Building Capacity for Health Informatics in the Future
F. Lau et al. (Eds.)

doi:10.3233/978-1-61499-742-9-211

How to Reach Evidence-Based Usability Evaluation Methods

Romaric MARCILLY[1,a] and Linda PEUTE[b]
[a]*EVALAB, Université Lille Nord de France; France*
[b]*Academic Medical Center, University of Amsterdam, The Netherlands*

Abstract. This paper discusses how and why to build evidence-based knowledge on usability evaluation methods. At each step of building evidence, requisites and difficulties to achieve it are highlighted. Specifically, the paper presents how usability evaluation studies should be designed to allow capitalizing evidence. Reciprocally, it presents how evidence-based usability knowledge will help improve usability practice. Finally, it underlines that evaluation and evidence participate in a virtuous circle that will help improve scientific knowledge and evaluation practice.

Keywords. Human engineering, usability, evaluation, evidence, health technology

1. Introduction

There is no more need to advocate that taking into account usability during the design process can facilitate the usage of Health Technology (HT), can contribute to fulfilling HT medical intention and can support the prevention of use-errors leading to patient harm. In order to integrate usability in the design process, it is recommended to apply a User-Centered Design (UCD) process [1]. Numerous documents have been published to support the application of the UCD: standards, reference books, publications etc.

Despite the availability of those resources, numerous instances highlight the erroneous application or the non-application of the UCD: methods are incorrectly applied (e.g. no scenarios used during user testing), or methods applied do not fit the context of the evaluation (e.g. heuristics evaluation for summative evaluation). Those failures in the design process may lead to usability issues that, in turn, may lead to usage problems (e.g. use errors, miscommunication, workarounds, non usage, ordering errors) and negative outcomes on the work system and even on the patient (e.g. care process hampered and changed, implementation failure, patient harm) [2;3]. One cause of those design failures is the difficulty experienced by designers and project' stakeholders in understanding and applying correctly the UCD and in being convinced of the necessity of applying it. Building and sharing evidence-based usability knowledge is needed to provide designers and stakeholders with persuasive proofs of the necessity of applying correctly usability methods, processes, and knowledge during the design process.

By analogy to the definition of evidence-based medicine by Sackett [4], evidence-based usability can be defined as "the conscientious, explicit and judicious use of

[1] Corresponding author: romaric.marcilly@univ-lille2.fr

current best evidence in making decisions in design of interactive systems in health care by applying usability engineering and usability design principles that have proven their value in practice"[5]. Evidence-based usability practice refers to (i) evidence-based usability design principles (how to design in terms desirable characteristics of interface, behavior, and interaction), and (ii) evidence-based usability methods (how to design in terms of methods and process to apply).

The paper at hand focuses on a subset the second type of knowledge: evidence-based usability evaluation methods. It aims at presenting the main steps to building evidence-based knowledge about usability evaluation methods, current challenges in building this knowledge, and the positive consequences of applying evidence-based usability practice.

2. Ask what evidence is searched

Looking for evidence supporting the choice and the application of evaluation methods calls for evaluating those evaluation methods and their conditions of application. Accumulating results from those evaluations must demonstrate the value in practice of each evaluation methods. The first step towards evidence is to formulate the question that guides the search for evidence, such as: for each step of the design process (e.g. formative vs. summative evaluation), which (combination of) method(s) is the most cost-efficient? In which conditions of application? Who should perform the evaluations (e.g. usability experts, clinicians, or double experts)? How many participants take part to the evaluation? etc.

Several attempts to answers to those questions can be found in the literature, books, and guidelines [6-9]. However, those answers are mainly based on inferences drawn from mathematical models [10;11], analyses of the cognitive processes involved in the application of the methods, or single studies [6]. Yet, achieving evidence requires critically accumulating data from several high quality evaluation studies; currently, those answers are not evidence-based.

