklists, we recommend using more than one rater and an evaluation protocol that aligns raters' terms, methods, and mental models.

Funding

This research was funded by PreciDIAB, which is jointly supported by the French National Agency for Research (grant reference: ANR-18- IBHU-0001), the European Union's European Regional Development Fund (FEDER - grant reference: NP0025517), the Hauts- de- France Regional Council (grant reference: 20001891/NP0025517) and the European Metropolis of Lille (grant reference: 2019_ESR_11).

References

[1] Jaspers MWM. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inf 2009;78:340–53.
[2] Lewis C, Wharton C. Cognitive Walkthroughs. Handb. Hum.-Comput. Interact., Elsevier; 1997, 717–32.
[3] Nielsen J, Mack RL, editors. Usability inspection methods. New York: Wiley; 1994.
[4] Khajouei R, Hajesmaeel Gohari S, Mirzaee M. Comparison of two heuristic evaluation methods for evaluating the usability of health information systems. J Biomed Inform 2018;80:37–42.
[5] Monkman H, Griffith J. A Tale of Two Inspection Methods: Comparing an eHealth Literacy and User Experience Checklist with Heuristic Evaluation. In: Mantas J, et al., editors. Stud. Health Technol. Inform., IOS Press; 2021.
[6] Mugisha A, Nankabirwa V, Tylleskär T, Babic A. A usability design checklist for Mobile electronic data capturing forms: the validation process. BMC Med Inform Decis Mak 2019;19:4.
[7] Marcilly R, et al. Competitive Usability Evaluation of Electronic Health Records: Preliminary Results of a Case Study. Stud Health Technol Inform 2021;281:834–8.
[8] Schiro J, Pelayo S, Heyndels L, Marcilly R. Design of an Evidence-Based Checklist to Help Prevent Use Errors with Auto-Injector Pens. In: Melles M, Albayrak A, Goossens RHM, editors. Converg. Break. Barriers Discip., vol. 30, Cham: Springer Nature Switzerland; 2024, p. 275–85.
[9] Ansari B, Martin EG. Development of a usability checklist for public health dashboards to identify violations of usability principles. J Am Med Inform Assoc 2022;29:1847–58.
[10] Dowding D, Merrill J. The Development of Heuristics for Evaluation of Dashboard Visualizations. Appl Clin Inform 2018;09:511–8.
[11] Checklist - Health Literacy Online | health.gov n.d. https://health.gov/healthliteracyonline/checklist/ (accessed January 23, 2024).
[12] Gwet KL. Handbook of inter-rater reliability: the definitive guide to measuring the extent of agreement among raters. Fourth edition. Gaithersburg, Md: Advances Analytics, LLC; 2014.
[13] Nogier J-F. UX Design et ergonomie des interfaces. 6e édition. Paris: Dunod; 2016.
[14] Landis, JR, Koch GG. "The measurement of observer agreement for categorical data." *Biometrics* vol. 33,1 (1977): 159-74.