2016 49th Hawaii International Conference on System Sciences

Design Implications of User Experience Studies: The Case of a Diabetes Wellness App

Bengisu Tulu
Worcester Polytechnic Institute
bengisu@wpi.edu

Diane Strong
Worcester Polytechnic Institute
dstrong@wpi.edu

Lei Wang
Worcester Polytechnic Institute
lwang1@wpi.edu

Qian He
Worcester Polytechnic Institute
qhe@wpi.edu

Emmanuel Agu
Worcester Polytechnic Institute
emmanuel@cs.wpi.edu

Peder Pedersen
Worcester Polytechnic Institute
pedersen@wpi.edu

Soussan Djamasbi
Worcester Polytechnic Institute
djamasbi@wpi.edu

Abstract

We developed a health and wellness smartphone app to support diabetes self-management. The target users are older, less healthy adults with advanced type 2 diabetes. We illustrate the design implications of a health and wellness smartphone app for such users by presenting the process and results of our usability and user experience lab test with our app.

1. Introduction

Health and wellness apps for smartphones – in fact, thousands of them – are now available in app stores. Yet, many were developed as personal initiatives of a few developers and implement only a minority of the behavioral strategies that are commonly used in clinical health and wellness interventions [1].

We believe that health and wellness apps should follow evidence based clinical practices and user interface design and development best practices because health and wellness app usage is voluntary, and thus, an excellent user experience provided by a clinically safe app is critical. With that in mind, our team includes technical experts in user experience, user behavior, app development, and medical experts in diabetes and behavioral medicine.

Our app is designed to help users to self-manage their type 2 diabetes. While health and wellness apps for the young and healthy are common, the design challenge we face is designing an engaging smartphone app for older, less healthy users. We illustrate those design challenges by presenting findings of our usability study.

2. Background

To understand the design choices we made about our app and its user experience lab study, we provide some background about the ways in which aspects of our app and our user experience study are similar to, and different from, other information technology (IT) and user experience studies.

2.1. Our App and its Intended Users

Our app is designed to support self-management of type 2 diabetes by those with advanced type 2 diabetes. As such, our intended users are older, less mobile, with possibly poor eyesight.

To assure that our app is technically, medically, and behaviorally sound, our design and development team is interdisciplinary. On the technical side, our team includes specialists in image processing, app design and development, and user interface and experience design. On the medical side, our team includes diabetes, diabetic wound, and behavioral medicine specialists.

We followed Scrum software development process with short iterative cycles of a few weeks each of design, implementation, and testing [2]. Thus, we were continually designing, implementing, and testing that the app was technically, medically, and behaviorally sound.

The app runs on Android smartphones. It is a stand-alone app in the sense that it is not connected to a healthcare provider's computing infrastructure. While smartphones can be challenging for our target population to use, more of the elderly are adopting

1530-1605/16 $31.00 © 2016 IEEE
DOI 10.1109/HICSS.2016.434

smartphones [3]. Furthermore, our design is tailored to this audience, e.g., it includes large buttons, no scroll bars, and minimal typing.

In terms of functionality, the app allows users to track glucose levels, weight, and physical activity. To engage and motivate users, it facilitates goal setting, provides reminders, daily tips, and encouraging feedback messages based on the data users enter about their glucose, weight, and physical activity [4]. To support the needs of those with advanced diabetes, the app also captures images of diabetic foot ulcers using the phone camera, analyzes them, and reports feedback on their healing progress [5-6]. The functionality for motivating and engaging users and that for tracking foot ulcers are the contributions of our app beyond the many apps available for diabetes self-management.

2.2. Healthcare Context and Patients as Users

IT to support business organizations and their processes or electronic health record (EHR) systems used in healthcare organizations are designed to support tasks that users must perform in their jobs. Thus, their use is generally not voluntary. In such contexts, usability and a good user experience, while desirable, may not be critical for ensuring use.

In contrast, patient-facing healthcare IT, such as our app, is voluntary use. First, patients' performance of health maintenance tasks is voluntary, e.g., exercising and healthy diet. Second, even if patients are taking good care of themselves, they can choose to do so without the support of IT. Thus, apps such as ours must engage users through value added features and excellent user experience.

Health maintenance apps must also conform to good medical practices in the information they provide to, and ask of, patients and the health behaviors they encourage. Ideally, health maintenance apps motivate users to engage in healthy behaviors. The goals of being engaging and providing an excellent user experience may sometimes conflict with the goals of operating in a way that is medically and clinically sound, resulting in the need for design tradeoffs. Thus, health maintenance app developers must consider the typical requirements for an excellent user experience and balance those with the needs of users as patients with health problems and clinician recommendations.