3. Perform high quality evaluations

The protocol of the evaluation studies must be designed so that it allows to answer the question asked: it is necessary to compare methodological options (e.g. types of methods, types of performers, type of HT evaluated) and to control as far as possible others factors. Similarly, the measures collected must be chosen and defined so that to answer the question asked: e.g. the efficiency of a method can be measured by the number of usability issues detected in relation to the cost of the method.

Evaluations performed during case studies are a valuable source of information. However, the methodological design of case studies does not allow comparing methodological options (e.g. number of participants, types of heuristics used) in controlled and standardized conditions. Therefore they cannot be used as a basis for building evidence-based usability knowledge. Experimental and quasi-experimental designs (e.g. case control) must be favored [12]. Their paradigm provides the opportunity to control some factors while manipulating the ones that must be compared. For instance, if one wants to know which of heuristic evaluation and cognitive walkthrough is the most efficient method at detecting severe usability flaws, both

methods must be compared while the profile of the evaluators, the material used during the evaluation and the technology-itself should be exactly the same in both conditions of evaluation.

Scientific literature provides a few instances of evaluation studies designed to evaluate evaluation methods. Even if those attempts in the field of HT are increasingly reported, they remain seldom (cf. Table 1).

Table 1. Non exhaustive list of publications comparing usability evaluation methods applied to HT.

Reference	Purpose
Yen *et al.* 2009 [13]	To compare heuristic evaluation and think-aloud protocol on the number of usability flaws detected.
Khajouei *et al.* 2011 [14]	To assess the effectiveness of cognitive walkthrough and think aloud for identifying usability problems and to compare their performance in identifying different types of usability problems.
Lacerda *et al.* 2014 [15]	To compare the impact of using a structured reporting form on the satisfaction and the number of usability flaws detected.
Peute *et al.* 2015 [16]	To compare the performance of the concurrent and retrospective think aloud methods and to assess their value in a formative usability evaluation.
Khajouei *et al.* 2016 [17]	To compare heuristic evaluation and cognitive walkthrough for evaluating usability of Health Information Technology.

4. Report full and precise results of evaluation studies

The current usability literature in HT often lacks of complete and precise descriptions of the methods applied (e.g. type of method, profile and number of participants and / or evaluators, material, scenario, and heuristics used) and of the results obtained (e.g. only a small set of usability issues is reported [18]). A huge amount of studies is wasted because the incompleteness of their publication prevents from re-using their results in order to capitalize knowledge and build evidence.

In order to improve the completeness of the reports of evaluation studies, several reporting guidelines have been published, including one specifically dedicated to usability evaluation studies (TRUE-HIT [19]). Only the adherence to such reporting guidelines will improve the quality and completeness of the reports and will increase the amount of studies from which evidence could be capitalized.

5. Compare and synthesize results

As for other types of evidence-based knowledge, systematic reviews and meta-analyses must be performed to gather and compare relevant data from previous evaluations. Data gathered can be analyzed from a quantitative perspective (e.g. in terms of number of usability issues detected, cost of the evaluation). Qualitative analyses must not be overlooked: they help identify operational characteristics to apply a method (e.g. types of usability issues detected, advantages and difficulties to apply a method).

Recently, a few researches attempted to synthesize knowledge about specific methodological questions through systematic reviews or meta-analyses:

- Pereira de Araujo *et al.* [20] intend to identify advantages and problems of usability evaluation methods applied to health collaborative systems
- Wills *et al.* [21] intend to evaluate the impact of usability evaluation and subsequent redesign on the task-completion time.

The very few number of publications available highlights that evidence-based usability knowledge is still at its very infancy. The evidence-based knowledge that comes from systematic reviews and / or meta-analysis has a limited validity in duration: the knowledge may evolve at fast pace depending on the results of new publications and on changes in the context of application (e.g. type of technology, legislation). Therefore practitioners who apply this knowledge must continuously update it in order to always use the best available knowledge.

