

An evaluation of patients' experienced usability of a diabetes mHealth system using a multi-method approach



Mattias Georgsson^{a,b,*}, Nancy Staggars^{a,c}

^a Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA

^b Faculty of Computing, Blekinge Institute of Technology, Karlskrona, Sweden

^c College of Nursing, University of Utah, Salt Lake City, UT, USA

ARTICLE INFO

Article history:

Received 8 June 2015

Revised 14 October 2015

Accepted 19 November 2015

Available online 27 November 2015

Keywords:

Usability

mHealth

Diabetes

Multi-method evaluation

Framework analysis

Usability Problem Taxonomy

ABSTRACT

Objective: mHealth systems are becoming more common to aid patients in their diabetes self-management, but recent studies indicate a need for thorough evaluation of patients' experienced usability. Current evaluations lack a multi-method design for data collection and structured methods for data analyses. The purpose of this study was to provide a feasibility test of a multi-method approach for both data collection and data analyses for patients' experienced usability of a mHealth system for diabetes type 2 self-management.

Materials and methods: A random sample of 10 users was selected from a larger clinical trial. Data collection methods included user testing with eight representative tasks and Think Aloud protocol, a semi-structured interview and a questionnaire on patients' experiences using the system. The Framework Analysis (FA) method and Usability Problem Taxonomy (UPT) were used to structure, code and analyze the results. A usability severity rating was assigned after classification.

Results: The combined methods resulted in a total of 117 problems condensed into 19 usability issues with an average severity rating of 2.47 or serious. The usability test detected 50% of the initial usability problems, followed by the post-interview at 29%. The usability test found 18 of 19 consolidated usability problems while the questionnaire uncovered one unique issue. Patients experienced most usability problems (8) in the Glucose Readings View when performing complex tasks such as adding, deleting, and exporting glucose measurements. The severity ratings were the highest for the Glucose Diary View, Glucose Readings View, and Blood Pressure View with an average severity rating of 3 (serious). Most of the issues were classified under the artifact component of the UPT and primary categories of Visualness (7) and Manipulation (6). In the UPT task component, most issues were in the primary category Task-mapping (12).

Conclusions: Multiple data collection methods yielded a more comprehensive set of usability issues. Usability testing uncovered the largest volume of usability issues, followed by interviewing and then the questionnaire. The interview did not surface any unique consolidated usability issues while the questionnaire surfaced one. The FA and UPT were valuable in structuring and classifying problems. The resulting descriptions serve as a communication tool in problem solving and programming. We recommend the usage of multiple methods in data collection and employing the FA and UPT in data analyses for future usability testing.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Diabetes is a condition affecting 29.1 million people in the United States [1] with concomitant health care expenditures of an estimated \$198 billion [2]. The most common form of diabetes

is Type 2 (T2DM) affecting a majority, 90%, of those with the condition [3]. A healthy diet, regular physical exercise, and maintaining a normal body weight in addition to medication treatment are seen as important [3] whereas poorly regulated glycemic measurements and poor self-management practices are contributors to a worsening condition [4,5]. Due to its complexity, Type 2 diabetes puts heavy demands on both patients and providers [6].

Different types of support systems for diabetes self-management have been developed recently using Information

* Corresponding author at: Faculty of Computing, Blekinge Institute of Technology, 371 79 Karlskrona, Sweden. Tel.: +46 708257134.

E-mail address: mattias.georgsson@bth.se (M. Georgsson).

and Communication Technology (ICT) [7,8]. Studies show that mobile health (mHealth) technology has been particularly successful in improving glucose management for these patients [9–11] and has aided in regulating lifestyle changes [11]. Recent interventions having beneficial effects are, for example, those focusing on self-coaching by allowing users to interact with algorithm-based systems that provide individually tailored messages [12] and feedback messages with automatically updated text messages based on clinical guidelines and patients' lifestyles [13].

Although mHealth applications have been beneficial in diabetes self-management, researchers mention the need for improved usability to allow patients to feel confident in their interactions with mHealth applications [14]. Unfortunately, relatively few studies exist on mHealth usability [15]. Of the available studies, most lack methodological rigor [16,17] and use a single method for data collection [17,18]. Also, qualitative data analyses in mHealth studies typically lack a structured approach, making study reproducibility difficult, especially when using multi-method approaches for large data sets with multiple researchers. No usability studies are yet available using the Framework Analysis (FA) method or the Usability Problem Taxonomy (UPT) to address this gap. Our approach acknowledges these current limitations and tests the feasibility of a multi-method approach for data collection and novel, structured techniques for data analyses.

Methodological improvements in mHealth usability studies could result in more comprehensive identification of usability issues, more specific redesign recommendations based upon user-centered data, more reliable data analyses and they could potentially improve the reproducibility of results across studies. For users, improved mHealth designs could result in improved interactions, user performance, adoption of mHealth applications and perhaps even increased adherence to suggested interventions for chronic care conditions.

1.1. Review of the literature

Recent reviews of mHealth usability studies point to the need for more rigorous studies and the use of multiple methods to help validate findings [16,17]. Authors commented that each usability testing method makes its own unique contribution to the overall identification of usability issues [17]. Thus, a robust usability evaluation should employ a combination of methods [16,17]. Specifically, Zapata et al. emphasized the need for (1) employing two or more different types of usability methods, (2) using more standardized methods and tools in mHealth evaluations and (3) using open-ended inquiry methods and qualitative analysis to identify more comprehensive user concerns and recommended that future researchers conducting usability evaluations need to standardize processes and use a combination of more than one method [17,18].

1.1.1. Multi-method approach to data collection

Multiple methods in data collection can increase the depth of inquiry while improving the reliability and validity of findings [19]. The use of multiple methods also assists with methodological triangulation because it can allow for a more comprehensive understanding of a phenomenon [20,21]. This type of triangulation was first identified by Patton who advocated data collection using observation, field notes and interviews [20,22]. These kinds of qualitative methods produce rich data [20,23] and are especially suitable in health informatics research [24]. Pertinent to mHealth, multiple methods in data collection could allow for a more comprehensive understanding of the mHealth user experience and pinpoint specific usability problems for redesign.

In the current study, we test the feasibility of overcoming common methodological issues in data collection during a mHealth usability study. We completed a multi-method approach to data

collection with user-based testing, Think Aloud protocol, open-ended interviews and a short answer questionnaire.

1.1.2. Structured approaches for data analyses

The main shortcomings of qualitative research methods are a lack of standardization in data analytic techniques, meaning that replicating the results of analyses can be difficult [25]. Two methods can assist in overcoming this issue, the Framework Analysis (FA) method and the Usability Problem Taxonomy (UPT) classification scheme. Neither has yet been employed in mHealth or other health informatics usability evaluations to our knowledge.

The FA, originally created by Richie and Spencer in 1994, is a standardized yet somewhat flexible framework for analyzing qualitative data. Its purpose is to assist with the analysis of descriptive, textual source data to produce reliable and valid qualitative results [26,27]. Importantly, the FA is seen as scientifically robust [28]; it has been used in social sciences research and to a lesser extent in health care research [28,29]. Recent applications of the FA have been mainly in health care (nursing research) and in multidisciplinary studies to manage large sets of qualitative data [28,29].

The FA consists of five stages (1) familiarizing oneself with the data, (2) identifying a thematic framework to be used, (3) indexing and applying the framework to the data, (4) charting the data and (5) mapping and interpretation [26]. This analysis method is especially useful in organizing, reducing, and interpreting data because it provides clear steps to follow and it produces more structured output with summarized data [29]. Moreover, the process can be used by several researchers simultaneously [28] such as multidisciplinary team members [29]. The FA promotes, in particular, data/decision transparency [28]. It is useful for both experienced and less experienced coders due to its available audit trail [28]. The framework can accommodate different analytic tools such as paper, post-it notes, Microsoft Word, or NVIVO [28]. Primary difficulties with this method are those usually inherent to qualitative analysis e.g., it is time consuming and requires a committed manner of analysis [29].

Finding ways to structure and classify usability problems has been explored within the Human–Computer Interaction (HCI) field. Keenan et al. [30] found no framework for classifying usability problems on an individual problem level, particularly using various perspectives that would allow problems to be compared, analyzed and described. Such an analysis would provide better information for problem correction [30]. These authors thought a new approach would either have a trained individual examine the whole set of usability problems to look for trade-offs, contradictions, and consistency issues [31], or an expert might think about the problems from a more global perspective to look for problem clusters [32]. Keenan et al. considered using heuristic analysis [33,34], but found that technique lacked sufficient problem distinguishability, mutual exclusiveness and specificity [30,35]. Thus, the UPT was built empirically using over 400 usability problem descriptions collected on real-world development projects [30].

UPT is a classification scheme and framework for characterizing usability problems according to their dimensions, providing a clear structure in usability problem definition. It was initially used to classify usability problems found on graphical user interfaces with textual components where usability problems were easily detected, classified, and analyzed [30]. The UPT is based on the notion that usability problems should be examined from two perspectives: the task-artifact approach proposed by Carroll et al. [36] and the Object-Action Interface Model by Shneidermann [37] to enhance problem definition. The *artifact component* defines usability problems arising when the user interacts with the interface while categories under *task component* focus on usability problems that surface when a user moves through a task [30]. Problem

classification occurs from both an *artifact* and a *task* component perspective and is then divided into five primary categories: *Visualness*, *Language* and *Manipulation*, *Task-mapping* and *Task-facilitation* (see Fig. 1).

