Journal of Biomedical Informatics xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

From expert-derived user needs to user-perceived ease of use and usefulness: A two-phase mixed-methods evaluation framework

Mary Regina Boland [a], Alexander Rusanov [b], Yat So [a], Carlos Lopez-Jimenez [c], Linda Busacca [d], Richard C. Steinman [c,f], Suzanne Bakken [a,e,f], J. Thomas Bigger [c,f], Chunhua Weng [a,f,*]

[a] Department of Biomedical Informatics, Columbia University, New York, NY, USA
[b] Department of Anesthesiology, Columbia University, New York, NY, USA
[c] Department of Medicine, Division of Cardiology, Columbia University, New York, NY, USA
[d] The Clinical Trials Office, Columbia University, New York, NY, USA
[e] School of Nursing, Columbia University, New York, NY, USA
[f] The Irving Institute for Clinical and Translational Research, Columbia University, New York, NY, USA

ARTICLE INFO

Article history:
Received 31 July 2013
Accepted 6 December 2013
Available online xxxx

Keywords:
Evaluation studies
Clinical trials
Workflow
Needs assessment
Medical informatics

ABSTRACT

Underspecified user needs and frequent lack of a gold standard reference are typical barriers to technology evaluation. To address this problem, this paper presents a two-phase evaluation framework involving usability experts (phase 1) and end-users (phase 2). In phase 1, a cross-system functionality alignment between expert-derived user needs and system functions was performed to inform the choice of "the best available" comparison system to enable a cognitive walkthrough in phase 1 and a comparative effectiveness evaluation in phase 2. During phase 2, five quantitative and qualitative evaluation methods are mixed to assess usability: time-motion analysis, software log, questionnaires – System Usability Scale and the Unified Theory of Acceptance of Use of Technology, think-aloud protocols, and unstructured interviews. Each method contributes data for a unique measure (e.g., time motion analysis contributes task-completion-time; software log contributes action transition frequency). The measures are triangulated to yield complementary insights regarding user-perceived ease-of-use, functionality integration, anxiety during use, and workflow impact. To illustrate its use, we applied this framework in a formative evaluation of a software called Integrated Model for Patient Care and Clinical Trials (IMPACT). We conclude that this mixed-methods evaluation framework enables an integrated assessment of user needs satisfaction and user-perceived usefulness and usability of a novel design. This evaluation framework effectively bridges the gap between co-evolving user needs and technology designs during iterative prototyping and is particularly useful when it is difficult for users to articulate their needs for technology support due to the lack of a baseline.

1. Introduction

Evaluation is important to all innovations [1], including Health Information Technology (HIT) interventions. However, there are significant barriers for evaluating HIT, such as the lack of a reference HIT gold standard [2], the paucity of knowledge regarding user needs prior to the development of an HIT intervention [3], and the complexity of socio-technical systems and multi-stakeholder teams, which can affect the intended outcomes of the HIT intervention [4].

Five levels of usability evaluation have been described in the literature: task-based, user-task, system-task, user-task-system, and user-task-system-environment [5]. The first three levels occur early in prototype development, focusing on task identification, how users perform their tasks, and if a system supports the task it was designed for [5]. The fourth level addresses how users perform a set of tasks using the system and how users perceive the usefulness of the system [5]. Building on these, the fifth level evaluates how the task, user, and system interact within the workplace environment [5]. The fifth level usually occurs after system deployment [6], while the fourth level occurs during the prototype development stage.

Abbreviations: IMPACT, Integrated Model for Patient Care and Clinical Trials; CRC, Clinical Research Coordinator; CTMS, Clinical Trial Management System; SUS, System Usability Scale; UTAUT, Unified Theory of Acceptance of Use of Technology; PRN, *Pro Re Nata* (as needed); HIT, Health Information Technologies; STARE-HI, STAtement on Reporting of Evaluation studies in Health Informatics.

* Corresponding author. Address: Department of Biomedical Informatics, Columbia University, 622 W 168th Street, VC-5, New York, NY 10032, USA. Fax: +1 212 305 3302.
E-mail address: cw2384@columbia.edu (C. Weng).

1532-0464/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jbi.2013.12.004

Often, HIT prototypes are not fully comparable with existing systems because of their inherent novelty and uniqueness. Identifying an appropriate baseline or comparison system (when a baseline is lacking) for evaluation purposes is difficult for emerging HIT interventions [7]. However, it is important to overcome this problem and select the "best available" system as a reference standard for comparing the usability and effectiveness of various systems against.

Various evaluation methods and strategies have been developed [8]. Evaluations that mix methodologies are considered robust [9,10] and particularly useful in the medical setting [11,12]. There are many ways to combine methods, such as mixing qualitative and quantitative methods [13], involving users of varying perspectives for data collections [14], or using various data collection methods to achieve greater data validity. The mixed-methods approach is superior to either qualitative or quantitative research methods alone [15] because it ensures comprehensive data collection and avoids unnecessary *a priori* assumptions often made by researchers [13]. In a mixed-methods evaluation, qualitative data can be used to identify unmet needs [16–18], while quantitative data can measure workflow impact [18–22]. Data triangulation further allows verification of derived user needs [23].