6. From evidence-based knowledge to evidence-based practice

Evidence-based knowledge is not an end per se. It cannot stand in for the methodological and usability expertise of the evaluator. On the contrary, the best available evidence-based knowledge must be adapted and contextualized to each specific case of application (e.g. technology under evaluation) through the expertise of the evaluator (cf. Figure 1). Therefore, once the evidence-based knowledge is available, it must be incorporated judiciously and correctly into the practice of the evaluator.

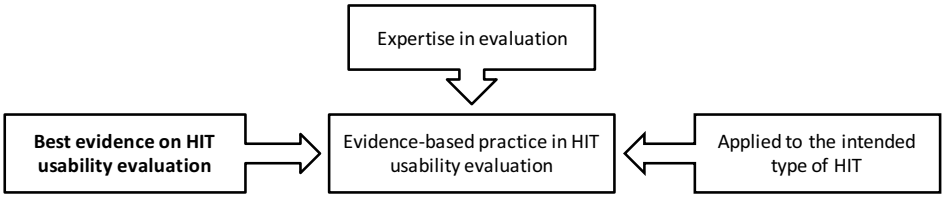

Figure 1. Schematic representation of how evidence-based usability knowledge feeds practice.

7. Discussion: evidence and practice, the virtuous circle

Evidence-based usability knowledge and more precisely evidence-based usability evaluation knowledge are at their infancy. It is a recent research topic that suffers from a lack of well-designed evaluations of evaluation methods and from a poor report of their results. Still, evidence-based usability knowledge and practice must remain a (long-term) objective for researchers in usability in HT.

The way towards evidence-based usability practice faces barriers but it is a laudable and necessary goal. As soon as the quality, the relevance, and the reporting of evaluation studies of usability evaluation methods will be improved, the amount and the quality of data available for systematic reviews and meta-analyses will increase. It will allow building evidence-based usability knowledge, not only about usability evaluation methods, but also on design principles (i.e. what are the desirable characteristics of the interface, of the behavior of the technology). Finally, once this knowledge is built and accessible to designers and evaluators, it will allow to base usability practice on evidence. In turn, this evidence-based usability practice will improve the quality of evaluation studies. In summary, targeting evidence-based knowledge and practice will lead to a virtuous circle of evaluation for and by the best available evidence (cf. Figure 2). This virtuous circle will support the improvement of both usability practice and usability scientific knowledge.

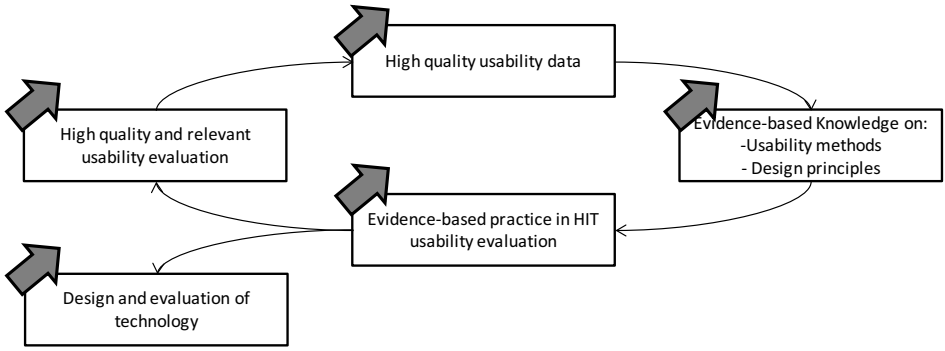

Figure 2. Representation of the virtuous circle of usability evaluation for and by evidence-based knowledge. Arrows represent the improvement of an item. Those improvements consequently improve the following item.

Building evidence-based usability knowledge and practice is very time-consuming and requires collaborative efforts. It is important that international collaborations on this topic be funded and that researchers and practitioners be taught and trained to evidence-based practice in order to start the virtuous circle. Ultimately, the development of evidence-based knowledge will help convince project's stakeholders of the relevance and necessity to consider usability during the design process.