The UPT classification could also aid developers in several ways (1) the categories are based on problem characteristics versus only problem type, meaning classifications occur on two levels (artifact and task), (2) UPT offers a method of identifying problem clusters using different ways and levels of categorizing problems, (3) it is beneficial for examining problem sets at varying levels of abstraction, and (4) problems distributed across UPT categories make visible the kinds of issues encountered most often. It can thus be used by developers to assess problem scope and frequency as well as defining the types of problems that both global and local solutions may be considered. The UPT has been used only sparingly since its creation, but its classification methods could provide an excellent description of the dimensions in detected usability problems. We concluded the UPT is a lesser known but robust technique.

1.2. Study aim

Our purpose was to test the feasibility of (1) using a multi-method approach for data collection during a mHealth usability evaluation and (2) applying structured approaches to data analyses by using the FA method and UPT. These comprise a novel, multi-method approach to data analyses and, to our knowledge, are the first applications of both FA and UPT in health informatics research.

2. Materials and methods

2.1. The evaluated system

The system evaluated in this study is an interactive, SMS-based mobile intervention for patients with diabetes, designed as a personalized self-care management tool. It is a commercially available tool in current use in several organizations. The system consists of a combined mobile phone solution and web service. Patients interact by either sending in or being prompted to send in their Type 2 diabetes, self-management values, e.g., morning blood glucose via text message. Using the web service, they can enter and/or review their results across various parameters. Patients can, for example,

obtain their glucose readings or blood pressure values visualized in a meter format, see medication adherence levels, exercise and weight progress and view scheduled medical appointments. When sending in their measurements patients receive personalized coaching and tailored responses from the system to track the progress of their disease. The web portal contains different features, main views and sub-views for patients to perform different actions and track results (see Fig. 2). In this study we concentrated our evaluation mainly on the web service part of the solution because of its inherent complexity and because it had no previous usability evaluation.

2.2. Study sample and setting

Institutional Review Board approval was obtained (WIRB, Olympia, Washington, USA). A larger randomized controlled trial for a diabetes mHealth intervention study involving 18 primary care clinics in the Salt Lake City metropolitan area served as the study population for our usability study. A set of 2317 patients met the inclusion criteria for the larger study [38]. Ten patients were randomly selected from this larger randomized controlled trial on the mHealth intervention by computer randomization and invited to take part in our usability evaluation. Inclusion criteria for our usability study sample included (1) patients diagnosed with Type 2 diabetes (2) no cognitive impairment; (3) familiarity and some knowledge and use of computers, the Internet, and cell phone; and (4) the ability to speak and understand the English language. Patients had no previous exposure to the mHealth web system evaluated in our study. The evaluation sessions were conducted in a quiet laboratory setting at HealthInsight, a U.S. Beacon Community in Salt Lake City, Utah.

2.3. Data collection methods

Usability can be evaluated by several different methods. Inspection methods such as heuristic evaluation [39,40], and cognitive-walk through [41,42] are expert methods meaning that experts go through the system to identify usability issues. Think Aloud protocol [43,44] is a user-related method for evaluating usability where users express their perceptions out loud as they interact with the system. Other user-related methods include administering

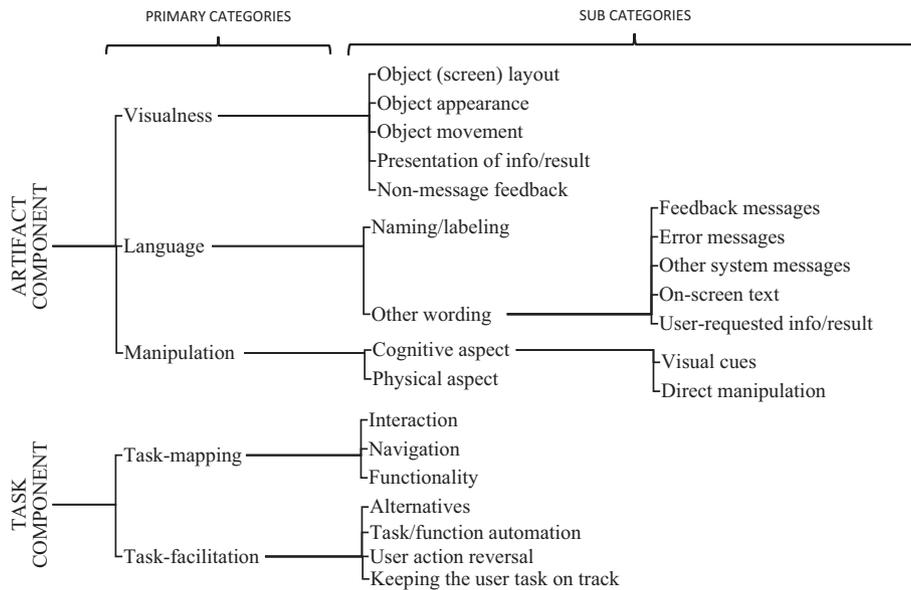


Fig. 1. Classification of categories according to the Usability Problem Taxonomy [30].

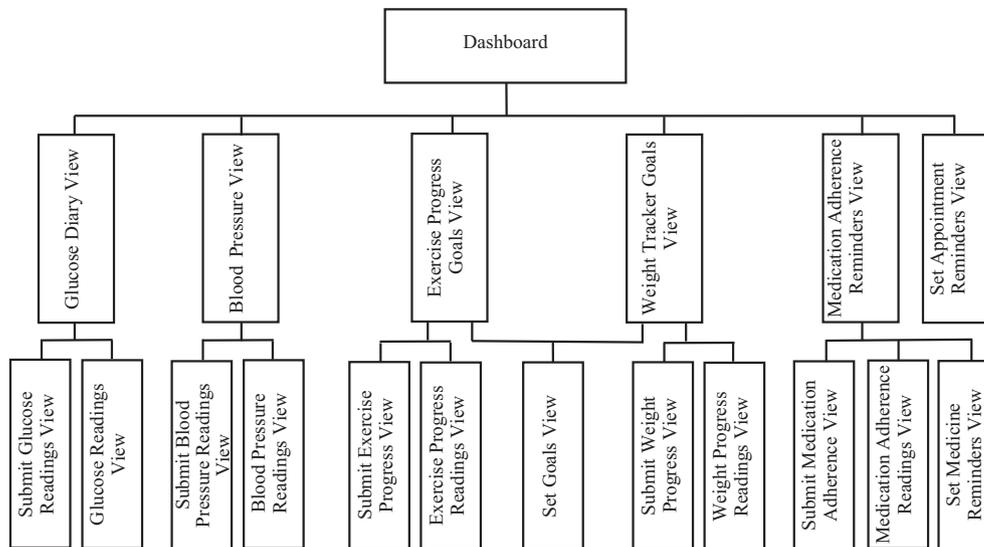


Fig. 2. Overview of system views.

in-depth interviews and questionnaires about patients' experiences with the system [24]. The main goal of user-related methods is to involve actual users in the evaluation and obtain their perspectives [45], to gain insight into how the intervention needs to be adapted for different users' abilities and experience levels [46,47] and to identify usability problems for correction.

2.3.1. Usability test with think aloud protocol

Think Aloud is a usability assessment method commonly employed to determine users' thoughts and opinions while they perform a list of specified tasks with a system. The method originated in 1984 in psychology when Simon and Ericsson thought of verbal reports as data. Revised in 1993 [48], the technique is well established within the Human Factors field [39]. Think Aloud asks users to talk aloud during their interactions, to express their reactions and thinking and to explain what they are doing as they perform specific, representative tasks [24]. The resulting data are normally audio- and/or video-recorded and/or an observer takes thorough, written notes [49]. Minimal intervention from the usability tester assures users' thought processes are not interrupted except to remind them to keep talking [49]. The focus is on understanding users' decision making processes and on how users experience the system in their own words [39,49]. Because the method provides extensive, detailed data, only a small sample of five to eight users is needed in usability testing to detect 80–85% of usability problems [50–52] to gain a thorough understanding of task behavior [48] and to identify the main usability problems [53]. Representative tasks for the specific domain are also essential, and they should be as realistic as possible [54]. Also, testing should be conducted in the actual user's context or one as close to the natural environment as possible [49,55]. Authors indicate that Think Aloud provides complete and detailed descriptions of patients' thought processes during system interactions and the technique generates many usability problems [56].

2.3.2. Representative tasks

Tasks for this study were based on common patient user interactions with the system; they were disease-specific and had varying levels of difficulty to simulate patient usage in a clinic or at home. Tasks were validated by a panel consisting of a physician and a nurse whose specialties were diabetes, a public health professional with chronic patient intervention systems expertise, and a diabetes patient. The specific tasks patients had to perform

consisted of (1) uploading glucose values into the system, (2) interpreting a glucose measurement in a graph view, (3) correcting a recorded glucose measurement value, (4) exporting glucose measurement value trends to a PDF to simulate material to take to a provider visit, (5) interpreting a blood pressure measurement in a graph view, (6) setting personal tracking goals for exercise and weight, (7) setting medication reminders, and (8) setting a physician appointment reminder.

