Usability of Patient-Centered Health IT: Mixed-Methods Usability Study of ePill

Manuel SCHMIDT-KRAEPELIN[a], Tobias DEHLING[a], and Ali SUNYAEV[a,1]
[a]*Faculty of Management, Economics and Social Sciences, University of Cologne, Germany*

Abstract. To facilitate use of patient-centered health IT applications in everyday life, a high degree of usability is required. Based on the example of a patient-centered web application, we propose a usability study design enabling developers and researchers to assess usability of patient-centered health IT applications and derive implications for their improvement. Our study design integrates tasks that subjects have to process, an associated questionnaire based on Perceived Ease of Use, Perceived Usefulness, Attitude Toward Using, and Behavioral Intention to Use, a System Usability Scale questionnaire, and focus groups. Application of the usability study design demonstrates its feasibility and provides insights for assessment of usability in related projects in research and practice.

Keywords. Health Information Technology, Research Design, Medication Adherence

1. Introduction

Everyone has to take medication, albeit to varying extent. To facilitate adequate use of medication and to avoid potentially harmful repercussions of wrong medication use, proper medication information is essential [1]. Recent studies show that there is a need for improvement of conventional, written information on pharmaceuticals [2], [3]. The patient-centered web application ePill (http://epill.uni-koeln.de) is designed to address this problem through provision of medication information and supplementary services [4]. A detailed description of ePill's architecture and functionality can be found in [5]. Usability is a core concern in health IT application development because users must be able to use health IT applications in order to gain benefit. Furthermore, users should not be discouraged from using or continuing use of otherwise promising health IT applications by avoidable deficiencies like usability issues. We conducted a usability study to identify usability problems and examine whether users are able to use core features of ePill.

In healthcare, IT has to be easy to use and leverage interactivity [6]. Designing patient-centered health IT applications without analyzing and tailoring usability to user expectations leads most likely to applications that fail to generate true value for users [7]. On the other hand, designing interfaces on the basis of usability principles inspires user confidence and results in usable and useful interfaces [8]. Due to different understandings and definitions of usability in literature [9] and a multitude of available

[1] Corresponding Author: Ali Sunyaev; Faculty of Management, Economics and Social Sciences, University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany; Email: sunyaev@wiso.uni-koeln.de

methods for usability testing, it is not trivial to select appropriate methods that assess desired aspects of usability. Several authors deal with usability in context of health information on the World Wide Web (e.g., [6], [10], [11]). Studies from other fields, for example, design and usability assessment of the 'Nutrition Label' for privacy [12] offer valuable insights as well. Results of the studies show that usability testing can deliver important information to detect potential for optimization. In this paper, we present a mixed-methods research design for usability studies based on the example of ePill [5].

2. Methods

To assess usability of ePill, we derived a context specific definition of usability, instructed subjects to solve simple tasks to familiarize themselves with ePill before answering questionnaires, and conducted focus groups to elicit qualitative feedback.

2.1. Usability

Although available definitions of usability (e.g., [13], [14]) have some aspects in common, there is no generally applicable definition of usability. Effectiveness, efficiency, and satisfaction are however three aspects common and central to the majority of usability definitions. Hence, we focus on effectiveness, efficiency, and satisfaction in the ePill usability study. Effectiveness is "the ability of users to complete tasks using the system, and the quality of the output of those tasks" [15]. Efficiency is the relation between effectiveness and the amount of effort needed. Satisfaction refers to users' "subjective reactions to using the system" [15]. In the context of ePill, effectiveness assesses whether users are able to use core ePill functionality: finding medication information, reading medication information, and leveraging supplementary services (e.g., comparing information on two or more medications and discovering drug interactions). Effectiveness assessment is of particular importance in the case of ePill: If users do not recognize functionality offered by ePill, they will not be able to leverage ePill features transcending those of conventional patient information leaflets and, as a consequence, they will not be able to benefit from use of ePill. An objective measuring of efficiency could be accomplished by comparing time required to find some specific medication information using ePill to the time required to find the same information using conventional patient information leaflets. However, we do not focus on the factor of time because, when providing medication information, where inaccuracies can obviously lead to serious repercussions, quality of the provided information and effectiveness are more important. To collect information on perceived efficiency we elicit users´ subjective impression of efficiency. Satisfaction refers to users' subjective view of ePill. Several aspects of satisfaction are assessed (e.g., users' subjective impression of efficiency, whether they liked using the system or whether they find the system useful). To elicit answers on these questions we used a questionnaire based on established constructs [16]. All instruments used, including employed task sheets, scenario descriptions, and questionnaires, are available from the authors upon request.

