EXPLORING THE INFLUENCE OF PROTOTYPING FIDELITY ON FEEDBACK FOR AN APP WITH MULTI-SIDED INTERACTION MODEL

by

Yaman Sanobar

M.S., Kobe Institute of Computing, 2020

B.S., Princess Sumaya University for Technology, 2016

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Computer Science)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

November 2023

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Exploring the Influence of Prototyping Fidelity on Feedback for an App with Multi-Sided Interaction Model

submitted by Yaman Sanobar in partial fulfilment of the requirements for

the degree of Master of Science

in Computer Science

Examining Committee:

Ian Mitchell, Professor, Computer Science, UBC
Supervisor

Dongwook Yoon, Associate Professor, Computer Science, UBC
Supervisory Committee Member

Abstract

The fidelity of mobile app prototypes affects the feedback received from participants testing the prototype. The relationship between fidelity and feedback relies on context, and it can be hard to choose the right fidelity when implementing a prototype. Exploring the differences in feedback between two prototypes of different fidelities is a step toward allowing this decision to be made in a more informed and less context dependent manner. In this study, we tested two prototypes of a multi-sided booking app: a high-fidelity fully functional mobile app developed using Flutter and an equivalent low-fidelity paper prototype presented on phone-size pieces of paper. In our between-group pilot experiment, we recruited 12 participants to carry out predefined tasks, and each group tested one of the prototypes. We collected quantitative data using the System Usability Scale survey, and we collected qualitative data using semi-structured interviews. The analysis of collected data showed that the effect of fidelity on quantitative rating of usability is minimal. However, it showed differences in the qualitative feedback between the two groups of participants. For instance, the low-fidelity groups commented on text-based content and their suggestions were about introducing new ideas to expand the current design; in contrast, the high-fidelity group commented on colour and image-based content, and their suggestions were about enhancing the current design.

Lay Summary

This is a pilot study to explore the influence of prototyping medium on feedback received from participants when testing a mobile booking app. We used two prototypes that were tested by two groups of participants: A low-fidelity paper prototype and a high-fidelity fully functional mobile app. Both prototypes represent the same app, which is an app for renting outdoor adaptive wheelchairs. The prototypes are identical in functionality but are different in their format or fidelity. We collected different types of data while testing the prototypes with participants, then we analyzed the data to present our results. Analysis showed that usability ratings from a standard survey were similar for both groups, but when participants were given the opportunity to provide open-ended feedback, each group commented on different aspects than the other group and the nature of suggestions was different for each group.

Preface

The work presented in this thesis was conducted in the Collaborative Robotic Lab at the University of British Columbia under the supervision of Dr. Ian Mitchell. I designed and implemented the prototypes, designed and conducted the study, analyzed the data, and wrote the thesis with feedback from Dr. Mitchell. The experiment presented in Chapter 4 was reviewed and approved by UBC's Behavioural Research Ethical Board under certificate number: H23-01482. Two graduate students, Justin Reiher and Yuan Tian, helped in testing the experiment and provided valuable feedback to enhance the procedure of the experiment.

Table of Contents

List of Figures

Acknowledgements

I offer my enduring gratitude to the faculty, staff and my fellow students at UBC, who have inspired me to continue my work in the field of Human-Computer Interaction. I owe thanks to Dr. Ian Mitchell, whose discussion taught me to question more deeply, fostering a critical mindset and encouraging a broader perspective.

I thank Dr. Dongwook Yoon for offering insightful comments on my thesis. I also thank Dr. Karon MacLean for providing coherent feedback to my questions. I thank all the staff at the CS department for their support.

Special thanks are owed to my family, who have supported me throughout the years of my education.

Chapter 1: Introduction

A prototype is an essential part of the design process. A prototype is a representation of the final product that enables designers to communicate and evaluate design choices with stakeholders. Prototypes are manifestations of design ideas, and due to their incomplete nature, they reveal certain aspects of design ideas and focus on certain qualities that are under exploration by the designer [1]. Hence, prototypes are not used for proving ideas, but to discover problems and to generate new ideas. Fidelity is one of the most important aspects that define a prototype, and it is used to characterize in what level of detail the concept of the final product is illustrated; in other words, whether the prototype needs to be complete and realistic (so-called “high fidelity“) or low-cost and quick to implement ("low fidelity") [2]. More rigorously, researchers have identified five dimensions of fidelity that characterize a prototype of a human-computer interaction: level of visual refinement, breadth of functionality, depth of functionality, richness of interactivity, and richness of data model [3]. Each of these dimensions can be manipulated independently, which enables the design of mixed-fidelity prototypes by choosing a different level of fidelity for each dimension.