2.3.3. In-depth interview and usability questionnaire

Both interviews and questionnaires have been used extensively in usability research to determine users' opinions about the difficulties they experience in an evaluated system [49]. In our study patients completed both an open-ended interview and a post-experience questionnaire [57]. The open-ended interview guide asked users to talk about aspects of the system with good or poor usability. This kind of interview is especially useful in uncovering comprehensive information from participants [58]. The three questions in the interview asked patients to comment on sections of the system they thought were well designed, sections that were inadequately designed and any further comments they might have about system usability (see Appendix A).

Patients also completed a post-interaction questionnaire on the mHealth system (see Appendix B). The first section consisted of short answer questions about: patients' IT/computer, mobile phone, and internet experience and use; their experience and perceptions about web and mobile service systems in health care; what they thought about the specific mHealth system evaluated in the session in terms of usability and any further comments they might have about these topics. Patients were also asked to rate their preferences in technology usage for work and leisure time using a Likert scale with 4 points ranging from strongly agree (1) to strongly disagree (4).

The second part of the post-questionnaire included open-ended questions about the specific system patients used in this study. Patients specified in writing their thoughts about what they found difficult and easy about the system and its navigation, they listed usability/user experiences they found satisfying or dissatisfying and then a final question asked for any further comments about the system.

Both the interview guide and post-experience questionnaire were assessed for face validity by a panel of three health care

professionals, three usability experts and one patient. The resulting format was finalized via discussion and consensus.

2.4. Study procedure

First, patients were asked for informed consent. Then, patients were walked through the different steps of the evaluation procedure and asked if they had any questions. The evaluation started with patients filling out a brief demographic questionnaire. Topics included age, gender, educational level, occupation, and how long they had been diagnosed with diabetes.

Next, standardized training was performed to simulate an actual patient educational process in a health clinic. This was important to decrease individual variability and to ensure that patients all had the same information about the system [59] because none had used the system before. After the training was completed by the first author, users interacted with the system on their own to get familiar with it for an average of about 10 min.

The Think Aloud and usability evaluation session was conducted by the first author. Patients were given a booklet outlining the specific tasks to perform. During the session patients were asked to think aloud as they completed the prescribed tasks in the system. If they became silent, the researcher encouraged them by asking what they were thinking or by clarifying actions, but otherwise any other interference with patients' thought processes was avoided. Interactions were digitally audio- and video recorded using Morae™ software [60]. The recording showed patients' navigation on screens, their facial expressions and captured their voices. The researcher made observations and notations about the individual task performances directly into Morae™.

Afterwards, patients were interviewed about their experiences and perceptions about the system, by the first author, using the in-depth interview guide with interview topics. Patients were provided the opportunity to express freely what they thought was easy and difficult with the system and add any further comments. This session was also digitally audio-recorded and lasted for 15–20 min. Finally, patients completed the post-test questionnaire on their perceptions about the usability of the system.

The complete testing procedure for all the steps averaged approximately two hours with a range of 1.5–2 h. Patients received a gift card for \$20 after completing the session.

2.5. Data analyses steps utilizing the Framework Analysis (FA) method and the Usability Problem Taxonomy (UPT)

The audio- and video recordings from the sessions, Think Aloud comments and observations, post-test interview and post-test questionnaire usability data were transcribed, checked for accuracy and imported into Nvivo™ 10 Qualitative Data Analysis Software [61]. Data analyses included content analysis in addition to applying the FA and UPT.

Coding and analysis included the five steps in the FA method [26] (1) data familiarization, (2) identifying themes and framework used and (3) classifying usability problems using the UPT, (4) the results were organized or charted into the place of occurrence within the mHealth application and then (5) mapped and interpreted. Descriptive statistics were used to summarize data.

The first author transcribed, imported the data and completed the initial coding which were verified by the second author. Both authors conducted step 2–5 in the analysis together.

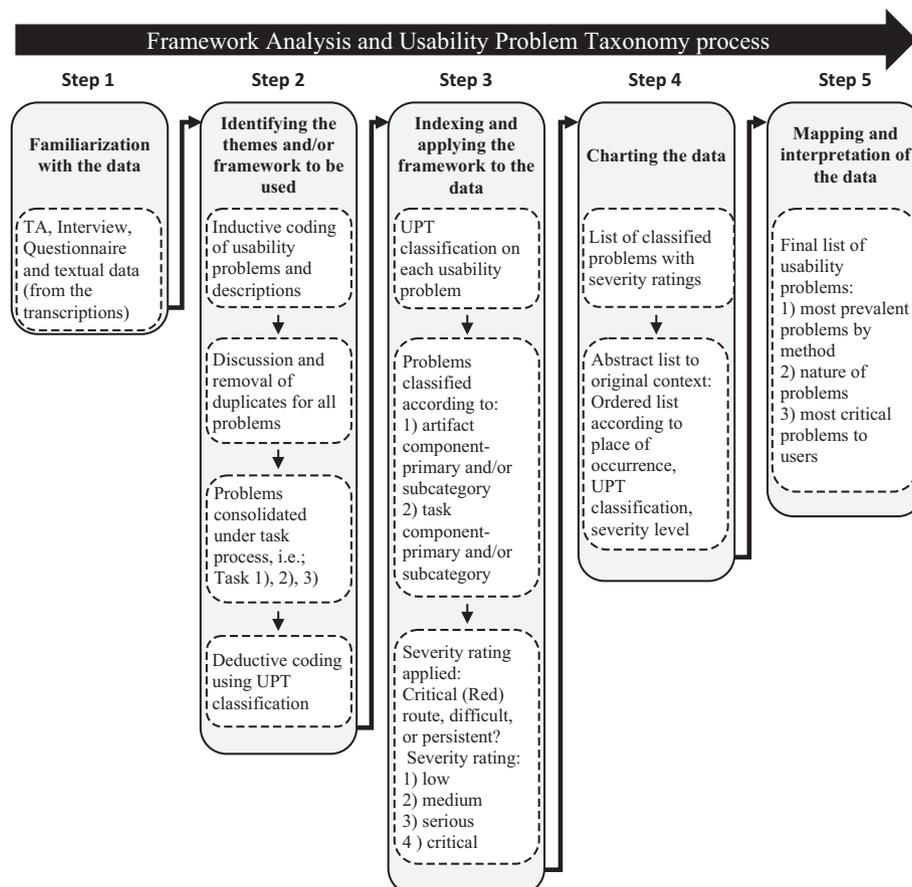


Fig. 3. Application of the FA method and UPT.

See Fig. 3 and the accompanying description of the analysis process below.

(1) Familiarization with the data

The first step was to become familiar with the transcribed textual data through immersion. This occurred by reading the uploaded textual documents several times.

(2) Identifying the themes and/or framework to be used

The second step was to identify themes and apply a framework to code the key issues. In the original FA method, researchers can select either an inductive or deductive content analysis approach to perform the coding. We modified the FA method slightly. We used FA to generate usability problems inductively across the various data collection methods, but we used the UPT deductively as well to classify the problems (see Fig. 3). The usability problems can be considered themes derived from the data. After a discussion and removal of duplicates, we consolidated the usability issues under their appropriate tasks. The application of the UPT is

described in more detail below, and we provide examples of two coded usability problems in Table 1.

(3) Indexing and applying the classification to the data

The third step consisted in indexing and applying the UPT classification framework as well as assigning severity ratings to our final usability problem list. We classified each usability problem into an artifact component (*Visualness, Language and Manipulation*) and/or a task component (*Task-mapping and Task-facilitation*) proceeding in the classification scheme as far as possible. The categories and subcategories within the artifact and task components at any level are mutually exclusive resulting in one final categorization [30]. An example of a classified usability problem from our data is shown in Table 2.

For the severity rating, we used a process defined by Travis that asks three questions about each usability problem (1) Does the problem occur on a red route; i.e., is it a frequent or critical task the system needs to support? (2) Is the problem difficult for users to overcome? and (3) Is the problem persistent and does it keep recurring [62]? Ratings were assigned using the scale (1) low, (2)

Table 1
Examples of raw data from transcripts used for coding usability problems.