In addition, evaluation designs can involve different types of evaluators, including usability experts and intended users. Several evaluation methods utilize usability experts. Cognitive task analysis (CTA) is an evaluation method performed by usability experts for assessing usability and has been successfully applied in healthcare settings [24]. Cognitive walk-through also involves usability experts but is less intensive than CTA. In a cognitive walk-through, an expert who is already familiar with the system performs a set of predefined tasks and notes the number of steps required by certain tasks and any usability and design problems with the interface [25]. Other evaluation methods make use of the intended end-users themselves. Time-motion analysis is a quantitative method that measures the amount of time users spend performing a task [26]. Results of time-motion analyses provide insight into the likelihood of system adoption and can be used to identify areas in users' workflow amenable to an informatics intervention [27]. The advantages of surveys, emails, and think-aloud protocols in evaluating informatics interventions is well-established [28]. Software log analysis is another useful evaluation method that can capture behind-the-scenes interactions with the system and is not intrusive to evaluators [29]. Questionnaires can assess users' perceptions of a system's usability [30] and the likelihood of acceptance of the technology [31]. Qualitative information in the form of unstructured interviews and think-aloud protocols are especially useful during system evaluation because they allow users to provide additional information not specified *a priori* via a structured questionnaire [32]. Mixing qualitative and quantitative evaluation methods to further enhance the evaluation result is a well-established approach [15].

To address the evaluation challenges with emerging HIT, where user needs are vague and clinical workflow is complex, we describe a two-phase mixed-methods evaluation framework to bridge the gap between co-evolving user needs and technology designs during iterative prototyping. This novel evaluation framework enables an integrated assessment of both expert-derived user needs satisfaction and the user-perceived usefulness and ease of use of emerging HIT interventions [33]. It supports formative evaluation of HIT before the release of a fully-fledged system. We applied our methodology to evaluate the prototypes of a novel clinical research decision support system called Integrated Model for Patient Care and Clinical Trials (IMPACT), which is designed to provide decision support for scheduling research visits [34]. We followed the STAtement on Reporting of Evaluation studies in Health Informatics (STARE-HI) guideline for reporting evaluation studies where applicable [35] since our framework was ideally suited for formative evaluations of software prototypes. We then describe this evaluation framework and its use in evaluating IMPACT prototypes.

2. Materials and methods

Our evaluation framework consists of two phases. In phase 1, a usability expert collects user needs and compares the intervention with related systems by aligning system functions with derived user needs for each system. This enables the selection of a suitable comparison system followed by a cognitive walk-through involving a task analysis and a comparison of interface design differences between the innovation and the comparison system. Phase 2 involves the system's end-users, Clinical Research Coordinators (CRCs) to collect quantitative and qualitative data. Fig. 1 illustrates our mixed-methods evaluation framework.

Table 1 shows the types of data collected at each phase. Two measures are assessed during phase 1: the number of steps required by each task and interface features used while performing each task (e.g., screen transitions and pop-ups). Analysis during phase 1 allows developers to assess how well the system performs in a laboratory setting. If phase 1 identifies many critical system functions that require improvement, the system can be refined prior to testing with end-users. This approach prevents end-users from being adversely affected by a system requiring critical improvements. Since phase 1 of the IMPACT evaluation revealed no such deficiencies, we were able to proceed directly to phase 2 of the evaluation.

2.1. The IMPACT system and its environment

Columbia University Medical Center (CUMC) is an academic medical center where many patients are also research participants. The IMPACT system, developed at CUMC, was designed to integrate information from both patient care and clinical research to facilitate the scheduling of research visits and coordination of patient care and research workflows. It incorporates temporal constraints from the research protocol's visit schedule and availability of research resources (e.g., rooms, equipment, and personnel) into a calendar interface. Designed for use by CRCs and schedulers, IMPACT automatically calculates resource availability and recommends suitable dates and times for the next research visit. IMPACT's complete functionality has been published elsewhere.

2.2. Phase 1: Usability expert component

2.2.1. Cross-system feature vs. derived user needs alignment

We recruited a usability expert to derive comprehensive user needs for scheduling decision support. This usability expert was independent from the design team but was present in the participatory design meetings to understand user needs. To guide user needs identification, the expert surveyed existing scheduling systems and anticipated problems that the user is likely to encounter using knowledge of CRCs' workflow. Each system's features (including those of IMPACT) were compared to this set of usability-expert derived user needs. Four relevant systems currently being used for scheduling at CUMC were included to quantify how well user needs were satisfied by each system: Microsoft Outlook Calendar, AllScripts Study Manager [36], Velos eResearch [37], and WebCAMP [38].

2.2.2. Comparison system selection

The cross-system feature alignment was used to identify a competent system to compare IMPACT with. This was done by

Please cite this article in press as: Boland MR et al. From expert-derived user needs to user-perceived ease of use and usefulness: A two-phase mixed-methods evaluation framework. J Biomed Inform (2013), http://dx.doi.org/10.1016/j.jbi.2013.12.004