The importance of the fidelity question arises from the early steps that designers take to create a prototype. Choosing the right fidelity for a given task is context-dependent and often far from straightforward [2]. In this study, we are trying to answer the fidelity question for a certain context, but most importantly we are trying to reproduce a connection reported in the literature (for similar but not identical user interfaces) between the fidelity of a prototype of a mobile app and the feedback received from testers. We wanted to investigate how the fidelity of a prototype would affect the responses received from testers. Hence, the process of choosing the right fidelity initially could be more informed and less dependent on context.

In this study, we tested a paper prototype and a mobile app with a fully functional user interface across two groups of participants, where each group tested one of the prototypes. The app is a multi-sided booking app for renting outdoor wheelchairs which was developed using Flutter for Android and iOS. Following the testing, participants answered a quantitative usability survey, and then we conducted a semi-structured interview with each participant, asking about their interaction with the proposed app. We analyze the quantitative data statistically, and we analyze the qualitative data using content analysis methods to answer our research questions.

In the next chapter, we explore current literature concerned with the fidelity question for apps. Then we define the scope of the study and state our research questions. In Chapter 3, we examine the proposed application and the technologies used to develop the high-fidelity mobile app, and present the paper prototype. In Chapter 4, we note the details of the experiment: the participants, the tasks used in the testing sessions, the procedure of conducting the sessions, and the methods of data collection. In Chapter 5, we illustrate the results of the analyzed data and discuss our findings and the study's limitations.

Chapter 2: Related work

It is widely accepted that differences in prototype fidelity influence feedback. Prototypes are commonly designed to target one of the two ends of the spectrum in most or all dimensions; consequently, they can be roughly labeled as "low-fidelity" or "high-fidelity." Both types have their value and place in the design space [2], [4], [5]. Studying the effect of fidelity on user feedback is not new, and several studies have explored this effect from different perspectives.

2.1 Low-fidelity vs high-fidelity

In general, high-fidelity prototypes take too long to build and modify in contrast to low- fidelity prototypes, which maximizes the number of design refinements that could be applied to the low-fidelity prototype before committing to code [6]. High-fidelity prototypes provide a more realistic and immersive experience, while low-fidelity prototypes provide a more abstract canvas that may lead to the generation of more creative ideas during testing [7]. On the other hand, feedback received from high-fidelity prototypes might be oriented toward aesthetics instead of content, while the design suggestions from low-fidelity prototypes might turn out to be inapplicable for reasons obscured by their abstraction. In particular, the difference in visual fidelity for a game prototype did not affect the magnitude of usability issues discovered by usability testers [8]. The visual stimuli (in the form of text or pictures) affected the results of participants asked to generate new ideas for a design concept; the presence of pictures increased the rarity and non-obviousness of ideas but did not affect quality, quantity, or originality [9]. In a study of testing and comparing low and high-fidelity prototypes of a mobile messaging feature, none of the participants testing the paper prototype suggested new design ideas, in contrast to participants testing other conditions [10].

Most notably, aesthetics has the biggest influence on feedback bias: The greater fidelity, the stronger the critical perception of users becomes [11]. As specific examples, child testers of a game design concept considered the lack of colour for a low-fidelity prototype as a negative aspect [12], while testers of a messaging mobile app overrated the aesthetic qualities of the lower-fidelity prototype [13].

Also observed: user expertise and the presence of the observer influenced the evaluation of prototypes. In one study, using reduced fidelity prototypes to determine the behaviour of users with real appliances led to an overestimation of control settings. Interestingly, in this study expert users reported more usability issues than novices, but the issues identified by novices were more severe [14]. In another study, the presence of an observer had a more negative impact on the performance of testers when testing a fully functional system than an early prototype [15]. Lastly, in a study about eliciting user requirements from low-fidelity and AR-enhanced prototypes of a fan product, researchers demonstrated that both prototypes achieve similar performance regarding eliciting user requirements. However, the AR-enhanced system shows potential in providing a better overall experience and in stimulating the engagement of participants [16].

