Journal of Biomedical Informatics 56 (2015) 284–291

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

The use of think-aloud and instant data analysis in evaluation research: Exemplar and lessons learned

Jonathan Joe [a,*], Shomir Chaudhuri [a], Thai Le [a], Hilaire Thompson [b], George Demiris [a,b]

[a] Biomedical and Health Informatics, University of Washington, Box SLU-BIME 358047, 850 Republican St, Building C, Seattle, WA 98109, USA
[b] School of Nursing, University of Washington, Box 357266, 1959 NE Pacific Street, Seattle, WA 98195-7240, USA

ARTICLE INFO

Article history:
Received 11 June 2014
Revised 24 April 2015
Accepted 2 June 2015
Available online 10 June 2015

Keywords:
Instant data analysis
Usability testing
Usability research
Older adults
Wellness tools

ABSTRACT

While health information technologies have become increasingly popular, many have not been formally tested to ascertain their usability. Traditional rigorous methods take significant amounts of time and manpower to evaluate the usability of a system. In this paper, we evaluate the use of instant data analysis (IDA) as developed by Kjeldskov et al. to perform usability testing on a tool designed for older adults and caregivers. The IDA method is attractive because it takes significantly less time and manpower than the traditional usability testing methods. In this paper we demonstrate how IDA was used to evaluate usability of a multifunctional wellness tool, discuss study results and lessons learned while using this method. We also present findings from an extension of the method which allows the grouping of similar usability problems in an efficient manner. We found that the IDA method is a quick, relatively easy approach to identifying and ranking usability issues among health information technologies.

1. Introduction

Usability testing is an important component in design, as it aims to assess ease of use and identify learnability issues within a tool. Performing usability testing typically involves users of the target user group, as these "real" users may think or act differently than expected by the designers or developers [1]. Often, these issues are only identified when testing with real users, reinforcing the importance of doing "real world" usability testing. Furthermore, this testing can be done during early stages of development, leading to easier and cheaper fixes compared to finding issues after the product has been built and released. However, the usability testing process can be time consuming and labor intensive, which may lead designers to omit testing, as the upfront cost is perceived to be too high even though the process could be useful. Instant data analysis may be one solution to address this challenge providing real world testing while reducing the time and labor involved.

1.1. Usability testing

Traditional usability testing involves a think-aloud protocol combined with a video recording of a user from the target group while they interact directly with the device or tool in question to complete specified tasks [2–4]. This recorded video allows for observation of the user to identify points of frustration, confusion or other issues. The video is transcribed and often analyzed qualitatively or referenced for issues. These issues are then reconciled between researchers and scored by severity, depending on the frequency of the issue and how much it delayed or frustrated the user on completing the tasks. While such observational analysis identifies what causes the user to be frustrated or delayed, the reason or *why* this causes frustration is not evident. In order to better understand the users' thought process, this observational method is often combined with a think-aloud protocol. A think-aloud protocol asks the user to verbalize their thoughts as they perform the tasks required in a usability test giving insight into their mental model, and has its roots in Ericsson and Simon's work [1,5]. With these data, researchers can then examine the differences between the participants' mental model and the system's interaction model to identify errors and changes that need to be made. These thoughts can address what users like, what they dislike or how to improve the interface and tool from their perspective. Combining these two techniques with qualitative analysis of a transcript comprises the traditional method for usability testing. At the end of the analysis, researchers or designers are able to generate a list of usability issues and a related a score/severity ranking for each issue. Such usability tests have been used successfully to assess the usability of home-based telemedicine systems [6],

* Corresponding author. Tel.: +1 206 569 8403.
E-mail address: jjoe@washington.edu (J. Joe).

http://dx.doi.org/10.1016/j.jbi.2015.06.001
1532-0464/

medical diagnostic and research tools [7], and online self management tools [8], among others [9–12].

Traditional usability testing however, is not without its own challenges. While such an approach is very thorough, it can require significant amounts of manpower and time. Transcription of user comments and verbalizations, along with specifying user actions in relation to the interface can require significant amounts of manpower, which is then followed up by qualitative coding and analysis. Thus, the time between when the actual usability tests occur and when the final results are generated can span several weeks. For example, Jeffries et al.'s empirical usability study, with 6 users each participating in a 2 h usability session took 199 man-hours to analyze [13]. This may delay or discourage system improvements.