References

[1] International Standardization Organization, *Ergonomics of human system interaction - Part 210: Human Centered Design for Interactive Systems* (Rep N°9241-210), International Standardization Organization, 2010.
[2] D. Ash, Lessons from epinal, *Clin Oncol (R Coll Radiol)* **19** (2007), 614-5.
[3] M. Grissinger, Pen injector technology is not without 'Impending' risks, *Pharm Ther* **35** (2010), 245-266.
[4] D. L. Sackett, W. M. Rosenberg, J. A. Gray, R. B. Haynes, W. S. Richardson, Evidence based medicine: What it is and what it isn't, *BMJ* **312** (1996), 71-72.
[5] Marcilly R, Peute LW, Beuscart-Zephir MC, Jaspers MW, Towards evidence based usability in health informatics? *Stud Health Technol Inform,* **218** (2015), 55-60.
[6] Jaspers MW, A comparison of usability methods for testing interactive health technologies: Methodological aspects and empirical evidence, *Int J Med Inform,* **78** (2009), 340-353.
[7] J. Nielsen, *Usability Engineering,* Academic Press, 1993.
[8] D. Mayhew, *The Usability Engineering Lifecycle,* 1st edition ed. Elsevier, 1999.
[9] The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, 2d Edition ed. Lawrence Erlbaum Associates, 2007.
[10] S. Borsci, R. D. Macredie, J. L. Martin, T. Young, How many testers are needed to assure the usability of medical devices? *Expert Rev Med Devices,* **78** (2014), 513-525.
[11] M. Schmettow, W. Vos, J. M. Schraagen, With how many users should you test a medical infusion pump? Sampling strategies for usability tests on high-risk systems, *J Biomed Inform,* 46 (2013), 626-641.
[12] Oxford centre for evidence-based medicine - levels of evidence, [cited 2016 Nov 10]. Available from: http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/.
[13] P. Y. Yen, S. Bakken, A comparison of usability evaluation methods: Heuristic evaluation versus end-user think-aloud protocol - An example from a web-based communication tool for nurse scheduling, *AMIA Annu Symp Proc,* 2009: 714-718.
[14] R. Khajouei, A. Hasman, M. W. Jaspers, Determination of the effectiveness of two methods for usability evaluation using a CPOE medication ordering system, *Int J Med Inform*, **80** (2011), 341-50.
[15] T. C. Lacerda, C. G. von Wangenheim, I. Giuliano, Does the use of structured reporting improve usability? A comparative evaluation of the usability of two approaches for findings reporting in a large-scale telecardiology context, *J Biomed Inform,* **46** (2014), 222-230.

[16] L. W. Peute, N. F. de Keizer, M. W. Jaspers, The value of retrospective and concurrent think aloud in formative usability testing of a physician data query tool, *J Biomed Inform,* **55** (2015), 1-10.
[17] R. Khajouei, E. M. Zahiri, Y. Jahani, Comparison of heuristic and cognitive walkthrough usability evaluation methods for evaluating health information systems, *J Am Med Inform Assoc,* online Aug 7, 2016.
[18] L. W. Peute, R. Spithoven, P. J. Bakker, M. W. Jaspers, Usability studies on interactive health information systems; Where do we stand? *Stud Health Technol Inform,* **136** (2008), 327-32.
[19] L. W. Peute, K. F. Driest, R. Marcilly, S. Bras Da Costa, M. C. Beuscart-Zephir, M. W. Jaspers, A framework for reporting on human factor/usability studies of health information technologies, *Stud Health Technol Inform,* **194** (2013), 54-60.
[20] J. Chan, K. G. Shojania, A. C. Easty, E. E. Etchells, Does User-Centred Design Affect the Efficiency, Usability and safety of CPOE order sets? *J Am Med Inform Assoc,* **18** (2011), 276-81.
[21] A. Liberati, D. G. Altman, J. Tetzlaff, C. Mulrow, P. C. Gotzsche, J. P. Ioannidis, et al., The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration, *PLoS Med,* **6** (2009), e1000100.