As for usability, while some researchers concluded that low-fidelity prototypes are sufficient to use at all design stages to extract usability problems [17], other researchers found that “the lower the fidelity, the more people idealize a product idea” [18]. In a study to test the effect of the prototype medium on usability testing outcomes, it was found that users were able to identify significantly more usability problems in a higher fidelity medium [19]. In another study testing mobile phone prototypes, it was found that testers of the high-fidelity prototype criticized the choice of fonts, colour, and button sizes; in contrast, users of the low-fidelity prototype commented on content (task flow, general layout, terminology) rather than appearance (presumably due to its simplified form) [20]; likewise, in a study testing medical devices researchers found that prototype format influenced the type of response provided by users [21]. That does not mean that creating higher fidelity prototypes is a wasted effort, as researchers have found that the extra design information gained more than offsets the extra prototype creation effort [4].

2.2 Inconsistent Outcomes

As can be seen, many studies have been done around exploring the effect of prototyping fidelity. However, each study has its own context and goal, which led to what appears to be inconsistent results. Potential causes of this inconsistency are [14]:

- the differences in the types of prototypes used and the target product model, or
- the factors examined, which are different in each study, such as aesthetics, usability errors, or efficiency measures[14].

In this study, our product model is a booking mobile app and our examined factors are feedback and idea generation. The study will clarify the role of prototyping fidelity in eliciting feedback and generating design concepts for the booking app. We expect that the study will enrich the literature and provide guidelines to better design prototypes to receive more informed feedback. Accordingly, we will not make assumptions about how fidelity will influence feedback, and we want to explore that effect in our context through the study that we conduct.

2.3 Scope and Research Questions

In this thesis, we are prototyping a multi-sided booking app. The context is a booking operation that requires the synchronization of the user, a volunteer, and the device to be rented within timeframes that are constrained with respect to one another. This model is "multi-sided" because it involves connecting or otherwise serving as the intermediary among two or more participant groups [22]. A successful app can create a sustainable and scalable ecosystem by connecting relevant groups and defining interactions among them [23]. Popular and successful examples of multi-sided booking apps are the apps for ride-sharing services such as Uber and Lyft.

In this study we wanted to include sketching which provides a rich medium that makes understanding concepts easier than using language [24]. Sketching also facilitates reflection and works as a potential complement for usability testing [25] as usability testing by itself is used to define problems not to provide solutions [26]. Hence, we provided participants with a pen and a piece of paper during the interviews and encouraged them to sketch while answering the questions to better convey any comments or design suggestions.

While there are many studies reported in the literature which explore the effect of fidelity on user feedback for other user interfaces, with the exception of [10], [13], [20]we were unable to find any academic studies specific to interfaces for mobile phone apps. In this study, we tested a fully functional mobile app beside a paper prototype. We examined what users say about the interaction after testing it, and we asked users to suggest design improvements to the multi-sided booking app. We wanted to examine if the feedback received from a high-fidelity prototype is different from the feedback received from a low-fidelity one, specifically in the case of a multi-sided booking mobile app. Our results may allow designers to better choose the fidelity in advance based on the desired type of feedback. The primary research question that we aimed to answer:

1. What are the differences in the feedback received from low-fidelity prototypes versus high-fidelity prototypes when testing a multi-sided booking mobile app?

The secondary research questions that we wanted to investigate relate to the suitability of this prototype for its intended audience:

2. How do users evaluate the usability of the prototypes for the tested multi-sided booking app?
3. What are possible improvements to the multi-sided booking app prototype?

Chapter 3: Prototype

The app is an adaptive equipment rental system that allows users to book all-terrain wheelchairs and other assistive mobility devices (which we generically abbreviate as “AMDs” below) and match them with volunteers who will help adjust the AMD appropriately before departure. Our low-fidelity prototype was created in Figma (https://www.figma.com/) and printed in greyscale onto phone-sized paper sheets which could be shown to the participant and which the participant could touch to simulate touch-based interaction with the app. In contrast, our high-fidelity prototype was a fully functional, full colour, touch-based mobile app developed for Android and iOS using the Flutter framework (https://flutter.dev/). The high-fidelity prototype was created first, and then the low-fidelity version designed to visually match, modulo some additional visual abstraction. As shown in Figure 3.1, both prototypes are identical in terms of layout and interaction process, but the low-fidelity prototype is visually slightly more abstract than the high-fidelity version.