2.2. Tasks and Task Questionnaires

To investigate usability of ePill, we designed different tasks requiring subjects to find specified medication information through use of ePill. A short storyline was employed enabling subjects to put themselves into the position of a patient suffering from hypertension. Every task consisted of several subtasks. The majority of tasks asked subjects to give a short description of their solution approach. All subjects were provided with the same tasks and no hints regarding what ePill functionality was supposed to be used was given in order to investigate which functions of ePill subjects were able to find and employ.

To elicit subjective feedback on subjects' experiences gained while performing the tasks, we employed questionnaires to be answered after every task. Questionnaires consisted of questions regarding Perceived Usefulness, Perceived Ease of Use, Attitude Toward Use, and Behavioral Intention to Use [16]. We used three items for every construct since previous studies show that three items per construct deliver reliable results [17]. Feasible questions for our questionnaire where chosen from Lai and Li [17], translated to German, and slightly modified so that they target users' impression of ePill with respect to a specific task. We used all items of the Behavioral Intention scale of Lai and Li because they demonstrated high reliability (Cronbachs' Alpha = 0.94) [17]. Additionally, we used three of their six items concerning Perceived Usefulness (Cronbachs' Alpha = 0.95). We could not use all the items of the Perceived Ease of Use and Attitude Toward Use constructs because both contain a question starting with 'Overall'. These questions were not feasible for our study because extensive modifications would have been required to make these questions target only a specific task. We used two questions for both scales from Lai and Li [17] and supplemented them with other items that proved to be reliable in studies by Davis [18] and Vijayasarathy [19]. The questionnaires used a seven-point Likert scale (*strongly disagree*(1) to *strongly agree*(7)).

After all tasks, associated questionnaires, and further questions regarding socio-demographics had been answered, subjects received the System Usability Scale (SUS) [15] questionnaire. The SUS questionnaire targets users' overall impressions regarding ePill and results in a point estimate of usability, which can be used for comparison with other systems or results of subsequent usability studies of the same system. Although there are alternatives to the SUS, it is a good choice for general usability practitioners [20]. The SUS is a very simple-to-use questionnaire to measure usability of any tool or system [15]. It consists of ten items and measures usability with a score between 0 and 100. Additionally, the SUS is non-proprietary and thus cost effective [20]. Bangor et al. [20] investigated 200 studies which used the SUS to characterize a 'good' SUS score. Their results can be used as reference. We translated the original SUS items [15] and replaced the word 'the system' by the word 'ePill'.

2.3. Focus Groups

The goal of focus groups is to elicit a range of opinions regarding a certain object of investigation by collecting qualitative data. Focus groups can be a promising extension for usability studies because they facilitate learning more about users' impressions as well as elicitation of ideas for improvement [21].

Prior to conducting focus groups, the number of participants must be chosen. Literature recommends groups of six to ten participants, if participants are