Usability problem description	Sample comments and notes		
	Think aloud usability test	Interview	Questionnaire
(1) Difficult knowing to exit the table view and click the "Add data" button to be able to add in a new value	<p>(Observation) Subject finds it hard to remember where to go after the value has been deleted in order to insert the new, corrected glucose value. (close the current table view, choose the "Add data" button to get to the "Add data" view instead)</p> <p>"There really should be a button back on the previous screen to...because it is a little bit confusing..." (Observer comment/s: The subject is circulating with the cursor around the "Add data" button and refers to the Glucose Readings View)</p> <p>"There should be a button somewhere here that would tell you how to add...but there is no...I can't find a button here that says to add...so it's a little bit difficult..."</p>	<p>"first you have to delete it and then you have to exit out of that back to the dashboard and then click on add if I remember right...um... but that seemed a little awkward, like excess steps, too many steps"</p> <p>"so what was cumbersome for me was to change the date or values, an appointment or glucose readings to where... rather than just hovering over it and editing it with the new value you need to delete it first and then go out of that dialogue box into an add box, it seems like it was several steps required to just do one process, it wasn't that continuous flow"</p> <p>"if they had the icons there that would be delete and then add right next to each other that would make it a lot easier so you wouldn't have to really change screens and then have to go back, to me that's that doesn't make any sense"</p>	<p>Going from screen to do one procedure was a bit frustrating</p> <p>When changing/updating information I need to go into a few other screens to do so. It was not simple to do</p> <p>There should have been an add button on the same page. It made no sense to have to change screens when everything could have (should have) been on the same page</p>
(9) Difficult to know to select "Delete data" for the table view to change, delete, export, or print the value list	<p>(Observation) Subject finds it hard to remember, understand that one has to press the "Delete data" button in order to get to the table view with values and there choose the "Export tab" to export data</p> <p>"So what's really weird is that in order to export anything um...you have to click on delete data which does not make sense...at all"</p> <p>"Oh you are going to delete data...which by the way makes absolutely no sense" (Observer comment/s: The subject comments this before clicking the delete data button before exporting the values on the Glucose Reading View)</p> <p>"You need to go to delete data which I don't know why...it should be save instead" (Observer comment/s: The subject comments this before clicking the delete data button before exporting the values on the Glucose Reading View)</p>	<p>"I was going to export data but it said to delete when I see delete it's like a big red flag you know you always hope that you... that you get that question are you sure you want to do this (?) because once it's gone it's pretty much gone."</p> <p>"...um the first thing I noticed on the export page was I just thought it was strange that you had to go to delete and so I got...so I got a little wrapped around that...in my little brain I was thinking I wonder why you had to... doesn't make sense to me you know"</p> <p>"I mean yeah that button is very much not... I mean for somebody who's nervous about computers anyway they'd never click on that, they'd be worried they were going to erase something... very non-intuitive"</p> <p>"I don't want to delete it, I want to print it...so choosing the wrong terminology"</p>	<p>The wording was difficult like having to go to "delete" when I was to export</p> <p>I did not like that I had to use the delete button to get to the export data</p> <p>When exporting data I needed to click the "delete" button. It should be a simple Export button to select</p>

Table 2

Example of usability problem classified according to the Usability Problem Taxonomy process.

Usability problem description	Artifact component	Classification
	Task component	
It is not possible to put in time in minutes or half hours for the medication reminder – only the whole hour	Manipulation	Partial Classification (PC) Full Classification (FC)
	Task-mapping → Functionality	

medium, (3) serious or (4) critical for each problem [62]. The resulting severity scores were averaged per system view and for the whole system.

(4) Charting the data

This step consisted of abstracting our final list of usability problems back into their original context as is consistent with the FA method. We arranged the order according to the problem's place of occurrence, classification and severity level. Descriptive statistics were used to summarize issues within each method and per patient.

(5) Mapping and interpretation of the data

The final step involved mapping and interpreting the usability problems. After the resulting list of problems were charted, we were able to identify the most prevalent problems and their severity ratings in their respective views. By doing this, we could determine the nature of the problems and what their classifications implied. The latter served as guidance for designers in correcting specific usability issues.

3. Results

The sample of 10 patients included six women and four men (see Table 3). Most were between the ages of 40–59 with a range of 40–69. The majority were university educated and half were employed. Sixty percent had a Type 2 diabetes diagnosis of five or more years.

The sample used technology regularly. For example, 80% used a computer daily, 70% used the internet daily, and 90% indicated they used their mobile phone to make and receive calls every day. A majority agreed or strongly agreed that using IT/Computers and mobile phones in health care is a positive development and thought it was positive for their own diabetes self-management.

3.1. Usability problems

A total of 117 initial usability problems were detected by the ten patients across the different data collection methods where 59 problems were detected during the usability test, 34 in the post-interview and 24 using the questionnaire. The problems were consolidated into a list of 19 unique usability issues (see Appendix C). The average severity rating for the whole site was 2.47 or serious.

Sample problem descriptions are presented in Table 4. They are organized according to the most critical issues and listed in their place of occurrence.

The usability problems and severity ratings were the highest for the *Glucose Readings View* which had eight (consolidated) problems and a severity rating of three. The *Glucose Diary View* and *Dashboard* had three problems each. The *Glucose Readings View*, the *Glucose Diary View* and the *Blood Pressure View* all had severity ratings of three. The views with the highest number of problems were those with several steps in a task. These were vitally

Table 3

Participant characteristics.

Participant characteristics (N = 10)		n (%)
Gender	Male	4 (40)
	Female	6 (60)
Age	40–49 yrs	3 (30)
	50–59 yrs	3 (30)
	60–69 yrs	4 (40)
Education	High school	2 (20)
	College/university	8 (80)
Profession	Employed	5 (50)
	Retired	5 (50)
Diabetes	0–4 yrs	4 (40)
	≥ 5 yrs	6 (60)
Computer/IT-knowledge	High	3 (30)
	Medium	5 (50)
	Small	2 (20)
Frequency of computer use	Every day	8 (80)
	Several times a week	1 (10)
	Once in a while	1 (10)
Frequency of internet use	Every day	7 (70)
	Several times a week	1 (10)
	Once in a while	1 (10)
Mobile phone knowledge	High	5 (50)
	Medium	4 (40)
	Small	1 (10)
Mobile phone use for phone calls	Every day	9 (90)
	Several times a week	1 (10)
Mobile phone use for text messaging	Every day	8 (80)
	Several times a week	2 (20)
Mobile phone use for e-mails, surfing	Every day	5 (50)
	Several times a week	1 (10)
	Never	4 (40)
Enjoy using computer during work/leisure time	Strongly agree	7 (70)
	Agree	3 (30)
Enjoy using mobile phone during work/leisure time.	Strongly agree	6 (60)
	Agree	4 (40)
Web services, mobile services for support, information within health care is a positive development	Strongly agree	8 (80)
	Agree	2 (20)
I can see advantages personally using web services, mobile services within health care	Strongly agree	8 (80)
	Agree	2 (20)

important diabetes management functions; they received high severity ratings (Fig. 4).

3.1.1. Critical and severe usability problems

The two most critical usability problems (severity level 4) were located in the *Glucose Readings View* and the *Glucose Diary View*. These required correcting a glucose value by removing the erroneous value and adding a new one. To complete the task, patients had to exit the *Glucose Readings View*, navigate to the *Glucose Diary View* and add the new value by clicking the *Add data* button. This sequence was confusing. Patients were also confused about the

Table 4
Example of usability problems and classifications for the most severe issues.

Usability problem description	Place of occurrence	% of pat. detecting per method			UPT classification		Severity
		TA ^a	I ^a	Q ^a	Artifact	Task	
Difficult knowing to exit the table view and click the “Add data” button to be able to add in a new value	Glucose readings view	70	100	50	Manipulation-Cognitive aspects-Visual cues (FC)	Task mapping-Navigation (FC)	4
Difficult to know to select “delete data” for the table view to change, delete, export, or print the value list	Glucose Diary View	70	100	100	Language-Naming/labeling (FC)	(NC)	4
Difficult to understand and perform the exporting action	Glucose Readings View	30	40	40	Manipulation-Cognitive aspects-Visual cues (FC)	Task mapping-Interaction (FC)	3
The blood pressure graph shows only the systolic blood pressure value, and the diastolic blood pressure is missing	Blood Pressure View	10	20	–	Visualness-Presentation of Information/results (FC)	(NC)	3
Difficult to know how to adjust the range of values that are to be shown, retrieved	Glucose Diary View	90	–	–	Manipulation-Cognitive aspects-Visual cues (FC)	Task mapping-Functionality (FC)	3
No action support to specify where the exported file is to be saved	Glucose Readings View	20	–	–	Visualness-Non message feedback (FC)	Task mapping-Functionality (FC)	3
Difficult to know how to navigate within and adjust the table view to get to or show the right value range, value	Glucose Readings View	10	–	–	Visualness-Presentation of Information/results (FC)	Task mapping-Navigation (FC)	3
A confirmation is missing to indicate that the exported file was saved and its directory	Glucose Readings View	10	–	–	Visualness-Non-message feedback (FC)	Task facilitation-Keeping the user on track (FC)	3

^a TA = Think Aloud usability test, I = Interview, Q = Questionnaire.

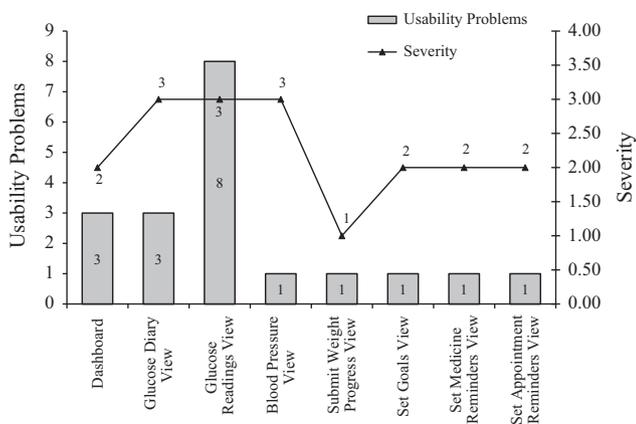


Fig. 4. Distribution of usability problems and assigned severity ratings across places of occurrence.