In terms of Alter's HCI topics framework [7], a health maintenance app likely has design elements to handle all four types of user behaviors: engineered, guided, emergent, and undesirable. The app seeks to engineer behavior to the extent that the design is seeking conformance with health guidelines. Primarily, our app seeks to guide users' behavior in assisting them in health maintenance tasks. In the longer term, we expect our app to lead to emergent behavior as users determine how to integrate use of the app into their daily activities. Lastly, our app seeks to avoid undesirable behaviors in the long term as users understand how their behaviors affect health outcomes.

A true assessment of the user experience of a health maintenance app requires monitoring user behaviors over time. In the short term, however, it is important to ensure that the app is sufficiently usable and provides a good user experience. Otherwise, the app will not be used long enough to generate positive changes in health maintenance behaviors.

2.3. Smartphone App User Interfaces

As Galletta and Dunn [2] note, smartphone user interfaces differ substantially from those of PCs, largely due to the small screen size and the use of the touchscreen swipes, etc. for most input. This means that standard usability and user experience study design, or at least some aspects of them, may not be appropriate for smartphones. Norman and Nielson, acknowledged experts in usability and usability studies, go a step further and claim that smartphone user interfaces are a step backward [8-9]. They argue that smartphone developers have ignored the many findings of HCI researchers and thus have violated many good interface design principles and guidelines.

Whether smartphone user interfaces are a step forward or backward can be debated, but their user interfaces clearly do not rely on the knowledge users have built up about how to use PCs. Thus, smartphone apps and their user interfaces should be studied in controlled lab settings, especially in our case of older users less likely to be familiar with smartphones, to uncover any violations of basic usability principles.

2.4. Eye-tracking User Experience Studies

Using eye-tracking equipment to aid in usability lab studies is becoming more common [10]. Eye-tracking provides accurate data about what users are viewing – more accurate than the traditional speak aloud protocols – because it captures viewing that users may not even be aware of. The captured data are objective and can be quantified, e.g., exactly how long a user spent on a screen or looking at a particular button or part of a screen [11]. With eye-tracking equipment, detailed, accurate data can be collected about each user in the study.

One downside of eye-tracking is that eye-tracking studies handle one subject at a time. Although detailed data are collected for each subject, fewer subjects can be tested. Furthermore, eye-tracking equipment is expensive so although eye-tracking has many advantages, it is not yet standard procedure for usability and user experience studies.

Eye-tracking studies have focused on the design of web sites, especially those used to convey information to users [10]. The design goals for these sites are to keep users at the site so that they view and absorb as much of the information on the site as possible [12]. A key result from such studies is that users follow an “F-shaped” reading pattern for web sites, looking primarily at the left and the top of a screen and often missing information on the right or the bottom, especially if that requires scrolling [13]. Eye-tracking has been used with mobile devices, e.g., [14-15], but often in studies examining using mobile devices to access web sites.

One value of these previous eye-tracking studies is that they developed metrics for translating raw eye-tracking data into data useful for analyses of a user’s experience with an interface. For example, eye-tracking equipment with a 60Hz sampling rate is capturing a gaze point (where the user is looking) every 16.6 milliseconds [10], but this level of detail is not useful. The concept of a fixation is a steady gaze at the same point for at least 100 milliseconds [10]. These fixations and their sequence or pattern can be useful data for assessing a user interface.

A number of these metrics, however, were developed to support the goals of web sites, i.e., to keep the user on the web site looking at key information. Because we are testing for ease-of-use of our app user interface, we are not seeking long gazes. In fact, long gazes indicate problems, which Djamasbi [10] notes as an alternative explanation for a long gaze, i.e., a long gaze may indicate confusion about what to do next.

Thus, our use of eye-tracking is primarily to collect accurate time on task measures so that we can detect screens that are confusing to most users or screens that have high time variability among users.

3. Methodology

We conducted a formative usability study to evaluate the design of our app and to collect, analyze, and report the findings back to the designers and developers as quickly as possible to inform the next iteration of the project. We used a laboratory setting where we had the ability to control the environment and focus solely on user experience.

3.1. Subject Enrollment

We enrolled a total of five subjects through email advertisements using the employee mailing list of a university in the northeastern U.S. Given the richness of eye movement data, formative eye tracking studies with 5-6 participants are quite common in industry research [13]. Subjects were required to meet the following criteria: (1) 35 years of age or older and (2) have some first-hand knowledge of Type 2 diabetes management, either as a patient or as a caregiver. The target audience for our app is those with advanced Type 2 diabetes or their caregivers. Therefore, some knowledge of what it means to live with and manage Type 2 diabetes was important to collect meaningful user experience feedback.

3.2. Laboratory Environment

We used the user experience lab on campus for this study. This laboratory is equipped with commercially available eye tracking devices that are used to track gaze and to map this gaze information on the screens that are being viewed. Figure 1 shows the device we used to conduct our study, specifically a remote eye tracker, the Tobii X2 with sampling rate of 60HZ.