The app accommodates three types of roles: the user, the volunteer, and the administrator. Each can log in with their credentials to access the corresponding interface. The user interface allows for booking AMDs and the volunteer interface allows to set time availabilities. The administrator interface is customized for the web, and it allows the admin to modify AMDs on the system, view bookings based on device or date, and configure the profiles of users and admins.

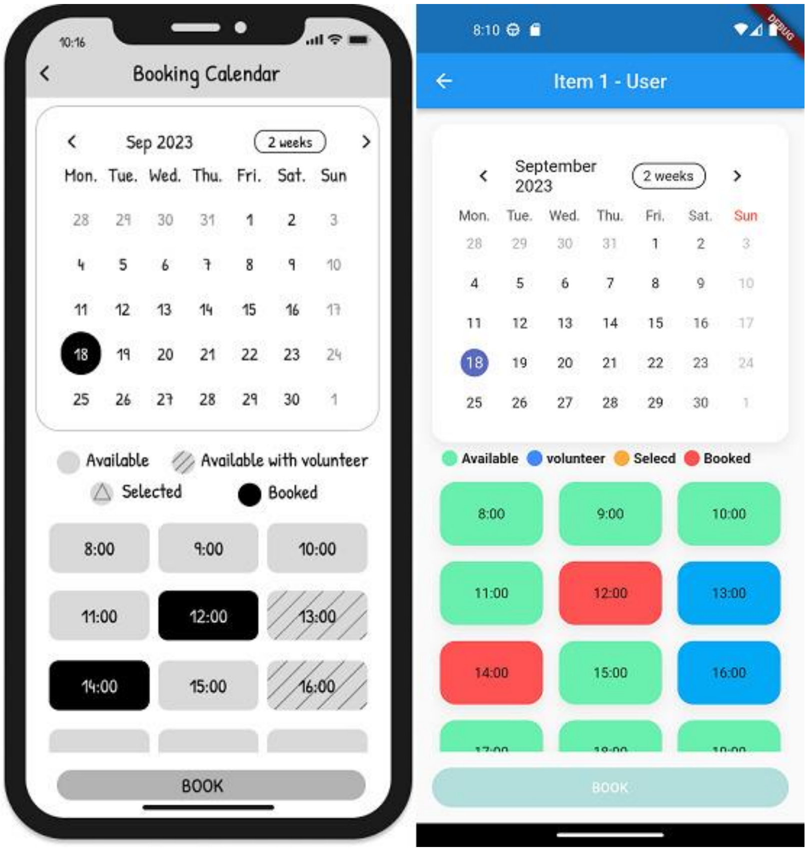


Figure 3.1 Left: Low-fidelity prototype designed in Figma and printed greyscale onto paper. Right: High-fidelity prototype designed in Flutter (screen capture from an Android simulator).

3.1 High-fidelity prototype

Through the app, users can choose an AMD from a list (Figure 3.2), and then choose from available time slots for that AMD in a calendar. The colour of time slots conveys different meanings: whether the slot is booked or available, and whether a volunteer is available at that time or not. Volunteers enter their availability for each device through a different interface where they can select dates and time intervals. The high-fidelity prototype is a fully functional app running on an Android device for the volunteer interface and an iOS device for the user interface during the study sessions.