Delete data button as it had a dual function of allowing data to be exported and printed in the *Glucose Diary View*. The latter was an example of a label that did not make sense to them. These problems were classified into both (artifact) *Manipulation, Cognitive aspects/navigation*, and *Naming/labeling* deficiencies. These were issues brought up most frequently by patients across the different data collection methods.

Other serious issues (level 3 severity ratings) were in the *Glucose Readings View*, the *Glucose Diary View* and the *Blood Pressure View*. Several examples illustrate these issues. Problems occurred in the *Glucose Readings View* for the exporting task which included many difficult, hard-to-understand steps. The logic of exporting was not apparent, and system support was not available for this process. Then, if patients discovered how to export a file, they were unable to specify where the file would be saved. Patients also had difficulty adjusting the range of values in the *Glucose Diary View* and its graph. In another example, patients experienced difficulty navigating within the table in the *Glucose Readings View* (which allowed the display of correct value ranges or values). No

navigation support was available. The *Blood Pressure View* was notable because only the systolic blood pressure was visible in the graph although patients also entered their diastolic blood pressure. This was an incomplete and potentially harmful display.

Other difficulties included the various ways of accessing reminders in the *Dashboard View* due to labels and multiple locations. The *Medication Reminder* could be accessed through the *Medication Adherence* tab, the *Exercise, Weight and Medicine* pane, the *Message Settings* and *Medication Reminders* pane. The *Appointment Reminder* had no separate tab but was listed instead under the menu item titled *Message Settings and Reminders*. The “Tracking Goal” (exercise and weight) item had three different paths: the *Exercise and Weight Progress* tab, the *Exercise, Weight and Medicine* pane and the *Message Settings and Reminders* pane.

3.2. Results for methodological approaches

The different methods contributed to the 117 different initial usability problems. The usability test detected the majority (59) or 50% of the problems followed by the in-depth interview which identified 34 or 29% of the problems. The questionnaire identified substantially fewer usability problems at 24. Further, when looking at how the methods performed in combination with one another, the usability test plus the in-depth interview detected the highest number of problems (93) at 80%, while the usability test and post-questionnaire uncovered 83 or 71% of the usability problems. The combined questionnaire and the in-depth interview identified only 58 or 50% of the problems. These data indicate that the usability test detected most of the usability problems both on its own and in combination with the other methods while the post-test questionnaire detected the least volume of problems on its own and in combination.

Fig. 5 shows the unique and shared consolidated 19 usability problems between the different methods as depicted in the Venn diagram. As can be seen, the usability test and in-depth interview allowed us to identify 18 and 9 of the resulting usability problems, respectively. The usability test alone detected eight consolidated usability problems. This was more than the others at zero or one. All three methods converged on five usability problems.

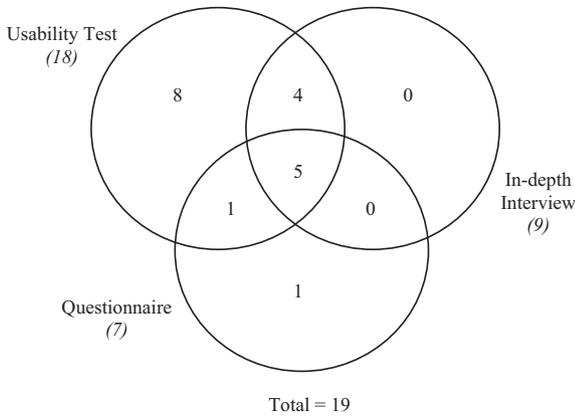


Fig. 5. Venn diagram showing the usability problem partitions for each data collection method.

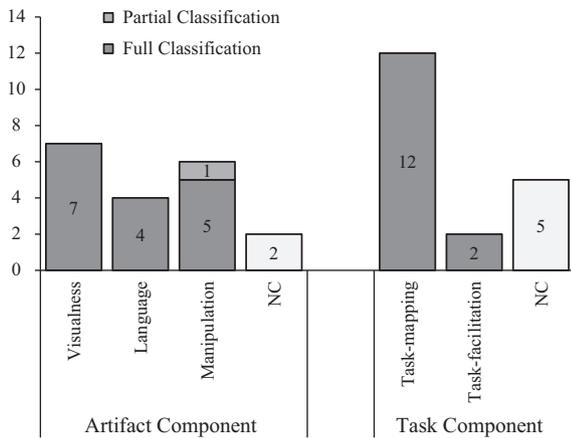


Fig. 6. Distribution of Artifact and Task Component classifications on primary level.

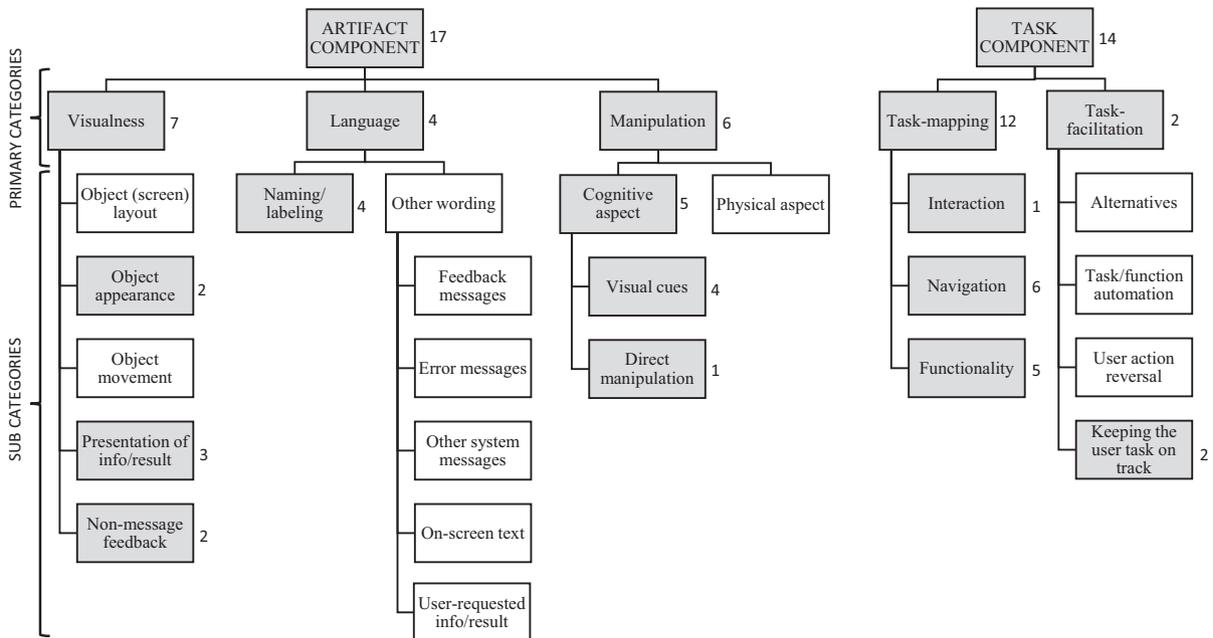


Fig. 7. Classification of identified usability problems according to the Usability Problem Taxonomy.

3.3. Problem classifications

Fig. 6 indicates the distribution of problems for each UPT classification level and the depth of classification. The most critical usability problems were classified into the artifact components *Manipulation*, *Cognitive aspects/Visual cues*, and *Visualness*, *Non-message feedback/Presentation of information/results*. Task classifications involved a variety of functionality/navigational concerns as well as one non-classification. The additional problems were classified under the artifact components *Language and Naming/labeling* and task components *Task-Mapping* and *Navigation*. In terms of the distribution of problems, we were able to classify all problems into the artifact or task component or both. However, two problems in the artifact component and five problems in the task component could not be classified. *Visualness* received the highest volume of full classifications (7) in the artifact component while in the *task component*, 12 problems received a full classification in *Task-mapping*.

As can be seen in Fig. 7, most subcategories for the whole system related to the presentation and visualization of information. For the six total *Manipulation* problems, five were categorized in the subcategory of *Cognitive aspect* and further broken down into *Visual cues* and *Direct manipulation*. The *Language* problems all related to the *Naming/labeling* subcategory. Most of the *Task-mapping* problems were associated with *Navigation* and *Functionality*. This indicates that most usability problems were related to system design and structure as a whole as it was not logically organized from a user perspective.

4. Discussion

In this study we conducted an in-depth usability assessment of a mHealth system using a combination of data collection methods and by employing two new frameworks for data analyses to structure and classify data. The multi-method approach in data collection resulted in a more comprehensive understanding of patients' interactions with the system and provided triangulation on severe usability issues. The structured data analyses provided an audit

trail for analytic decisions and resulted in clear descriptions of the usability problems.

