

Insights and limits of usability evaluation methods along the health information technology lifecycle

Romarc MARCILLY^{1a}, Andre W. KUSHNIRUK^b, Marie-Catherine BEUSCART-ZEPHIR^a and Elizabeth M. BORYCKI^b

^aINSERM CIC-IT 1403, Lille ; Univ Lille Nord de France ; CHU Lille ; UDSL EA 2694 ; F-59000 Lille, France

^bSchool of Health Information Science, University of Victoria, Victoria, Canada

Abstract. A great variety of usability evaluation methods exist but they do not provide the same kind of results and do not address the same stage of the Health Information Technology (HIT) lifecycle. This paper takes stock of the application of expert evaluation, usability testing, clinical simulation, clinical trials and post-implementation surveillance to provide an overview of their main similarities and differences. Results from this comparison will help in choosing methods that are best able to evaluate a HIT and improve its usability and ultimately its safety of use.

Keywords. Human engineering; Technology assessment; Medical errors; Evaluation studies;

Introduction

It is becoming increasingly more clear that the usability of Health Information Technology (HIT) must be considered during its design, its implementation and even after it is put into use [1]. However, there continue to be many reported issues that have poor usability in healthcare. Usability has been shown to affect the quality of health professional clinical reasoning, HIT adoption rates and the occurrence of medical errors [2]. Regulatory authorities are aware that this problem requires that usability engineering methods to be applied during the design process [3]. In this paper, we focus on methods that can be used to evaluate the usability of HIT. Usability evaluations aim to identify violations in usability design principles (*i.e.* usability flaws) and their consequences for the user (*i.e.* usage problems) and for the work system (e.g. negative outcomes) in order to prevent them [4]. A great variety of usability engineering methods exist, yet, they do not provide the same kind of results [5] and they do not address the same stage of the HIT lifecycle [6]. The purpose of this paper is to take stock of the application of usability engineering methods along the lifecycle of an HIT. Our paper aims to provide an overview of the similarities and differences in those methods in terms of usability results, lifecycle's stages concerned and their

¹ Corresponding author : Romarc Marcilly, CIC-IT Biocapteurs & e-santé, Innovation & usages, Maison Régionale de la Recherche Clinique, 6 rue du professeur Laguesse, 59037 Lille. E-mail : romarc.marcilly@univ-lille2.fr / romarc.marcilly@gmail.com

specific added value in comparison to other methods. Ultimately, our work aims to help individuals and groups working on HIT projects to choose those methods that are best able to evaluate their HIT and improve the usability of their products.

1. Background: HIT Lifecycle and Usability Evaluations

Usability evaluations methods can be applied throughout the HIT lifecycle (Table 1) as soon as this lifecycle meet usability engineering needs for evaluation (e.g. involving users, providing room for standardized evaluations). Approaches such as pure "Agile", that do not support involving end-users nor evaluating stabilized versions are excluded: they need to be adapted to meet those usability engineering requirements.

Usability methods' application ranges [1;7]:

- From early stages of design (e.g. evaluating alternate user interface mock-ups) to make informed usability design decisions, check the compliance of the system with usability requirements and reduce the risks of the use errors that could be induced by a poor usability
- Through implementation and evaluation of the usability of systems after they have been deployed to assess the impact of usability characteristics of the system on users and on the work system, and to get feedback to ultimately improve the usability of next versions of the systems.

Table 1. Applicability of usability evaluations methods throughout the HIT lifecycle.

Type of evaluation	Stage in the HIT lifecycle				
	First mock-up	Interactive mock-up	Near completed	Completed	Marketing authorization
Expert evaluation	X	X	X	X	X
Usability testing		X	X	X	X
Clinical simulation			X	X	X
Clinical trials				X	X
Post implementation surveillance					X

2. Usability Evaluation Methods

Descriptions and comparisons of the types of methods are based on lessons learned from the literature and the authors' own experience. Types of methods are presented from the less reliable in terms of real-world usage on the left side of the continuum to the ones with the highest fidelity on the right side of the continuum in Figure 1.

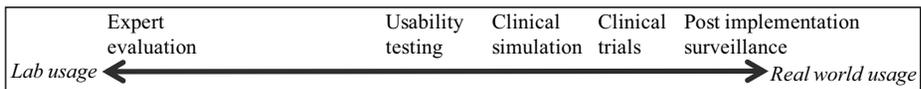


Figure 1. Graphical representation of the range of usability evaluation methods from least real-world usage (far left) to the most real-world usage (far right).