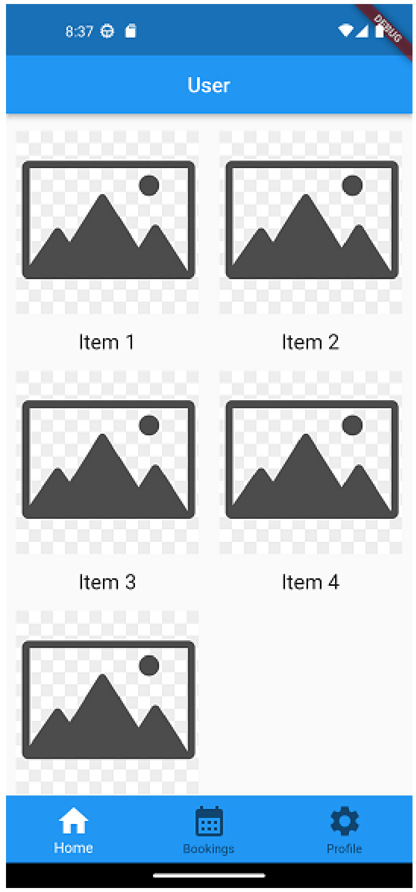


Figure 3.2 Home page of the mobile app for the user role interface (executing on the Android simulator).

3.1.1 UI Technology

Among the current hybrid app development frameworks, React Native from Facebook and Flutter from Google are considered the most popular [27], [28]. React Native uses JavaScript and Flutter uses Dart as their primary programming languages. Both frameworks enable developers to develop iOS, Android, and Web apps from a single code base. The main difference between the tools lies in their architecture [29]. React Native interacts with native controllers and APIs via the React Native Realm. It has a JavaScript realm as well that enables developers to write UI code in JavaScript, and the two realms can communicate with each other using the so-called JavaScript Bridge. In contrast, Flutter builds on ”widgets”, which are high-fidelity replicas of all the UI components found on the native platforms. The widgets follow the guidelines of Material Design on Android (https://docs.flutter.dev/ui/widgets/material) and the Cupertino specifications on iOS (https://docs.flutter.dev/ui/widgets/cupertino), which makes Flutter apps look like native apps without additional development effort.

Both frameworks allow rapid development of high-quality UI interfaces and have detailed and comprehensive documentation. React has exceptional performance, especially when it comes to web apps. Flutter is famous for its testing capabilities and rapid development. Choosing between the two is based on the situation [30], and because the focus in this project is on prototyping, we chose Flutter as its debugging tools make modifying and iterating on apps a much smoother task.

We started development from the main feature of the app: the user (as opposed to volunteer) booking process. Through the official package repository of Flutter, we examined several open-source booking-related packages that met our requirements for the booking process. We selected the *booking_calendar* package (https://pub.dev/packages/booking_calendar), which provides a nice looking calendar interface, and then modified its code to accommodate our imagined design. To handle state management, we used the popular *Provider* package (https://pub.dev/packages/provider). Flutter uses the concept of a "widget tree" where each UI element on the screen is a widget or a part of a widget in a hierarchy. Provider can be seen as the parent of all widgets [31], and "widgets listen to changes in state and update as soon as they are notified" [32] by Provider.

3.1.2 Backend Technology

We used Cloud Firestore (https://firebase.google.com/docs/firestore), a real-time NoSQL database provided by Firebase. Firebase also provided the following additional services to the high-fidelity prototype:

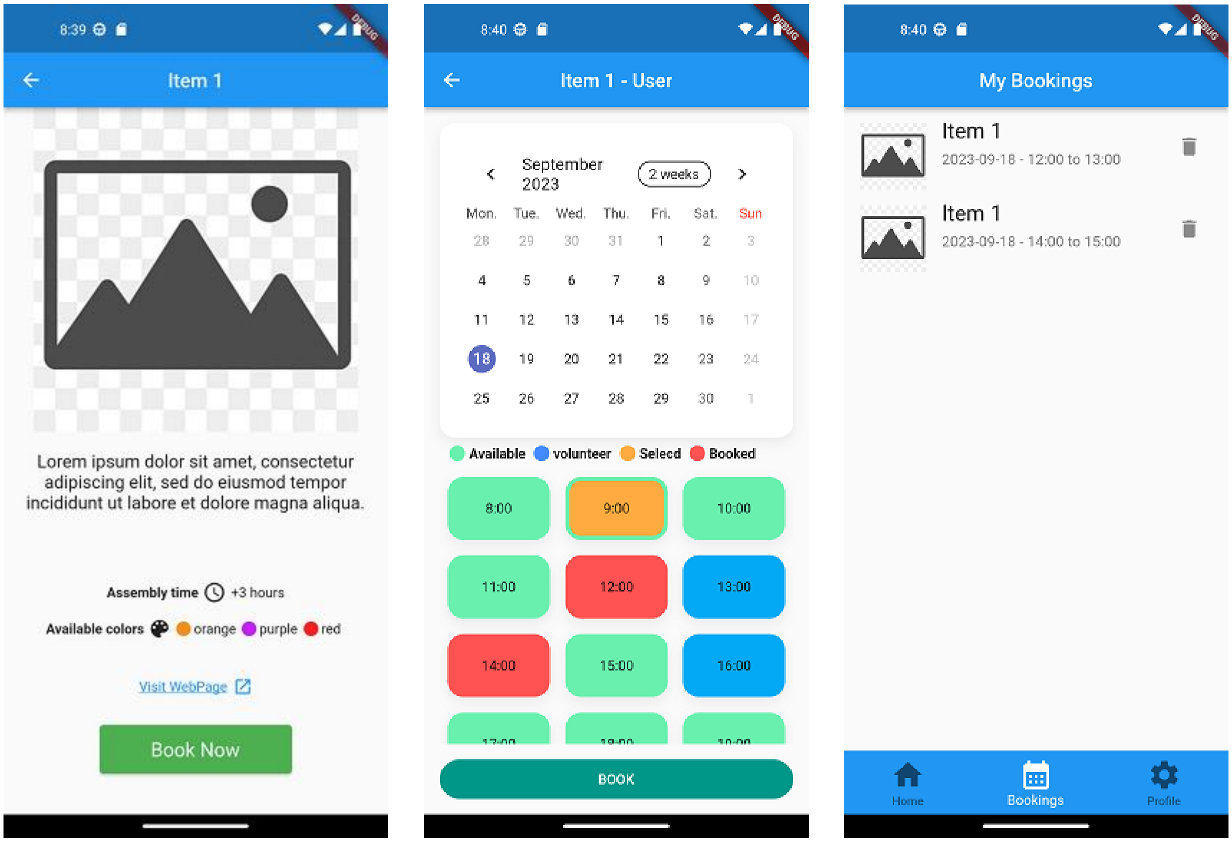
- Authentication to allow apps to save user data in the cloud and authenticate users to the app.
- Storage to store images of AMDs.

- Test Lab to run automatic tests that analyze the structure of the UI and simulate user activities.
- App Distribution to distribute app files to the devices of testers.

3.1.3 Distribution

We built versions of the app for Android, iOS, and the Web. The Web version's on-screen layout is optimized for the admin interface, while the mobile version's layout is configured to better support the users and the volunteers. From the single codebase of Flutter, we built APK, IPA, and Web installation packages. Then we distributed the APK to Android devices and the IPA to iOS devices through the App Distribution service provided by Firebase, while the Web package was installed at a GitHub Pages site.