4.1. Usability problem types and classifications

A total of 117 usability problems were identified across the different methods and were consolidated into 19 final classified usability issues. The majority of problems were located in the *Glucose Readings View* and had the highest severity ratings (serious or 3). Two other views, the *Glucose Diary View* and *Blood Pressure View*, had fewer usability problems but also high severity ratings. Usability problems included issues with deleting and entering glucose values, exporting and printing a glucose value (including salient blood pressure values) and information about how to save a file.

Based on the UPT classifications, the majority of problems were deficient visualization of *Information/results*. *Manipulation*, *Cognitive aspects* and *Visual cues* also dominated the usability issue classifications, reinforcing the visualization issues and cognitive aspects for system content. The results pointed to difficulties patients had completing common tasks. These were actions patients need to perform on a frequent basis for diabetes self-management, so these concerns need to be resolved quickly.

4.2. Method contributions

4.2.1. Data collection methods

Using a multi-method approach is beneficial for identifying usability problems, especially initial usability problems. Results from one method can be verified by another [63,64], contributing a form of problem validity through triangulation [20,65]. In our methods feasibility study, user testing identified the vast majority of problems followed by the interview and then the questionnaire. User testing plus the interview and questionnaire triangulated on five and four shared (verified) problems, respectively. The interview did not contribute unique consolidated usability problems in this study while the questionnaire produced one unique consolidated finding. However, each method provided a lens to view different usability issues, especially initial usability problems where usability testing contributed 59, interviews 34 and questionnaires 24 problems and thus contributed to the overall results [66,67].

Think Aloud was essential in usability testing in our study as it detected 18 of 19 consolidated usability issues. This is consistent with past literature indicating that user testing identifies more problems, identifies more recurring usability problems, and defines the underlying causes for usability problems [68]. The in-depth interview contributed the next highest volume of issues and nine of the 19 consolidated problems. Its contribution in particular was for triangulation on identified usability problems although the method did not contribute any unique consolidated issues. The reason for the latter is not completely clear but it may be that patients echoed only the issues they voiced during testing. Perhaps these particular issues resonated with them, and they did not have insights into other, new problems during the interview. Open-ended interview questions are beneficial in that they provide an opportunity for the user to express thoughts freely and provide perspectives about problems that may not surface during user testing. Of course, open-ended questions can vary from study to study but the primary value of interviewing in the evaluation process lies in users expressing perspectives in their own words. Also, multiple lenses might generate more unique issues in a more complex application than the one studied here. In the future, researchers may want to follow-up the high-level questions we used with more specific probes to uncover unique issues during interviewing. Future researchers may also categorize issues in a more granular fashion than in this study.

The post-questionnaire generated fewer issues but verified significant usability problems. A reason for the lower volume of issues with the questionnaire might be due to study sequencing. We began with the usability test and ended with the questionnaire which meant that some problems reverberated during each step of the process. Also, questions could have seemed similar to users or perhaps they became fatigued at the end of the two-hour testing session. Perhaps users mentioned the same issues each method because they were seen as especially severe to them. This may indicate the value of using all three methods. The significance of the post-questionnaire may lie specifically in eliciting users' views in writing as they had time to ponder and write about issues they thought were difficult.

Five shared usability problems were detected by all three methods. Four were critical issues while one was less so. Thus, different methods may surface different issues having varying levels of severity. The multiple methods produced a more comprehensive set of usability problems than did one individual method, although more so at the initial problem identification step. This notion is also highlighted by other authors in the health informatics technology literature. In particular, authors indicated that multiple evaluation methods are important because systems are becoming more complex [49]. Based on our results, we also recommend the use of multiple methods but advocate a minimum of three methods to capture a variety of issues of varying kinds and at differing severity levels. However, the levels of questions in the interview and questionnaire should be different. This recommendation will need to be tempered by the available time participants have for testing, e.g., few physicians will be available for two-hour usability evaluation sessions.

4.2.2. Structured data analyses

The FA method provided a beneficial way of structuring data from our multi-method approach and was particularly useful for coding large amounts of data as noted by previous authors [26]. Despite our need to modify FA to accommodate the UPT for usability problem classification, we found the FA very worthwhile overall. Researchers may find a similar modification for the FA suitable during future usability evaluations.

Sometimes vague problem descriptions or the problem nature meant a problem could not be fully classified in the UPT. As Keenan et al. noted, this can result in a null or partial (versus full) classification [30]. In our study several usability issues could only be classified in the artifact component level in the *Naming/labeling* category as they did not have sufficient descriptions that might lead to a deeper task level classification. Whether issues are fully or partially classified does not mean that issues are categorized as “better” or worse but merely that the description is more or less granular.

We found the UPT of significant value. Determining the UPT classification for each problem assisted in defining the usability problems on a more detailed as well as a hierarchical level. Similar to Keenan et al. [30], we found the resulting issues were well described overall, and we could group problems of a similar nature easily. We also found that adding a severity rating to the problems and aggregating the frequency of the problem would likely aid in problem analysis and future approaches for solutions.

4.3. Contributions to the literature

Our findings verify the usefulness of multiple data collection methods in generating more comprehensive lists of issues and serving as a means of triangulation for severe usability issues. Each method generated unique initial usability problems. For consolidated usability problems we were able to calculate the percentage of patients who raised a specific issue across the different methods,

providing an indication of its importance to users. This is a clear benefit as triangulation can assist developers in locating priority user problems. The multiple methods approach was somewhat less beneficial in identifying unique consolidated usability problems. Thus, the most significant contributions of our study lie in the data analysis methods. The FA method was useful for structuring qualitative data as it assisted in standardizing problem descriptions and promoted consistent analyses across researchers. Likewise, the FA was valuable because it (1) provided a systematic way of managing and mapping the data [26], (2) created reproducible steps [29], and (3) supported both our inductive (usability issue coding) and deductive (UPT) approaches [23]. By combining the FA method with the UPT framework we were also able to create in-depth, easy-to-follow classifications for each identified usability problem. We think this will aid in solution generation and could contribute to problem-solving for designers [30].

Overall, this study adds to the body of literature because it provides an example of a multi-method approach for collecting data on usability problems [18], for standardizing data analyses, and contributing to the reproducibility of results in qualitative usability studies. This more standardized and structured approach could assist in building a science around mHealth usability.

4.4. Lessons learned and recommendations

We found the FA analysis method valuable because the multi-method design generated a large amount of data even with a sample of only 10 users. However, coding using both FA and UPT was still a laborious process. Its utility is in generating a clear audit trail and decision path for data categories (usability problem classifications in our case).

The FA method is recommended for studies employing several researchers because of its more structured process [26]. We found it effective with just two researchers because of the linear steps in the coding process and the resulting descriptions, allowing researchers to work together in data analysis. We also found it advantageous to have detailed usability problems that we could then classify using UPT.

The UPT was also beneficial for classifying usability by the artifact and task components. However, it was helpful to have two researchers to discuss classifications before reaching consensus. Keenan et al. [30] indicate the UPT is easy to use for individuals familiar with the terminology in HCI and those with some background in usability. Classification is also easier when the researcher is familiar with the particular context of use of the system [30]. The two researchers in this study fit those criteria. Future researchers will want to take these criteria into consideration before analyses.

Once the analysis process was completed, we had a list of defined and classified usability problems with severity ratings that could be aggregated into similar UPT categories. We recommend using both frameworks for data analyses in future usability studies, especially those with multi-method approaches.

4.5. Limitations and future research

The sample size of 10 patients was small but fitting for usability and methods feasibility testing. The smaller sample is consistent with sample sizes recommended by Nielsen and Landauer [50], Virzi [51], Monk et al. [52]. For formative usability tests, such as this one, five to eight users are able to detect 80–85% of usability problems [50–52]. Moreover, an extensive amount of data were generated from the multi-method approach. Thus, 10 users was sufficient for methods feasibility testing in this study.

Our study involved a randomized sample from a larger, convenience sample of patients with diabetes. The convenience sampling

frame for the larger study may make it difficult to generalize findings to the diabetes population as a whole. For instance, the study included patients who are more highly educated and who use technology more than an urban, lower income or rural sample might. For future studies, researchers may want to use purposeful sampling to select users with different characteristics to maximize variability.

Patients were new to using the system. It is possible that more practiced users may adapt to the observed issues after they interact more with the system. However, the user interactions did surface critical usability issues and pointed to particular issues that made the current version non-intuitive. In future mHealth research, it is imperative that users be involved early and often in the development process of mHealth applications. Future research might include testing with larger samples, testing earlier in the development life-cycle, testing of other available classification schemes and frameworks and/or repeating this study to investigate how reproducible results are with the tools employed here. Future researchers may also want to test data collection methods using a larger sample and more methods to determine whether they yield more unique usability issues.

5. Conclusions

Recent systematic reviews of mHealth self-management tools in general and for diabetes in particular speak to the need for more studies on patient interaction and system usability. Finding standardized, structured and reproducible ways to work in usability evaluation is important for providing evidence. This study provides an example of a multi-method design for both data collection and data analyses. Multiple data collection methods resulted in a more comprehensive set of usability problems and helped triangulate data. The structured data analyses allowed reproducible steps and data validation (triangulation), a method of determining the most severe problems for users.