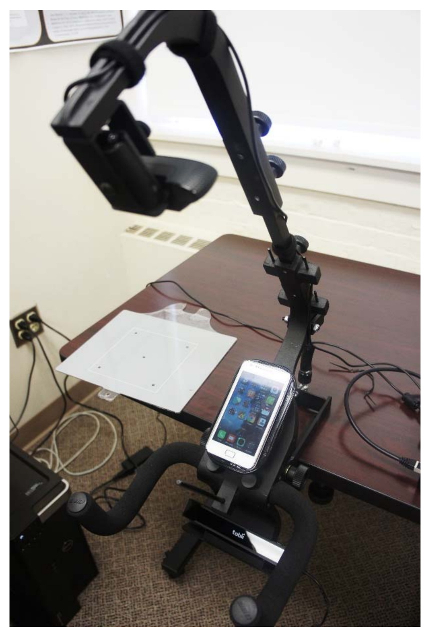

Figure 1 - Eye tracker used in the study

A remote eye tracker is unobtrusive and does not require subjects to be attached to devices such as helmets as in the case of old eye tracking technologies. The calibration process for these old devices is also typically long, an hour or more. In our study each participant completed a calibration process, which took less than 1 minute. Subjects were seated in front of the eye tracker to complete their tasks. We were able to view each subject's gaze over the mobile interface as we were conducting the experiments. Separate researchers were watching the gaze while another researcher worked with the subject.

3.3. Use Scenarios

The following use scenarios were presented to the subjects during the usability study.

3.3.1. Use Scenario 1. In this scenario, you are a recently diagnosed diabetic and your doctor asked you to measure your blood glucose level three times a day and track your weight once a week.

Today, you have measured your glucose, and your values were 125, 250 and 75.

You decided to record these values in the app so that you can show them to your doctor in your next visit.

After you record them, you decided to view the graph to make sure that all your values are entered. You also wanted to see how you did today compared to previous days.

Today you also weighed yourself, and your weight was 190.

You decided to record your weight in the app as well to keep all your information in the same place. You thought this would make your life easier when you see your doctor again.

After logging your weight, you were curious to see how your weight has changed over the past month. You thought the app could show you this change if you viewed the weight graph.

You finished the day in a good mood and wanted to record this also in the app.

3.3.2. Use Scenario 2. In this scenario, you are a patient who had been diagnosed with pre-diabetes. Your doctor told you that exercise is really important for you if you want to keep your blood sugar under control.

Today, you had a 45-minute brisk walk. You decided to add your activity in the app.

Since going for walks is one of your favorite physical activities, you decided to add "walking" to your "Favorite" list in the app.

You like to exercise during lunchtime. You realized that it might be helpful if the app reminded you that it is time to go for a walk every day.

You decided to explore the reminder functionality and add a reminder for 12:00pm with note "Time to go for a walk!"

3.3.3. Use Scenario 3. In this scenario, you are a patient with diabetes. You had this condition for a while. In your last visit, your doctor mentioned the importance of setting goals for physical activity and your doctor asked you to exercise 30 minutes every day.

You have been using the app for a while and decided to explore the goal setting functionality of the app. You found the goal setting option in the app, enabled the activity goal "Daily Active Time", and set the time to "30 minutes".

As you explore, you saw that you could also set a weight goal. You decided to lose two pounds this month and set this as your "Period Goal".

3.4. Data Collection

Each subject had an individual session in the lab. When they first arrived at the lab, we first explained the study and then obtained their consent. Subjects were paid $50 for participating.

Subjects were then asked to fill out a questionnaire that gathers information about their demographics, health management and personal health information management habits. After completing the questionnaire, they received a short (about 10 minute) tutorial on the app. Next, we asked the subjects to sit in front of the eye tracker and go through a calibration process to facilitate accurate data collection.

Once the eye tracker was calibrated for subjects, we asked them to accomplish the use scenarios while recording their visual gaze on the smartphone. Each subject was asked to accomplish seven scenario-based tasks in the app: 1) add three different glucose readings, 2) add one weight reading, 3) add one mood record, 4) add one activity record, 5) add an activity reminder, 6) add an activity goal and 7) add a weight goal. While they were doing the scenario tasks with the apps, the researchers could observe the process via video that recorded their actions and their gaze. The researchers also took notes during this observation.

After completing all tasks, the subjects were asked to fill out an exit survey that focused on usefulness and ease of use of the app. We also asked the subjects to provide any additional comments they may have about the app.

3.5. Data Preparation

Before analyzing the collected video data, we first broke down every scenario-based task into a list of detailed sub-tasks that represent atomic steps the user needs to take to successfully complete the entire scenario. For example, a task of adding a glucose record were broken down into following sub-tasks: 1) click the glucose button on the main screen, 2) click "+" button on the upper corner of the glucose screen, 3) click "+" or "-" button to enter the correct glucose value, 4) click "OK" button to confirm the input and 5) click "OK" button on the feedback screen that acknowledges the input. Each sub-task was assigned a task type (navigation, data entry, confirmation or acknowledgement) and was associated with a screen and a UI component (e.g., a clickable button).