Other methods, such as heuristic evaluation, rely on usability experts to compare a system against usability principles, in order to hopefully avoid major usability issues [14–17]. Once a device or application has been through a heuristic evaluation, various aspects of the tool will have been judged to be either in or out of compliance with recognized usability heuristics [18]. From this analysis, changes can be made to bring the device or application into compliance, and avoid user frustration. While this can save time compared to conducting the usability tests and can form an important component of the design lifecycle for tools, it lacks interaction between the system and real users. Additionally it is based on the expert's assumptions about user needs and preferences, rather than the users' perspective. Users may interact differently with the system than expected by the usability expert, with the result being many unidentified usability problems. Furthermore, the fact that multiple expert evaluators are needed to do a heuristic evaluation can be challenging within a single organization [19]. Heuristic evaluation can therefore be a useful complement to traditional usability testing, but is not a direct replacement.

1.2. Instant data analysis

Instant data analysis (IDA) aims to reduce the labor and time commitment required to perform and analyze a usability test [20]. In IDA, multiple individual sessions are held on a single day. After sessions are completed, those participating in the evaluation meet to discuss the usability issues that were identified. Meeting directly after the sessions allows a better recall of the events and allows thoughts and ideas that may not be at the forefront of one's memory to be prompted by the other person involved. The idea behind this initial brainstorming session is to list as many usability issues remembered or seen down on paper. After these issues are exhausted, they are ranked based on severity and frequency with which the issue arose. This method is designed to make usability testing more accessible while retaining the advantages of "real" user testing by cutting down on the amount of time needed for analysis [20]. The majority of time involved in usability testing goes into understanding what issues were identified during the tests. Instant data analysis reduces the amount of time needed for analysis significantly, potentially allowing results to be seen the same day as the usability testing sessions. Previous studies have shown that using IDA can reduce the amount of time needed for analysis by 90%, while achieving 85% overlap in critical usability issues compared to the traditional standard video analysis, while a second study found 76% overlap between the two methods [20,21]. However, this method is relatively novel. To date, it has been used successfully to improve the design of medication lists to reduce adverse drug events, personal health applications, and electronic meeting support systems [9,22–25].

This paper details our experiences using the novel IDA method together with analysis mapping methods. We use an exemplar of this method in the evaluation of a multifunctional wellness tool designed for older adults. We provide insight into the feasibility of the IDA method and discuss our experiences and insights of this method to inform future researchers, designers and other stakeholders who evaluate the usability of technology tools.

2. Case exemplar

The number of adults aged 65 or older in the United States is projected to grow quickly over the next few decades, climbing from 40 million in 2010 to 72 million by 2030 [26]. As people age, they are more likely to have health issues and multiple comorbidities, leading to an increased need for health interventions [27] while the healthcare workforce is not increasing at a similar rate. Information technology is emerging for the delivery of health related interventions targeting both health maintenance and disease management. While the use of technologies has generally grown, the usability of these technologies have lagged for older adults, who have their own unique needs [28,29]. Usability concerns will play a larger role, potentially leading to greater user dissatisfaction and reduced effectiveness.

This paper is based on a pilot study for testing the usability of a multifunctional, commercially available wellness tool for older adults, hereafter referred to as "device A" using IDA as the usability testing approach. The purpose of the pilot was to evaluate and assess usability issues with the device in an older adult population. Older adult participants ($N = 5$) were recruited at an independent retirement community via information sessions. Participants could not have had prior exposure to the device to be evaluated. All participants conducted usability sessions individually, and were given 3 tasks to complete using the device.

2.1. Design

Usability testing was accomplished with a think-aloud protocol that asks users to verbalize their thoughts as they complete various tasks, allowing investigators to gain insight on participants' thought processes in relation to the interface and task [1]. Sessions included a single participant and a facilitator and designated note-taker, who observed and took notes as the participant worked through the various tasks. Testing involved a short questionnaire which asked about demographics, eHealth literacy (eHEALS) [30] and other technology use questions, followed by 3 tasks for the participants to work through. A brief post-session interview was then conducted to solicit further feedback regarding their overall impressions of the system, suggestions for improvement, and any particular frustrations they wanted to emphasize. The University of Washington institutional review board approved all procedures in this study.

2.2. Device

This study focuses on usability testing a commercially available multifunctional wellness tool, Device A. Device A is a multifunctional, touchscreen wellness tool installed in over a thousand communities across the US. It has features that were selected to address many different dimensions of wellness, including social wellness (email, video chat, reminiscence features), cognitive wellness (brain exercises, puzzles), spiritual wellness (videos, relaxation), and physical wellness (exercise videos, aerobics), among many others.

Physically, the device consists of a touchscreen computer, with a keyboard, mouse and speakers on a movable stand. The entire device is mounted to allow user-adjustable height. The main navigation consists of a 3×3 grid, where each point is a button that specifies a category or folder, with a hierarchy that is several levels