Usability testing with Think Aloud was essential for surfacing usability issues. The in-depth interview and questionnaire allowed data triangulation for severe usability issues, but both uncovered a smaller volume of consolidated usability issues. For data analyses, the more structured method, using the FA, provided a more standardized and feasible way to derive usability problems from a large volume of qualitative data. The FA method can guide analyses across multiple researchers. The UPT was advantageous as an in-depth classification scheme and for determining severity ratings for usability problems. It also assisted in categorizing specific types of problems which could be useful for designers. We recommend the use of multiple data collection methods to uncover a variety of problem types and severity levels. We also recommend the use of the FA and UPT methods during data analyses.

Conflict of interest

The authors declare that there are no conflicts of interest.

Appendix A

Start-up topics/questions for in-depths interviews

- What parts of the system did you think were well designed (?)
- Which parts of the system did you think were inadequately designed (?)
- Do you have any other comments about the system functions and regarding its usability (?)

Appendix B

1. How do you consider your level of computer-/IT-knowledge?
 High Medium Small None
2. How often do you use a computer (approximately)?
 Every day Several times a week Once in a while Never
3. How often do you use the Internet (approximately)?
 Every day Several times a week Once in a while Never
4. How do you consider your mobile phone knowledge level?
 High Medium Small None
5. How often do you use your mobile phone to make, receive phone calls?
 Every day Several times a week Once in a while Never
6. How often do you use your mobile phone for text messaging?
 Every day Several times a week Once in a while Never
7. How often do you use your mobile phone for e-mails, surfing?
 Every day Several times a week Once in a while Never

To what level do you agree or disagree with the following statements?

8. I like to use the computer in my work/leisure time.
 Strongly agree Agree Disagree Strongly disagree
9. I like to use my mobile phone in my work/leisure time.
 Strongly agree Agree Disagree Strongly disagree
10. Web services, mobile services such as different patient support and information services within health care are getting more common. I believe that this is a positive development.
 Strongly agree Agree Disagree Strongly disagree
11. I can see advantages for me personally in using web services, mobile services within health care.
 Strongly agree Agree Disagree Strongly disagree

12. What usability/user experiences were you satisfied with regarding the mHealth system?

.....

13. What usability/user experiences were you dissatisfied with regarding the mHealth system

.....

14. Any other comments? Please specify.

.....

Appendix C

Usability problem descriptions and classifications (complete list)

Usability problem description	Place of occurrence	% of pat. detecting per method			UPT Classification		Severity
		TA ^a	I ^a	Q ^a	Artifact	Task	
(1) Difficulty knowing to exit the table view and click the “Add data” button to be able to add in a new value	Glucose Readings View	70	100	50	Manipulation-Cognitive aspects-Visual cues (FC)	Task mapping-Navigation (FC)	4
(2) Difficult to understand and perform the exporting action	Glucose Readings View	30	40	40	Manipulation-Cognitive aspects-Visual cues (FC)	Task mapping-Interaction (FC)	3
(3) No support to specify where the exported file is to be saved	Glucose Readings View	20	–	–	Visualness-Non-message feedback (FC)	Task mapping-Functionality (FC)	3
(4) Difficult to know how to navigate within and adjust the table view to get to or show the right value or value range	Glucose Readings View	10	–	–	Visualness-Presentation of Information/results (FC)	Task mapping-Navigation (FC)	3
(5) No message about the lack of capability to save an exported file	Glucose Readings View	10	–	–	Visualness-Non-message feedback (FC)	Task facilitation-Keeping the user on track (FC)	3
(6) Difficult to know to go to the “Export” tab to view, export, and/or print the data table or its values	Glucose Readings View	50	30	30	Manipulation-Cognitive aspects-Direct manipulation (FC)	Task facilitation-Keeping the user task on track (FC)	2
(7) Difficult to know to choose the Action-tab for deleting a value	Glucose Readings View	50	–	–	Manipulation-Cognitive aspects-Visual cues (FC)	Task mapping-Navigation (FC)	2
(8) Difficult to find the Export button for the specific export command	Glucose Readings View	20	–	–	Visualness-Object appearance (FC)	(NC)	2
(9) Difficult to know to select “Delete data” for the table view to change, delete, export, or print the value list	Glucose Diary View	70	100	100	Language-Naming/labeling (FC)	(NC)	4
(10) Difficult to know how to adjust the range of values that are to be shown, retrieved	Glucose Diary View	90	–	–	Manipulation-Cognitive aspects-Visual cues (FC)	Task mapping-Functionality (FC)	3
(11) Incorrect rendering of graph values for the last 30 or 90 days as it lacks the correct delimiters	Glucose Diary View	20	–	–	Visualness-Presentation of Information/results (FC)	(NC)	2
(12) Difficult to find, access “Medication Reminder” as four different paths existed with similar but different names	Dashboard	40	20	–	Language – Naming/labeling (FC)	Task-mapping-Navigation (FC)	2
(13) Difficult to find, access “Appointment reminder” as it was listed under a pane with a different name	Dashboard	30	–	10	Language-Naming/labeling (FC)	Task-mapping-Navigation (FC)	2
(14) Difficult to find the “Tracking goal” (exercise, weight) item as three different paths existed with similar but different names	Dashboard	40	–	–	Language-Naming/labeling (FC)	Task-mapping-Navigation (FC)	2
(15) The blood pressure graph shows only the systolic blood pressure value; the diastolic blood pressure is missing	Blood Pressure View	10	20	–	Visualness-Presentation of Information/results (FC)	(NC)	3
(16) Weights lack conversion to the metric system	Submit Weight Progress View	10	10	–	(NC)	Task-mapping-Functionality (FC)	1
(17) Difficult to detect and distinguish the Update and Submit buttons	Set Goals View	–	–	10	Visualness – Object appearance (FC)	(NC)	2

(continued on next page)

Appendix C (continued)

Usability problem description	Place of occurrence	% of pat. detecting per method			UPT Classification		Severity
		TA ^a	I ^a	Q ^a	Artifact	Task	
(18) Times are only available for whole hours, not minutes or half hours	Set Medication Reminders View	10	10	–	Manipulation (PC)	Task-mapping-Functionality (FC)	2
(19) Only dates and not times are available for appointment reminders	Set Appointment Reminders View	20	10	10	(NC)	Task-mapping-Functionality (FC)	2

^a TA = Think Aloud usability test, I = Interview, Q = Questionnaire.