Using the data gathered by the equipment, we converted this into three milestone time-stamps for users' actions in each sub-task: 1) the time when the associated screen was first viewed; 2) the time the associated UI task component was first noticed (based on gaze fixation), and 3) the time the associated UI task component was clicked. Each time stamp was manually recorded in the following format: minute:second:frame. Using these data, we calculated several key measures for each sub-task: 1) noticeability (how long it takes for users to notice a UI component), 2) click delay (how long it takes for users to click on the button) and 3) time spent on each screen (how long a user stayed on a screen).

Originally, we planned to use the visual gaze fixation to a UI component to calculate noticeability. Noticeability of an object is measured as the amount of time that it takes for a subject to look at that object for the first time (the amount time elapsed until the first fixation on that object) [12]. In our study, fixation was defined as a steady gaze that lasted at least 100 milliseconds [10].

Unfortunately, visual gazes and fixations on UI components were not successfully recorded for the first three subjects. This may be due to the glasses the subjects were wearing, their eye anatomy, or movements that occurred after the calibration that resulted in subjects changing their position. Thus, we changed our analysis strategy for noticeability. Instead of calculating the time subjects took to fixate on the UI components after entering the UI screen, we calculated the time it took click on the UI components after they entered the associated UI screen. Since our screens are not as busy as a web site might be, this loss of detail differentiating noticeabililty and clicking is unlikely to be a problem.

We also noticed that the detailed sub-task lists were not exactly the same for all 5 subjects. Some of this variation may be due to users overlooking some of these tasks or following alternative (but not optimal) paths to achieve the same goal.

3.6. Data Analysis

Once we prepared the raw data and calculated the measures that were meaningful for us to understand issues related to usability, we compared results from the five subjects based on task accomplishment time, scenario accomplishment time, time to click, and time on screen. Task accomplishment time is calculated as the difference (in seconds) between the time when the first user screen for starting a task was viewed and the time when the last UI component was clicked to complete that same task. We also compared the time subjects spent on each type of sub-task (navigation, data entry, confirmation or acknowledgement). For each sub-task, the time spent was calculated as the difference between the time the first user screen was entered and the time the last user screen for that sub-task was left.

4. Usability Study Personas

Because our results are influenced by the background, attitudes and knowledge of our subjects about their health and its management and about smartphones, we present brief subject personas before presenting our results.

4.1. Subject 1

4.1.1. Basic Demographics. Subject 1 was a female, 55-64 years old, in good overall health (self reported), a 4-year college graduate, with less than 30K annual income. She used a basic mobile phone and had never used a smartphone at the time of the study. She was not sure how comfortable she would be using one, but was excited to getting her first smartphone in the next few weeks.

4.1.2. Health Management. Over the past 3 years, Subject 1 visited her primary care provider 2-3 times a year. She saw a specialist for diabetes care 4-6 times a year and a nutritionist once a year. She did not have diabetic wounds and hence did not see a wound specialist. She felt in control of her health and had a good understanding of her condition (diabetes) and available treatments. She has been taking measures to maintain a healthy lifestyle. She trusted her doctors and followed their instructions.

4.1.3. Personal Health Information Management. Subject 1 used an online PHR to store her primary doctor contact information and information about her appointments. She kept paper notes on glucose levels, weight, physical activity, daily diet logs, and mood. She also stored diabetes related information on paper.

4.2. Subject 2

4.2.1. Basic Demographics. Subject 2 was a female, 45-54 years old, in very good overall health (self reported), with some college experience or a 2-year degree, with annual income between 30-50K. She used a basic mobile phone and had used a smartphone on rare occasions. She was not sure how comfortable she would be using one.

4.2.2. Health Management. Over the past 3 years, Subject 2 visited her primary care provider 2-3 times a year. She did not see any specialists (for diabetes care, nutritionist, for wounds). She felt in control of her health and had a good understanding of her condition (diabetes) and available treatments. She has been taking measures to maintain a healthy lifestyle. She trusted her doctors and followed their instructions.

4.2.3. Personal Health Information Management. Subject 2 used her mobile phone and paper to store her primary doctor contact information. She stored information about her appointments on paper. She kept paper notes on glucose levels, daily diet logs, and mood. She also stored diabetes related information on paper. She did not keep any record of weight or physical activity.

4.3. Subject 3

4.3.1. Basic Demographics. Subject 3 was a female, 35-44 years old, in very good overall health (self reported), with a graduate degree and an annual income between $70-90K. She had been using a smartphone for more than 6 years at the time of the study and she described her use frequency as “all the time”. She was an experienced smartphone user.