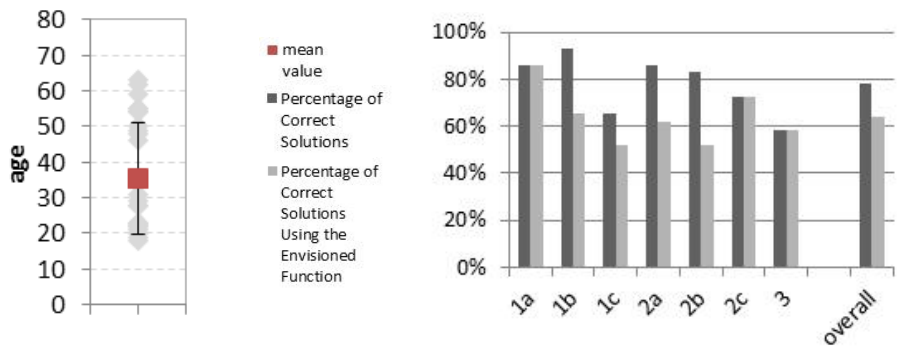

Figure 1. Age Distribution and Percentage of Correct Solutions

inexperienced with the topic [21]. Accordingly, we conducted focus groups with a target number of eight participants and conducted four focus group sessions with different participants. Two different focus group strategies are commonly used: Topic Guide and Questioning Route [22]. The Topic Guide is a strategy that can be developed very fast, is more spontaneous, allows the moderator to use previous comments in future questions, and is often used by professional moderators. The questioning route is a more structured strategy consisting of several types of questions asked in a fixed arrangement [22]. It is often used in academic research because it quickly provides precise answers and is easy to analyze. Hence, we employed the questioning route strategy.

3. Results

3.1. Sample Description

Four sessions were conducted with six to nine participants each. In total, 29 (female=12) subjects participated. Figure 1, left displays the age distribution of our sample (min=18, max=63, mean=35.31, std. dev.=15.75). Subjects had different professions (e.g., information systems, banking, law etc.) and 11 (37.93%) were

Table 1. Mean Value, Standard Deviation and Cronbach's Alpha for the Scales of Task Questionnaires

Task	scale	mean value	standard deviation	Cronbach's Alpha
1	Perceived Ease of Use	5.26	1.20	0.87
	Perceived Usefulness	5.53	1.21	0.90
	Attitude Toward Using	5.37	1.38	0.95
	Behavioral Intention to Use	4.75	1.53	0.95
2	Perceived Ease of Use	4.82	1.45	0.91
	Perceived Usefulness	4.95	1.50	0.90
	Attitude Toward Using	5.02	1.56	0.93
	Behavioral Intention to Use	4.70	1.69	0.96
3	Perceived Ease of Use	4.49	1.20	0.96
	Perceived Usefulness	4.97	1.21	0.96
	Attitude Toward Using	4.94	1.38	0.96
	Behavioral Intention to Use	4.62	1.53	0.96

Table 2. Correlation Matrix of SUS Score, Age and Percentage of Correct Solutions(PCS)

	SUS Score	Age	PCS
SUS Score	1		
Age	0.1341 (p = 0.496)	1	
PCS	**0.5078 (p = 0.006)**	**-0.5389 (p = 0.003)**	1

students. Besides gender and age, ten demographic questions measuring subjects experience with computers and health information were employed. Participants were randomly assigned to computers in a computer lab of the University of Cologne.

3.2. Usability of ePill

28 SUS questionnaires could be analyzed. The mean SUS Score was 64.91 with a standard deviation of 18.35 and a 95% confidence interval ranging from 58.11 to 71.71. Participants performed eight subtasks. Figure 1, right shows the percentage of correct solutions for every subtask. In addition, it shows what percentage of subjects used the functionality envisioned for a certain task. Overall, 159 of 203 tasks were solved correctly (77.83%). 130 tasks were solved correctly using envisioned functionality (64.04%). Table 1 presents results of the task questionnaires. Cronbach's Alphas of the four constructs range from 0.87 to 0.96. The highest mean values were achieved in task one (4.75 to 5.53) while task 2 and 3 achieved lower mean values for all constructs.