References

- Centers for Disease Control and Prevention, National Diabetes Statistics Report: Estimates of Diabetes and Its Burden in the United States, Department of Health and Human Services, Atlanta, GA, U.S., 2014
- P. Zhang et al., Global healthcare expenditure on diabetes for 2010 and 2030, *Diab. Res. Clin. Pract.* 87 (3) (2010) 293–301.
- World Health Organization (WHO), Diabetes Fact Sheet N°312. 2015 January 2015; Available from: <<http://www.who.int/mediacentre/factsheets/fs312/en/>> [cited 06 Sept 2015].
- The Diabetes Control and Complications Trial Research Group, The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. The diabetes control and complications trial research group, *N. Engl. J. Med.* 329 (14) (1993) 977–986.
- M.M. Funnell et al., National standards for diabetes self-management education, *Diab. Care* 34 (Suppl. 1) (2011) S89–S96.
- M.M. Funnell, R.M. Anderson, The problem with compliance in diabetes, *JAMA* 284 (13) (2000) 1709.
- P. Sutcliffe et al., Systematic review of communication technologies to promote access and engagement of young people with diabetes into healthcare, *BMC Endocr. Disord.* 11 (2011) 1.
- Y.S. Chung, Y. Kim, C.H. Lee, Effectiveness of the smart care service for diabetes management, *Healthc. Inform. Res.* 20 (4) (2014) 288–294.
- X. Liang et al., Effect of mobile phone intervention for diabetes on glycaemic control: a meta-analysis, *Diab. Med.* 28 (4) (2011) 455–463.
- B. Holtz, C. Lauckner, Diabetes management via mobile phones: a systematic review, *Telemed. J. E Health* 18 (3) (2012) 175–184.
- K. Waki et al., DialBetics: a novel smartphone-based self-management support system for type 2 diabetes patients, *J. Diab. Sci. Technol.* 8 (2) (2014) 209–215.
- S.A. Mulvaney et al., A pilot test of a tailored mobile and web-based diabetes messaging system for adolescents, *J. Telemed. Telecare* 18 (2) (2012) 115–118.
- A. Fioravanti et al., Automatic messaging for improving patients engagement in diabetes management: an exploratory study, *Med. Biol. Eng. Comput.* 53 (12) (2015) 1285–1294.
- E. Arsand et al., Mobile phone-based self-management tools for type 2 diabetes: the few touch application, *J. Diab. Sci. Technol.* 4 (2) (2010) 328–336.
- O. El-Gayar et al., Mobile applications for diabetes self-management: status and potential, *J. Diab. Sci. Technol.* 7 (1) (2013) 247–262.
- S. Wolpin, M. Stewart, A deliberate and rigorous approach to development of patient-centered technologies, *Semin. Oncol. Nurs.* 27 (3) (2011) 183–191.
- C.R. Lyles, U. Sarkar, C.Y. Osborn, Getting a technology-based diabetes intervention ready for prime time: a review of usability testing studies, *Curr. Diab. Rep.* 14 (10) (2014) 534.
- B.C. Zapata et al., Empirical studies on usability of mHealth apps: a systematic literature review, *J. Med. Syst.* 39 (2) (2015) 182.
- C.A. Green et al., Approaches to mixed methods dissemination and implementation research: methods, strengths, caveats, and opportunities, *Adm. Policy Ment. Health* (2014).
- M.Q. Patton, Enhancing the quality and credibility of qualitative analysis, *Health Serv. Res.* 34 (5 Pt 2) (1999) 1189–1208.
- D.F. Polit, *Nursing Research: Generating and Assessing Evidence for Nursing Practice*, 9th ed., Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, 2012.
- M.Q. Patton, *Qualitative Research & Evaluation Methods: Integrating Theory and Practice: the Definitive Text of Qualitative Inquiry Frameworks and Options*, SAGE Publications, Inc., Thousand Oaks, California, 2015.
- S. Elo, H. Kyngas, The qualitative content analysis process, *J. Adv. Nurs.* 62 (1) (2008) 107–115.
- A.W. Kushniruk, V.L. Patel, J.J. Cimino, Usability testing in medical informatics: cognitive approaches to evaluation of information systems and user interfaces, in: *Proc AMIA Annu Fall Symp*, 1997, pp. 218–222.
- D.J. Cohen, B.F. Crabtree, Evaluative criteria for qualitative research in health care: controversies and recommendations, *Ann. Fam. Med.* 6 (4) (2008) 331–339.
- J. Ritchie, L. Spencer, Qualitative data analysis for applied policy research, in: A. Bryman, R.G. Burgess (Eds.), *Analyzing Qualitative Data*, 1994, pp. 173–194.
- J. Ritchie, J.R. Lewis, *Qualitative Research Practice: A Guide for Social Science Students and Researchers*, SAGE, London, 2003.
- D.J. Ward et al., Using framework analysis in nursing research: a worked example, *J. Adv. Nurs.* 69 (11) (2013) 2423–2431.
- N.K. Gale et al., Using the framework method for the analysis of qualitative data in multi-disciplinary health research, *BMC Med. Res. Methodol.* 13 (2013) 117.
- S. Keenan et al., The usability problem taxonomy: a framework for classification and analysis, *Empirical Softw. Eng.* 4 (1) (1999) 71–104.
- R. Jeffries, Usability problem reports: helping evaluators communicate effectively with developers, in: R.L. Mack, J. Nielsen (Eds.), *Usability Inspection Methods*, John Wiley & Sons Inc., New York, 1994, pp. 273–294.
- J.S. Dumas, J. Redish, Rev. ed., *A Practical Guide to Usability Testing*, vol. 2, Intellect Books, Exeter, England ; Portland, Or., 1999. 404p.
- J. Nielsen, R. Molich, Heuristic evaluation of user interfaces, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People*, ACM, Seattle, Washington, United States, 1990, pp. 249–256.
- J. Nielsen, Finding usability problems through heuristic evaluation, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Monterey, California, United States, 1992, pp. 373–380.
- S.L. Keenan, *Product Usability and Process Improvement Based on Usability Problem Classification*, Virginia Polytechnic Institute and State University, 1996.
- J.M. Carroll, W.A. Kellogg, M.B. Rosson, The Task-artifact Cycle, in: M. Carrol (Ed.), *Designing Interaction: Psychology at the Human-Computer Interface*, Cambridge University Press, New York, 1991, pp. 74–102.
- B. Shneiderman, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, Addison Wesley Longman, Reading, Mass., 1998.
- K. Capozza et al., Going mobile with diabetes support: a randomized study of a text message-based personalized behavioral intervention for type 2 diabetes self-care, *Diab. Spectr.* 28 (2) (2015) 83–91.
- J. Nielsen, *Usability Engineering*, xiv, Academic Press, Boston, 1993. 358p.
- J. Nielsen, Heuristic evaluation, in: R.L. Mack, J. Nielsen (Eds.), *Usability Inspection Methods*, John Wiley & Sons Inc., New York, 1994, pp. 25–62.
- P.G. Polson et al., Cognitive walkthroughs – a method for theory-based evaluation of user interfaces, *Int. J. Man Mach. Stud.* 36 (5) (1992) 741–773.
- C. Wharton et al., The cognitive walkthrough method: a practitioner's guide, in: J. Nielsen, R.L. Mack (Eds.), *Usability Inspection Methods*, John Wiley & Sons, Inc., New York, 1994, pp. 105–140.
- K.A. Ericsson, H.A. Simon, *Protocol Analysis: Verbal Reports as Data*, MIT Press, Cambridge, Mass., 1984, p. 426.
- C. Lewis, J. Rieman, *Task-centered User Interface Design: A Practical Introduction*, University of Colorado, Boulder, Department of Computer Science, 1993.
- S.G. Shah, I. Robinson, Benefits of and barriers to involving users in medical device technology development and evaluation, *Int. J. Technol. Assess Health Care* 23 (1) (2007) 131–137.
- L. Yardley et al., Understanding reactions to an internet-delivered health-care intervention: accommodating user preferences for information provision, *BMC Med. Inform. Decis. Mak.* 10 (2010) 52.
- K.G. Dobson, P. Hall, A pilot study examining patient attitudes and intentions to adopt assistive technologies into type 2 diabetes self-management, *J. Diab. Sci. Technol.* 9 (2) (2014) 309–315.
- K.A. Ericsson, H.A. Simon, *Protocol Analysis: Verbal Reports as Data Revised Edition*, The MIT Press, 1993.
- A.W. Kushniruk, V.L. Patel, Cognitive and usability engineering methods for the evaluation of clinical information systems, *J. Biomed. Inform.* 37 (1) (2004) 56–76.

- [50] J. Nielsen, T.K. Landauer, A mathematical model of the finding of usability problems, in: *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, ACM, Amsterdam, The Netherlands, 1993, pp. 206–213.
- [51] R.A. Virzi, Refining the test phase of usability evaluation: how many subjects is enough?, *Hum Factors* 34 (4) (1992) 457–468.
- [52] A. Monk et al., *Improving your human-computer interface: a practical technique*, BCS Practitioner Series, Prentice-Hall, London, 1993, p. 99.
- [53] J. Nielsen, R.L. Mack, *Usability Inspection Methods*, xxiv, John Wiley & Sons, Inc., New York, 1994. 413p.
- [54] Nielsen Norman Group, *Turn User Goals into Task Scenarios for Usability Testing*, 2014 Jan 12; Available from: <<http://www.nngroup.com/articles/task-scenarios-usability-testing/>> [cited 25 May 2015].
- [55] A.W. Kushniruk et al., Emerging approaches to usability evaluation of health information systems: towards in-situ analysis of complex healthcare systems and environments, *Stud. Health Technol. Inform.* 169 (2011) 915–919.
- [56] M.J.V.D. Haak, M.D.T.D. Jong, P.J. Schellens, Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue, *Behav. Inform. Technol.* 22 (5) (2003) 339–351.
- [57] M. Maguire, Methods to support human-centred design, *Int. J. Hum Comput Stud.* 55 (4) (2001) 587–634.
- [58] S. Kvale, *InterViews : Learning the Craft of Qualitative Research Interviewing*, in: S. Brinkmann (Ed.), second ed., Sage Publications, Los Angeles, 2009.
- [59] B.T. Karsh et al., Health information technology: fallacies and sober realities, *J. Am. Med. Inform. Assoc.* 17 (6) (2010) 617–623.
- [60] TechSmith Corporation, *Morae Usability Testing and Market Research Software*, TechSmith Corporation, Okemos, MI, USA, 2015.
- [61] QSR International Pty Ltd, *NVivo 10 Research Software for Analysis and Insight*, QSR International Pty Ltd., Doncaster, Victoria, Australia, 2015.
- [62] D. Travis, *How to Prioritise Usability Problems*, 2009 October 15; Available from: <<http://www.userfocus.co.uk/articles/prioritise.html>> [cited 2015 August 26].
- [63] J. Protheroe et al., An intervention to promote patient participation and self-management in long term conditions: development and feasibility testing, *BMC Health Serv. Res.* 10 (2010) 206.
- [64] M. Kirwan et al., Design, development, and formative evaluation of a smartphone application for recording and monitoring physical activity levels: the 10,000 Steps “iStepLog”, *Health Educ. Behav.* 40 (2) (2013) 140–151.
- [65] L. Atack, R. Luke, E. Chien, Evaluation of patient satisfaction with tailored online patient education information, *Comput. Inform. Nurs.* 26 (5) (2008) 258–264.
- [66] J. Horsky et al., Complementary methods of system usability evaluation: surveys and observations during software design and development cycles, *J. Biomed. Inform.* 43 (5) (2010) 782–790.
- [67] 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, in: *AMIA Annu Symp Proc*, 2009, pp. 714–718.
- [68] R. Jeffries et al., User interface evaluation in the real world: a comparison of four techniques, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, New Orleans, Louisiana, USA, 1991, pp. 119–124.