4.3.2. Health Management. Over the past 3 years, Subject 3 visited her primary care provider 2-3 times a year. She saw a specialist for diabetes care 4-6 times a year and a nutritionist once a year. She did not have diabetic wounds and hence did not see a wound specialist. She did not feel in control of her health and she was not sure if she had a good understanding of her condition (diabetes) and available treatments. Although she has been taking measures to maintain a healthy lifestyle, she was not sure what else she could do for her health. She followed her doctors’ instructions and understood them, however, she was not sure about whether she trusted them or not.

4.3.3. Personal Health Information Management. Subject 3 used an online PHR to store her primary doctor contact information. She stored information about her appointments electronically (computer, mobile phone and PHR). While she stored her glucose levels only on her computer, she stored weight, activity, daily diet logs, and mood both on her computer and her mobile phone. She stored diabetes related information online in a PHR.

4.4. Subject 4

4.4.1. Basic Demographics. Subject 4 was a male, 55-64 years old, in good overall health (self reported), a 4-year college graduate, with more than 90K annual income. He was very comfortable using a smartphone and has been a user for 4-6 years.

4.4.2. Health Management. Over the past 3 years, Subject 4 visited his primary care provider 2-3 times a year. He did not see any specialists (for diabetes care, nutritionist, for wounds). He felt in control of his health but he did not have a good understanding of diabetes or available treatments. He has been taking measures to maintain a healthy lifestyle. He trusted his doctors and followed their instructions.

4.4.3. Personal Health Information Management. Subject 4 used a computer and an online PHR to store his primary doctor contact information. He stored information about his appointments and educational material about diabetes on his computer. He did not store glucose, weight, activity, diet or any other health related information.

4.5. Subject 5

4.5.1. Basic Demographics. Subject 5 was a female, 35-44 years old, in very good overall health (self reported), a 4-year college graduate with an annual income between $30-50K. She had been using a smartphone for 4-6 years at the time of the study and she described her use frequency as “all the time”. She was an experienced smartphone user.

4.5.2. Health Management. Over the past 3 years, Subject 5 visited her primary care provider 2-3 times a year. She did not see any specialists (for diabetes care, nutritionist, for wounds). She felt in control of her health and had a good understanding of available treatments, but was not sure if she understood her condition (diabetes) and its causes. She has been taking measures to maintain a healthy lifestyle. She trusted her doctors and followed their instructions.

4.5.3. Personal Health Information Management. Subject 5 stored her primary doctor contact information on paper. She stored information about her appointments on her mobile phone. She was not asked to keep track of glucose levels. She did not store any record of weight, activity, diet or any other health information.

5. Usability Study Results

The results of analyzing the time stamps we collected, in terms of noticeability times, time on screens, and task completion times, are presented below. In section 6, we discuss the variations we observed by tasks, screens, and subjects, and the implications of those variations for the design of our app.

5.1. Time on Task Screens

Figure 2 shows the time subjects took until they clicked on the first UI component on each screen. Figure 3 shows the time subjects stayed on each screen as they were completing use scenario tasks.

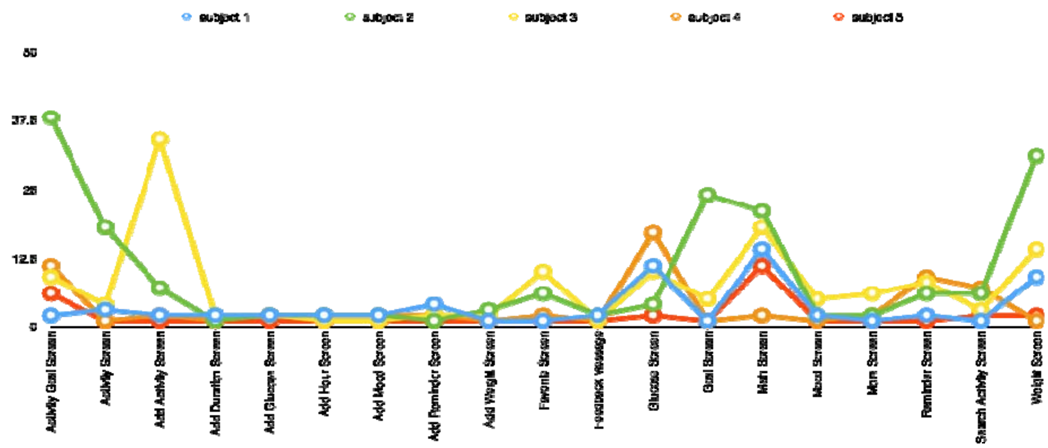

Figure 2 – Time until first click on each screen (in seconds)