Single factor ANOVAs revealed no significant connection between task type and task questionnaire responses. Table 2 presents Pearson correlations between age, percentage of correct solution, and SUS score. Two correlations are significant (marked as bold in Table 2). Correlation between Age and Percentage of Correct Solutions is negative (p=.003. Correlation between SUS Score and Percentage of Correct Solutions is positive (p=.006). Eleven participants (37.93%) suffered from a chronic disease or allergy. However, a two sample t-test revealed no significant aspect between this aspect and the SUS Score ($t(26) = 0.71$, $p=.48$) or Behavioral Intention to Use ($t(27)=0.57$, $p=.58$). The demographic item "How often do you worry about your state of health?" (DEM10) was the only demographic item showing a significant correlation with subjects' usability assessments (Table 3). There is a significant, positive correlation ($p=.05$) between DEM10 and the SUS Score. We could not find any significant correlation between DEM10 and SUS Score, Perceived Ease of Use, Perceived Usefulness, Attitude Toward Using or Behavioral Intention to Use. Figure 2 presents comments frequently made in the focus groups, which are addressed in more detail in the discussion section.

4. Discussion

ePill's mean SUS Score of 65 indicates a marginal acceptability rating, and an adjective rating lying between 'OK' and 'GOOD' according to the results of Bangor et

Table 3. Correlations between DEM10, SUS Score and Task Questionnaire Scales

Item one	Item two	Correlation r	p-value
DEM10	SUS Score	**0.3873**	**0.046**
DEM10	Perceived Ease of Use	0.1175	0.560
DEM10	Perceived Usefulness	0.0790	0.695
DEM10	Attitude Toward Using	0.1505	0.454
DEM10	Behavioral Intention to Use	0.1347	0.503

Layout and appearance
"I was not able to find all functions. They should be more eye-catching."
"At the beginning, I did not understand how ePill works. It would be nice to have a tutorial which guides you through the first steps with ePill."
"The Usage of ePill has to be more intuitional."
"I like the possibility to decide which information is displayed."
The different views of ePill
"It should be described which view enables which functions."
"The term "expert view" is very confusing. For me, only a doctor is an expert in this context."
"The standard view did not enable me to use all functions."
The Interaction Function
"I did not understand the danger levels."
"I would like to have a traffic light system for the different danger levels in the Interaction Function."
Displaying general medical information
"After scrolling down, I was not able to see which medicament I am regarding right now. It would be nice to have a fixed bar which displays this information."
"The information is not well edited."
"I would like to have a short summary of every medicament's main effect in a few words."

Figure 2. Quotes from Focus Groups Concerning the Usability of ePill

al. [20]. Thus, ePill's usability offers potential for further optimization. However, it has to be kept in mind that the SUS Score was rated after only one hour of system use. Participants reported having problems using ePill because they did not know which functions were offered by ePill and claimed that use of ePill should be more intuitive. Especially, supplementary services such as drug interaction discovery or medication comparison have not been used by many subjects.

These observations were confirmed during focus group sessions. Subjects reported that they were not able to find all functions while using the Standard View (ePill offers three different GUI design presets: standard view, expert view, and custom view [5]). Subjects wanted supplementary services to be presented in a more eye-catching way. Moreover, many subjects reported that the term 'Expert View' was confusing because they did not consider themselves as experts on medication. To alleviate this problem, the view design will be overhauled: Expert View will be renamed to be less misleading (e.g., Advanced View). Furthermore, more information on the different views will be offered on the start page enabling users to make a well-grounded decision when choosing a view.

Comparison of task successes with and without employment of envisioned functionality shows that subjects found alternatives to some functions. To increase usability of ePill, the functions where alternatives were not found will be made more salient. The positive correlation between DEM10 and the SUS Score indicates that people worrying more about their state of health think higher of ePill than people without such worries. This indicates that especially health-conscious people are interested in using patient-centered health IT applications like ePill. Further research is needed to investigate this assumption because our data did not reveal any significant correlation between DEM10 and the task questionnaire responses. During our investigation it became clear that some aspects of ePill are in need of improvement. Elimination of these problems will lead to a patient-centered web application that is easier to use and more beneficial for its users. Overall, the basic concept of ePill was positively conceived by subjects. However, prior to evaluating ePill's effectiveness in terms of medical outcomes in a large-scale clinical trial [4] further usability studies will be conducted with subsequent iterations of ePill to ascertain whether implications derived from the conducted usability study lead to a more usable program and to

reduce potential bias introduced by lacking usability for assessment of ePill's effect on medical outcomes.