2.1. Expert Evaluation

The main methods of usability evaluation that lie on the far left of the continuum include lab usage involving expert evaluations. This includes usability inspection

methods where one or more analysts systematically step through a user interface. The two main methods are heuristic evaluation (which involves analyzing a user interface by comparing it against a set of usability heuristics or guidelines) and cognitive walkthrough (which involves stepping through a user interface for a task, and noting goals, actions, system responses and potential problems). The advantages of these methods is that they do not require human subjects and can be done very cost effectively [8]. However, they do not involve observation of real users and cannot fully predict which usability flaws real users will actually face [9] or how they will interact with the system in real contexts of use (see Figure 1 and Table 2 for comparisons).

Table 2. Summary of main differences among main methods of usability evaluation.

	Place during the lifecycle	Fidelity	Insights/limits
Expert evaluation	As soon as a first mock-up is drawn	Poor: no end-users involved, scenarios inspired by context of use, default setting	Usability flaws, hypotheses on their impact on users and work system (incl. risks of usage)
Usability testing	As soon as a mock-up is sufficiently developed for the user to interact with (incl. wizard of Oz)	Weak: end-users involved, scenarios inspired by context of use, default setting	Usability flaws, usage problems, risks of usage and hypotheses on outcomes on the work system and on risks' consequences
Clinical Simulation (laboratory-based and In-situ)	As soon as a mock-up is sufficiently developed for the user to interact with (incl. wizard of Oz)	Medium: end-users involved, scenarios inspired by context of use, contexts and settings typical of real-world	Usability flaws, usage problems, risks of usage and hypotheses on outcomes on the work system and on risks' consequences
Clinical trials	Final version undergoing clinical trial	High: actual usage by end-users in a limited range of cases (no worth cases), actual setting for the ward	Usability flaws, usage problems, risks of usage, outcomes on the work system but only for intended usages
Post implementation surveillance	HIT actually in use (requiring marketing authorization)	Very high: actual usage (even unintended), actual setting for the ward	Usability flaws, usage problems, risks of usage, outcomes for (un-)intended usages

2.2. Usability testing

Usability testing involves observing representative end-users (e.g. physicians) interacting with representative systems (e.g. electronic health records) to carry out representative tasks (e.g. medication entry). The approach has been used for a wide range of usability studies in healthcare and such studies are often carried out as controlled observation and may involve the “think-aloud” method, where users are asked to “think aloud” as they carry out tasks (with their interactions recorded). Such studies are often carried out in usability laboratories. Although the fidelity of such studies is higher than that of expert evaluations, they are typically not carried out in real hospital or clinical settings, limiting their ecological validity and generalizability to real-world contexts. In terms of the HIT lifecycle, this method can be used for testing a range of systems, from early interactive mock-ups to (near) completed systems [10].

2.3. Clinical Simulation (Laboratory-based and In-situ)

Clinical simulations extend usability testing by locating the studies in real or realistic settings (*i.e.* contexts of use). Clinical simulations, like usability tests, typically involve observing end-users as they interact with a system. However, they typically involve conducting such studies in a simulation laboratory, where the fidelity of testing is higher than in a typical usability laboratory, or in real settings (*i.e.* in-situ testing)[11].

The advantage of this approach is that the fidelity of the evaluation is high. However, such studies require more planning or resources than pure usability testing (*i.e.* they require a simulation laboratory, or access to real settings such as hospital rooms after hours) and typically involves completed and near completed systems [12].

2.4. Considering usability during clinical trials of technology

One step further in the fidelity of usability evaluations is to take the opportunity to participate in clinical trials with a system in order to understand its actual usage by end-users in their work system with patients. However, clinical trials include only specific situations of use that match given inclusion criteria (e.g. patients' types) restricting the range of situations of observation (e.g. excluding worst case situations). In contrast to usability test/simulations, clinical trials focus on the number of patients treated with a technology, not on the number of users [13]. Both issues may ultimately limit the type of usability issues collected. Although important dependent outcome measures are collected from such trials (e.g. impact of a reminder system on error rates), such approaches are less useful in allowing investigators to understand how specific aspects of complex system or user interface interaction design lead to the observed outcomes. Besides, such evaluation requires systems that are completed in the HIT lifecycle. As far as we know, no published study reports taking the opportunity of clinical trials to evaluate the usability of a system and its software components.

2.5. Post implementation surveillance

High fidelity usability evaluation methods can be used after a system is implemented. Data collected during these evaluations are richer than data from in-lab evaluations [5]: they provide information about usability flaws and their consequences for the user and the work system (including use errors). Such usability flaws can be identified through direct observation [14], users' questionnaire or review of system reports [15]. In contrast to clinical trials, actual usage is not limited to the usage intended by the manufacturer, but includes unintended uses of the technology by users. However, as for clinical trials, the complexity of the work system in which the system is implemented can make it more difficult to determine how the usability of the system impacts users and clinical outcomes [13]. Moreover, this approach requires that systems be completed in the HIT lifecycle and get their marketing authorization (e.g. CE marked).