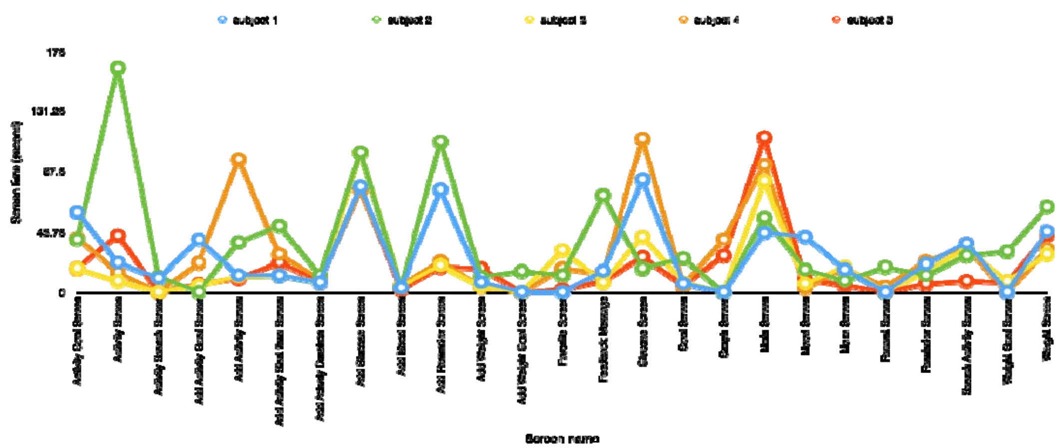

Figure 3 – Total time spent on each screen (in seconds)

We observed that some tasks required more time than the others, that is, although time to first click was fairly quick, time spent on that screen was high (e.g. add glucose and add reminder). We also observed the opposite where time to first click on some screens (e.g. activity goal and weight screen) was long but overall time spent on screen was low.

5.2. Task Completion

Figures 4 and 5 present task completion times and how they vary across subjects. Figure 4 shows completion times by subject for the seven scenario-based tasks discussed in the methodology (section 3.4). Figure 5 shows completion times by subject for the four task types defined in section 3.5.

6. Discussion

Using the task and screen time results, we discuss how and why task completion times vary consistently. Then we use the subject personas to explain how and why times vary between subjects. Finally, we discuss the design implications of these analyses.

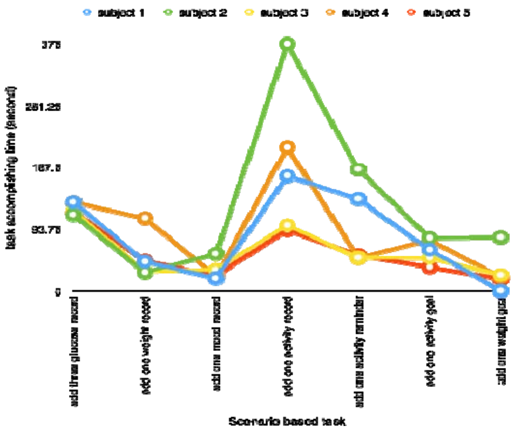

Figure 4 - Task completion time for use scenarios

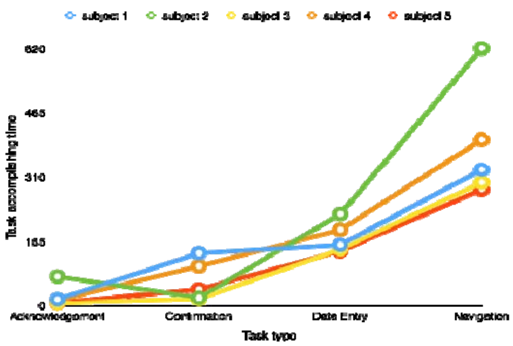

Figure 5 - Task completion time based on task type

6.1. Task Time Variations across Tasks

Some use scenarios were long by design. For example, the glucose adding task (task one in Figure 4) asks subjects to add three different glucose values that were very different from each other (125, 250 and 75). Because numbers are entered by incrementing or decrementing the previous entry, this takes some time, but is generally not confusing.

In contrast, adding data about one physical activity (task 4) is complex and thus is the most time-consuming task because it requires the subject to select an activity before entering the value for that activity. The design implications of these observed long task times are discussed in section 6.3.

Because the scenario tasks shown in Figure 4 are presented in the order subjects encountered them, these results also provide evidence of learning. For example, adding weight (the second task) took less time than adding glucose (the first task), although this is confounded by adding three glucose data points and only adding one weight data point. The third task, adding mood, shows some learning from adding glucose and adding weight, which is expected because data entry for these three are similar and this is their first use of the app. Not all the mood entries are faster, however. As we watched the subjects enter mood, we understood why data entry for mood is not entirely consistent with glucose or weight data entry, a fixable design problem.

The last three of the seven tasks in Figure 4 are a little different because they start from the "More" button on the main screen rather than having their own main screen button. Thus, we see learning for tasks 6 and 7 over task 5, the first of the tasks requiring the "More" button. Similarly, adding a weight goal (task 7) is faster than adding an activity goal (task 6).