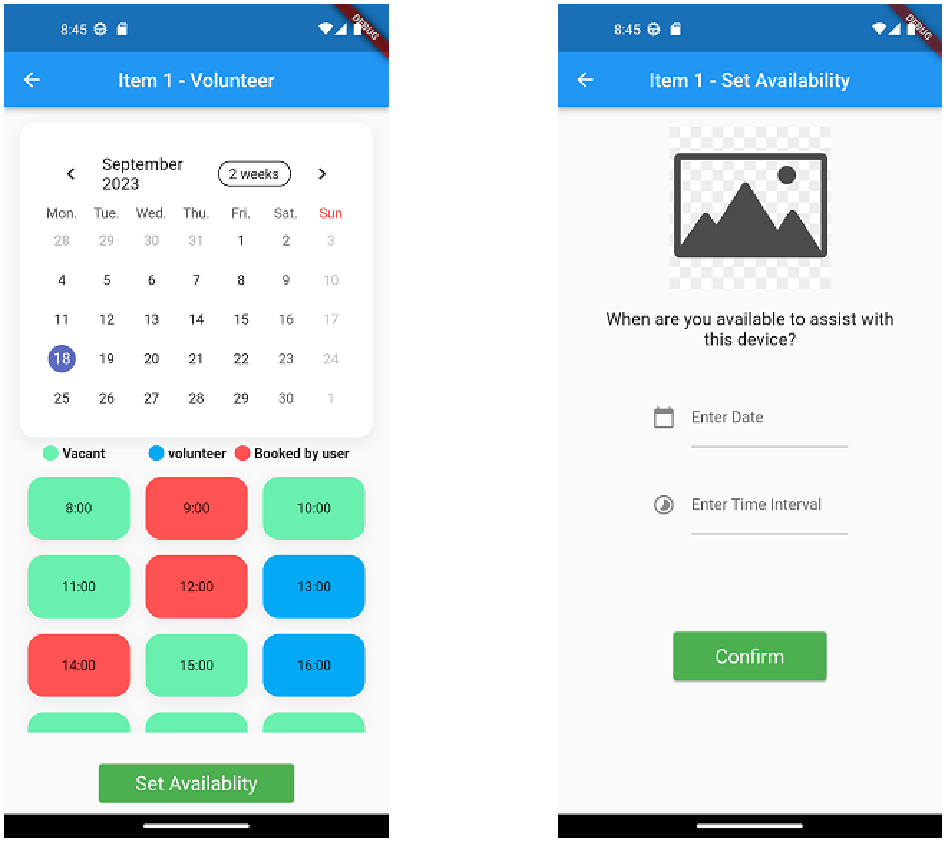
The user and the volunteer interfaces provide access to the list of AMDs (Figure 3.2). As shown in Figure 3.3 (a), after the user selects the desired AMD they see a brief description of the device, can access more information about the device through a weblink (typically to the manufacturer's website), and can book a time slot through a calendar (Figure 3.3 (b)). Depending on their needs, the user can select any "Available" or "Available with volunteer" slot and book it.



(a) (b) (c)

Figure 3.3 Screenshots from the user side interface to the app.
(a) Item 1 page. (b) booking calendar for item 1.
(c) "Bookings" screen where all bookings for the current user are listed.

For the volunteer interface, the main difference is that the volunteer does not "book slots" but "sets availabilities" instead. The volunteer's calendar does not allow selecting predefined slots like the user's calendar; instead, volunteers can enter their availability through a short form where they select a date and time interval as shown in Figure 3.4. This format was chosen both to distinguish the volunteer interface from the user interface—because volunteers may at times access the system as a user to borrow a device for themselves—and because we expect that volunteers will often be available over many booking slots in a row.

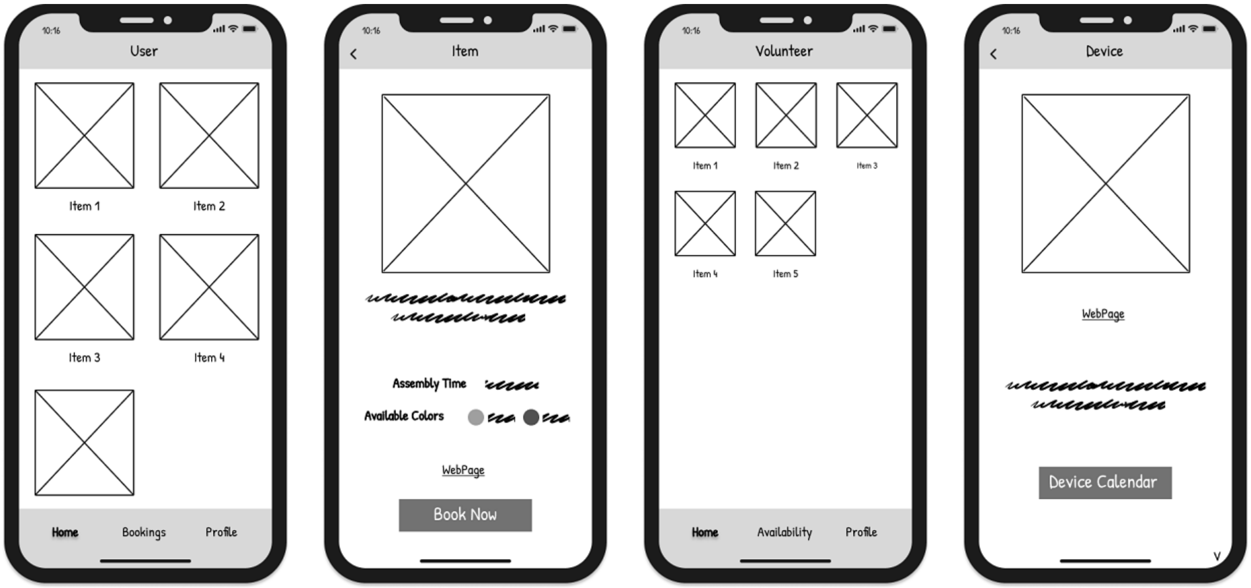


(a) **(b)**

Figure 3.4 Screenshots from the volunteer side interface to the app. (a) Calendar of the selected AMD. (b) Form to set the availability.