3. Discussion

There are a variety of methods for evaluating usability but they do not provide the same kinds of usability insights and cannot be used at the same stage of the HIT lifecycle. Post implementation surveillance provides a more complete coverage in terms of usability insights than other methods (useful to develop usability knowledge). However, the method can be applied only for HIT already in use and its usability feedbacks cannot be used directly by manufacturers: a new version must be developed. In contrast, expert evaluations can be performed at any stage of the HIT lifecycle but it highlights only usability flaws. Nonetheless, as long as the evaluation is performed as the beginning of the design process, feedback coming from this method can be considered quickly by the manufacturers without requiring a whole new version of the system.

Some methods do not allow for completely considering the characteristics of the real context of use and local settings (e.g. usability testing). Concluding that a HIT has good usability after applying such a method is no guarantee of a good usability once this system is put into use: a poor implementation can ruin a system with good usability. Results from those methods must be considered cautiously: moving from a default setting to a local setting can add new usability issues.

Each method has its own advantages and pitfalls but a determining factor in choosing a method is its cost-benefit ratio. As far as we know, there is no study published that compares all these methods from an economic perspective. Future research should tackle this point. To do so, for each type of method, the number of usability issues uncovered, their impact on users and the work system and risks of use errors they may cause must be considered. As for the costs, material costs must be assessed along with human costs (hiring experts and actual end-users), savings related to the risks identified and costs in terms of re-engineering. Results of those researches will support choosing method(s) providing the best coverage of usability issues for the least cost in order to improve HIT usability and ultimately their safety of use.

References

- [1] Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. *J Biomed Inform* 2004 Feb;37(1):56-76.
- [2] HIMSS Usability Task Force. Promoting Usability in Health Organizations: Initial Steps and Progress Toward a Healthcare Usability Maturity Model. Healthcare Information and Management Systems Society; 2011 Feb.
- [3] Council Directive 2007/47/EC, 2007:247:0021:0055, European Parliament Council, (2007).
- [4] Marcilly R, Beuscart-Zephir MC, Ammenwerth E, Pelayo S. Seeking Evidence to Support Usability Principles for Medication-Related Clinical Decision Support (CDS) Functions. *Stud Health Technol Inform* 2013;192:427-31.
- [5] Marcilly R, Vasseur F, Ammenwerth E, Beuscart-Zephir MC. Methods uncovering usability issues in medication-related alerting functions: results from a systematic review. *Stud Health Technol Inform* 2014;205:885-9.
- [6] International Standardization Organization. Ergonomics of human-system interaction -- Usability methods supporting human-centred design (Rep N° 16982). Geneva: International Standardization Organization; 2002.
- [7] Beuscart-Zephir MC, Elkin P, Pelayo S, Beuscart R. The human factors engineering approach to biomedical informatics projects: state of the art, results, benefits and challenges. *Yearb Med Inform* 2007;109-27.
- [8] Nielsen J. Usability Engineering. Boston: Academic Press; 1993.
- [9] Jaspers MW. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. *Int J Med Inform* 2009 May;78(5):340-53.
- [10] Bastien JM. Usability testing: a review of some methodological and technical aspects of the method. *Int J Med Inform* 2010 Apr;79(4):e18-e23.
- [11] Li AC, Kannry JL, Kushniruk A, Chrimes D, McGinn TG, Edonyabo D, et al. Integrating usability testing and think-aloud protocol analysis with "near-live" clinical simulations in evaluating clinical decision support. *Int J Med Inform* 2012 Nov;81(11):761-72.
- [12] Borycki E, Kushniruk AW, Anderson J, Anderson M. Designing and integrating clinical and computer-based simulations in health informatics: From real-world to virtual reality. Vukovar, Croatia: In-Tech; 2010 p. 31-52.
- [13] Wiklund M, Kendler J, Strohlic AY. Usability testing of medical devices. CRC press ed. 2010.
- [14] Russ AL, Zillich AJ, McManus MS, Doebbeling BN, Saleem JJ. Prescribers' interactions with medication alerts at the point of prescribing: A multi-method, in situ investigation of the human-computer interaction. *Int J Med Inform* 2012 Apr;81(4):232-43.
- [15] Magrabi F, Baker M, Sinha I, Ong MS, Harrison S, Kidd MR, et al. Clinical safety of England's national programme for IT: A retrospective analysis of all reported safety events 2005 to 2011. *Int J Med Inform* 2015 Mar;84(3):198-206.