Figure 4 indicates that some tasks are longer, and for those we must consider whether there are good ways to improve the design. Figures 2 and 3 provide more details at the screen level about which screens used in a task are longer or shorter. We also conclude that the app is relatively easy to learn because we see immediate learning rather than confusion as subjects performed tasks similar to a previous task.

Figure 5 illustrates a trend by task type, varying from the short time required before clicking OK after the app acknowledges data entry to the longer time for navigation. Data entry by nature takes extra time.

6.2. Task Time Variations Across Subjects

Despite the differences across personas among subjects, we did not observe a match between patterns and personas. Specifically, subject 2 (little smartphone experience) was slower on time to first click and overall task completion where as subject 1 with similar experience did not present similar time patterns.

6.3. Design Implications

The purpose of usability studies is to find any design problems or even minor issues that could be improved to provide a better user experience. Below we present the issues we uncovered from our usability study and how we might or did improve the app design based on this new knowledge. Because lab subjects were spread across several weeks, we were able to make most of the small design changes immediately and test them with later subjects.

6.3.1. UI Inconsistencies and Minor Errors. Because our app was relatively well tested before this study through our Scrum design, development, and testing cycles, the app worked well during our usability study, but we did discover a number of consistencies in the UI. For example, the order of presenting mood data differed on the two screens showing it, an easily fixable problem. In addition, the background color for some data entry boxes was orange, indicating a display value, rather than blue, indicating clickable for data entry, also easily fixable. Because the change in the color scheme and consistency was a relatively recent iteration, we found that the color scheme was not as well tested as functionality added in earlier iterations.

6.3.2. Navigation and the Return-to Screen. One difficulty with navigation was that when a task was finished the app did not return to the main screen but rather to the previous screen (the Android OS default), which was often a point partially through a set of related data entry screens. This confused subjects, who sometimes had difficulty getting back to the main menu. As a result, we determined where the end of an operation should return to and explicitly coded that, solving most navigation issues captured in our results.

6.3.3. Data Entry. When entering values for goals, subjects had difficulty because an old or default value was already there and subjects did not realize that in touching the data entry box, they were also picking a particular point in the old value for the cursor. Then they could not figure out how to move the cursor. We decided when a goal data entry box was selected, the app would clear the old value and the cursor would be set at an appropriate place for entering a new value.

A more challenging design problem was data entry related to selecting a physical activity before entering how much of that activity the subject did. The app uses a standard database of physical activities, which allows us to collect comparable data to other research. To select an activity subjects searched this database, which requires typing a search word using Android's pop-up keyboard. This was difficult for some subjects.

To address this typing issue, we took a longer term view. Once an activity is selected, it goes into a "recent" list and can be added to a "favorites" list. Thus, the typing problem mostly occurs once at setup, rather than being part of normal input. We now recommend that the app be setup for each individual user in consultation with their provider team. This would involve all the data entry for setting goals, as well as selecting a set of physical activities and putting them in the user's favorites list.

A final data entry concern is with the normal data entry of glucose, weight, and physical activity values. As noted in the results, adding values that differ significantly from previous values, e.g., glucose, can be slow. One solution we added is that long pressing rather than clicking the "+" or "-" resulting in fast incrementing or decrementing, which speeds entry of very different values. The other solution, which we had already implemented but did not test in our usability study, is automated entry of glucose and weight from a Bluetooth-enabled glucometer and scale (primarily because we did not seek IRB approval to use such devices in our study). These two devices solve the data entry problem for glucose and weight.

6.3.4. User Confusion Points. Two design choices about the app were confusing to subjects. The first, a minor easily fixable problem, was the app's default to metric units for entered values, e.g., kilograms rather than pounds for weight. This default was changed.

Subjects also found the physical activity favorites list confusing because they did not understand how it worked. Only one subject could add a physical activity to the favorites list without some additional direction. This subject was an experienced smartphone user and had seen such things before. We concluded that adding to the favorites list must be either part of setting up the app for a patient or explicitly taught (or both).

6.3.5. Requests for Added Functionality. In general, subjects found the app easy to use and something they would be interested in using. They also had requests for additional functionality. One request that we added quickly was a goal line on the graphs. Before and after entering glucose, weight, or physical activity values, the app shows a graph of past values. Adding a line showing a user's selected goal for that activity was consistent with our goals for the app in that it helped users see their progress toward their goals.

Subjects also suggested adding the ability to sync with other apps, diet tracking, and to automatically track physical activity rather than manually entering activity durations. While all of these are good ideas, they involve substantial development that must be separately planned and designed. In fact, the latter two ideas are significant research projects in themselves.

6.3.6. Design Tradeoffs. Some of the usability issues we observed in the usability study are the result of explicit design choices made to accommodate the preferences and abilities of our target users. For example, we use large buttons so that our users can see them and touch them without accidentally touching other things. As a result, we need a "more" button on the main screen for additional functionality. This is not necessarily a problem, because the functions available through "more" are mainly the setup functions, which users rarely need. In our usability study, however, the "More" button is a little confusing on first use.