The positive correlation between Percentage of Correct Solutions and SUS Score indicates that improvement of usability is worthwhile. If users are able to use core functionality offered by health IT applications they are also more likely to give higher usability ratings. Since this is a precondition for repeated use of health IT applications, it is also very critical for success and usefulness of health IT applications. The significant negative correlation between Age and Percentage of Correct Solutions shows that older users are not able to use health IT applications as easily as younger users. Since older people are more likely to be in need of medication, especially in the case of ePill, more effort is required to facilitate use of health IT applications for older users. Investigation of ePill's usability enabled us to derive implications for future development [23]: A short interactive tutorial will be implemented to show most important functionality to first time users. A tutorial could additionally increase users' confidence by explaining more about data sources and developers´ intentions. Alternatives to a tutorial could be a more detailed description of core functionality on the start page or asking users which objective they pursue by using health IT applications and giving them hints depending on their answer. Further usability studies will show whether a tutorial, provision of more information on getting started, and fixing further issues pointed out in the usability study will be sufficient to improve discovery of available functionality or whether further improvements are required.

To sum up, employing the SUS in usability studies of patient-centered health IT applications is useful to derive an approximate, overall classification of a system's usability. The resulting value can be translated into a verbal classification by using the scheme of Bangor et al. [20]. In addition, SUS scores can be used to compare systems with similar systems or subsequent versions. Tasks in combination with associated questionnaires inherently promote subjects' familiarity with the assessed system and enable a more detailed analysis of usability. Thus, specific aspects that cause difficulties for users and are consequently responsible for low perceived usability can be identified. Focus group discussions go even one step further and allow participants to exchange opinions and be creative so that ideas for improvement of deficient aspects are developed. Applying our mixed-method usability study design to assess usability of ePill can be considered successful: We were able to identify specific usability problems and developed various ideas to improve usability. Reliability of the used items and popularity of the SUS score show that these are feasible approaches to analyze usability. In addition, focus groups deliver useful feedback concerning usability and potential for improvement. Thus, combination of these methods seems to be adequate for assessing usability of patient-centered health IT applications. Usability evaluation is however an inherently iterative process: Amelioration of usability deficiencies identified in one study will have to be verified through a follow-up study, which will most likely identify further deficiencies. Yet, usability studies are worthwhile to ensure tailoring of health IT offerings to users' needs and preferences, counteract attrition, and weed out detrimental effects on the potential of health IT applications to improve users' state of health and quality of life.

References

[1] Joyce A. Cramer: Partial Medication Compliance: The Enigma in Poor Medical Outcomes. In: The American Journal of Managed Care. No. 2. Vol. 1, 1995, pp. 167-174.

[2] Adriana Chubaty, Cheryl A. Sadowski, Anita G. Carrie: Typeface Legibility of Patient Information Leaflets Intended for Community-dwelling Seniors. In: Age and Ageing, No. 4, Vol. 38, 2009, pp. 441-447.

[3] Almut G. Winterstein, Stephan Linden, Abigail E. Lee, Erica M. Fernandez, Carole L. Kimerlin: Evaluation of Consumer Medication Information Dispensed in Retail Pharmacies. In: Archives of Internal Medicine, No. 15, Vol. 170, 2010, pp. 1317-1324.

[4] Tobias Dehling, Ali Sunyaev: Improved Medication Compliance Through Health IT: Design and Mixed Methods Evaluation of the Application ePill. 34th International Conference on Information Systems, Milan 2013.

[5] Tobias Dehling, Ali Sunyaev: Architecture and Design of a Patient-Friendly eHealth Web Application: Patient Information Leaflets and Supplementary Services. In Proceedings of the 18th Americas Conference on Information Systems, Seattle, Washington, U.S., 2012.

[6] Nicol Nijland, Julia EWC van Gemert-Pijnen,Saskia M. Kelders, Bart J. Brandenburg, Erwin R. Seydel: Factors Influencing the Use of a Web-Based Application for Supporting the Self-Care of Patients with Type 2 Diabetes: A Longitudinal Study. In: Journal of Medical Internet Research. No. 3, Vol. 13, 2011, e71.