A second explicit design decision was avoiding typing whenever possible. As result, physical activity duration (also manually entered glucose and weight) require clicking a large "+" or "-" button rather than typing in a number. We did this at the request of participants in our patient focus groups who preferred such data entry. It does slow data input, but more importantly, it avoids typing. Thus, although efficient data entry is ideal for a good user experience, it is more important to accommodate the needs of target users.

7. Conclusion

Conducting our usability study in our user experience lab with its eye-tracking equipment provided us with very detailed and useful data for improving the user experience with our diabetes app. The data allowed us to analyze which tasks and screens took longer and why. With our subject personas, we were also able to understand why some tasks and screens had high variations among subjects. These analyses provided fairly direct design implications. Some of the design problems identified had relatively straight-forward solutions that could be easily implemented. Others were evidence of design tradeoffs for which there were no clearly optimal solutions. Even for problems that are not easily solved, collecting detailed user experience information allowed us to more fully understand the design problems and associated design tradeoffs.

Acknowledgements

This research was supported by the National Science Foundation under Grant IIS-1065298. Any opinions, findings, conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

8. References

[1] Pagoto, S., Schneider, K., Jojic, M., DeBiasse, M., and Mann, D.: 'Evidence-based strategies in weight-loss mobile apps', American journal of preventive medicine, 2013, 45, (5), pp. 576-582

[2] Schwaber K., "Scrum development process", Business Object Design and Implementation, Springer 1997, p. 117-134.

[3] Galletta, D.F. and B.K. Dunn, "Assessing Smartphone Ease of Use and Learning from the Perspective of Novice and Expert Users: Development and Illustration of Mobile Benchmark Tasks", AIS Transactions on Human-Computer Interaction 6(4), December 2014, pp. 74-91.

[4] Strong, D., B. Tulu, E. Agu, Q. He, P. Pedersen, L. Wang, R. Ignotz, R. Dunn, S. Pagoto, and D. Harlan, "Design of the Feedback Engine for a Diabetes Self-care Smartphone App.", Proceedings of the Americas Conference on Information Systems, Savannah, GA, August 7-9, 2014.

[5] Wang, L., P.C. Pedersen, D.M. Strong, B. Tulu, E. Agu, and R. Ignotz, "Smartphone-Based Wound Assessment System for Patients With Diabetes", IEEE Trans. on Biomedical Engineering, Vol. 62, No. 2, February 2015, pp. 477-488.

[6] Wang, L., P.C. Pedersen, D.M. Strong, B. Tulu, E. Agu, R. Ignotz, and Q. He, "An Automatic Assessment System of Diabetic Foot Ulcers based on Wound Area Determination, Color Segmentation and Healing Score Evaluation", Journal of Diabetes Science and Technology, 2015, Accessed Sept 11, 2015, http://dst.sagepub.com/content/early/2015/08/07 /1932296815599004.full.pdf+html.

[7] Alter, S., "Designing and Engineering for Emergence: A Challenge for HCI Practice and Research", AIS Transactions on Human-Computer Interaction 2(4), December 2010, pp. 127-140.

[8] Norman, D.A., "Natural User Interfaces Are Not Natural", ACM Interactions May+June, 2010, pp. 6-10.

[9] Norman, D.A. and J. Nielsen, "Gestural Interfaces: A Step Backward In Usability", ACM Interactions September +October 2010, pp. 46-49.

[10] Djamasbi, S., "Eye Tracking and Web Experience", AIS Transactions on Human-Computer Interaction 6(2), 2014, pp. 37-54.

[11] Buscher, G., E. Cutrell, and M.R. Morris, "What do you see when you're surfing? Using eye tracking to predict salient regions of web pages", Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2009, pp. 21-30.

[12] Djamasbi, S., M. Siegel, and T. Tullis, "Designing Noticeable Bricklets by Tracking Users' Eye Movements", in Proceedings of the Forty-Fifth Annual Hawaii International Conference on System Sciences (HICSS), IEEE Computer Society Press, 2012, pp. 525-532.

[13] Djamasbi, S., M. Siegel, and T. Tullis, "Visual Hierarchy and Viewing Behavior: An Eye Tracking Study", in Human-Computer Interaction Design and Development Approaches, Lecture Notes in Computer Science 6761, Springer Berlin Heidelberg, 2011, pp. 331-340.

[14] Biedert, R., G. Buscher, and A. Vartan, "Reading and Estimating Gaze on Smart Phones", Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA), ACM, 2012, pp. 385-388.

[15] Djamasbi, S., D. McAuliffe, W. Gomez, G. Kardzhaliyski, W. Liu, and F. Oglesby, "Designing for Success: Creating Business Value with Mobile User Experience (UX)", in HCI in Business, Lecture Notes in Computer Science 8527, Springer Berlin Heidelberg, 2014, pp. 299-306.