[7] Larry Goldberg, Bettijoyce Lide, Svetlana Lowry, Holly A. Massett, Trisha O'Connell, Jennifer Preece, Whitney Quesenbery, Ben Shneiderman: Usability and Accessibility in Consumer Health Informatics. In: American Journal of Preventive Medicine, No. 5S2, Vol. 40, 2011, pp. S187-S197.

[8] Rosy Tsopra, Jean-Philippe Jais, Alain Venot, Catherine Duclos: Comparison of Two Kinds of Interface, Based on Guided Navigation or Usability Principles, for Improving the Adoption of Computerized Decision Support Systems: Application to the Prescription of Antibiotics. In: Journal of the American Medical Informatics Association, Vol. 21, 2014, e107-e116.

[9] Timo Jokela, Netta Iivari, Juha Matero, Minna Karukka: The Standard of User-Centered Design and the Standard Definition of Usability: Analyzing ISO 13407 against ISO 9241-11. Proceedings of the Latin American Conference on Human-Computer Interaction, New York, NY, U.S., 2003, pp. 53-60.

[10] Gunther Eysenbach, Christian Köhler: How do Consumers Search for and Appraise Health Information on the World Wide Web? Qualitive Study Using Focus Groups, Usability Tests, and In-Depth Interviews. In: British Medical Journal. Vol. 324, 9 March 2002, pp. 573-577.

[11] Catherine H. Yu, Robinder Bahniwal, Andreas Laupacis, Eman Leung, Michael S. Orr, Sharon E. Straus: Systematic Review and Evaluation of Web-accessible Tools for Management of Diabetes and Related Cardiovascular Risk Factors by Patients and Healthcare Providers. In: Journal of the American Medical Informatics Association. Vol. 191, 3 February 2012, pp. 514-522.

[12] Patrick G. Kelley, Joanna Bresee, Lorrie F. Cranor, Robert W. Reeder: A "Nutrition Label" for Privacy, 5th Symposium on Usable Privacy and Security (SOUPS) 2009, July 15-17, Mountain View, CA, USA.

[13] International Organization for Standardization, 9241-11: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs)-part 11: Guidance on Usability, The Organization, 1998.

[14] Jakob Nielsen: Usability Engineering, Academic Press Inc., San Diego, 1993.

[15] John Brooke: SUS – A Quick and Dirty Usability Scale. In: Usability Evaluation in Industry, London: Taylor & Francis, 1996.

[16] Fred D. Davis: A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results, PhD Thesis, Massachusetts Insitute of Technology, 1985.

[17] Vincent S. Lai, Honglei Li: Technology Acceptance Model for Internet Banking: An Invariance Analysis. In: Information & Management, No. 2, Vol. 42, 2005, pp. 373-386.

[18] Fred D. Davis: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. In: MIS Quarterly, No. 3, Vol. 13, 1989, pp.319-340.

[19] Leo R. Vijayasarathy: Predicting Consumer Intentions to Use On-Line Shopping: The Case for an Augmented Technology Acceptance Model. In: Information & Management, No. 6, Vol. 41, 2004, p. 747-762.

[20] Aaron Bangor, Philip T. Kortum, James T. Miller: An Empirical Evaluation of the System Usability Scale. In: International Journal of Human-Computer Interaction, No. 6, Vol. 24, 2008, pp. 574-594.

[21] Richard A. Krueger, Mary A. Casey: Focus Groups: A Practical Guide for Applied Research, 3. ed., Sage Publications, Inc., Thousand Oaks, 2000.

[22] Richard A. Krueger: Developing Questions for Focus Groups, Sage Publications, Thousand Oaks, 1998.

[23] Till Blesik, Tobias Dehling, Ali Sunyaev: Improving Comprehensibility of Medical Information – Proof-of-Concept for Patient Information Leaflets. In: Proceedings of Multikonferenz Wirtschaftsinformatik, Paderborn, Germany, 2014.