3.2 Low-fidelity prototype

The low-fidelity prototype is a paper prototype designed in Figma to resemble the screens of the high-fidelity prototype (Figure 3.5). However, the paper prototype has a greyscale colour palette and a more abstract visual language, in the sense that any data that is not necessary for interacting with the app was replaced with dummy data; for example, the images and text description of the AMDs. We printed the designed screens onto paper and cut them to the size of a phone. Further, to ease the interaction with the prototype at the time of testing, we set up stacks of papers where each stack corresponds to a task and each individual page within that stack represents a sequential step of that task.



(a) (b) (c) (d)

Figure 3.5 Screens from the paper prototype.

(a) The user side home page. (b) The user side device page.

(c) The volunteer side home page. (d) The volunteer side device page.

Chapter 4: Experiment

We used a 2 x 1 design in this study. The main independent variable is the prototype fidelity, which is manipulated at two levels: paper prototype and fully functional application. The prototype fidelity is a between-group factor, in which each participant tested one fidelity and provided feedback for the user and the volunteer interfaces. Participants were randomly assigned to one of two testing groups prior to the study sessions. This design is ideal in our context because we wanted the same participant to view the booking process from the two sides (to better understand the booking process) while experiencing just one fidelity (to keep participants from providing their feedback in a comparative form).

4.1 Tasks

The booking process is done through a calendar, and intuitive interaction is based on past experiences with familiar features [33]. Participants carried out three tasks in total: two on the user side and one on the volunteer side of the multi-sided booking model. From the user side, the main task is to book a certain device on a certain date considering the availability of a volunteer. From the volunteer side, the main task is to set time availability for a certain device. It has been found that longer exposure to an early prototype leads participants to develop more reflective feedback by overcoming first impressions and developing a deeper understanding of the design conceptual model [34]. Therefore, we asked participants to carry out a proximate scenario, thereby increasing the exposure participants had to the tasks and hopefully leading to better design suggestions.

4.2 Procedure

Before the testing sessions started, we randomly assigned all participants to either the low-fidelity or the high-fidelity prototype group. At the beginning of each session, the experimenter gave an overview of the problem that the app sought to solve and briefed the participant about the goal of the study. After signing the consent form, participants filled in a short online demographic survey, and then the experimenter introduced the assigned prototype. Participants were told that there was another prototype of a different format in a different medium that would be tested by the second group, but the "fidelity" of either prototype was not mentioned at this point because the concept of fidelity might be familiar to some participants and might cause bias in their responses. However, at the end of the study participants were informed that the study was about testing participant responses to low-fidelity vs high-fidelity prototypes.

Testing began by providing participants with guidelines on how tasks would be performed (see Appendix B.1). We asked participants to start performing the tasks, while the experimenter observed and took notes. Each participant carried out three tasks, and each task had a clear and sequential set of steps to accomplish. If the participant made a mistake and entered an incorrect or unimplemented component of the prototype, the experimenter intervened and returned the participant to a previous screen.

At the end of testing, we asked participants to complete an online usability survey (Appendix B.2). Following the survey, we conducted a semi-structured interview with each participant (Appendix B.3). In the interview, we asked participants about their interaction with the prototype, and we requested design suggestions or ideas to improve the multi-sided booking interaction. Finally, we debriefed participants about the full purpose of the study and provided them with an opportunity to give any additional comments.

4.3 Data collection

After performing all tasks but before the interview, we asked participants to fill in a survey that is a global assessment of usability called the System Usability Scale (SUS) [35]. Ideally, a survey of this form might be able to identify quantitative differences in perceived usability between the different fidelity prototypes; however, the small size of the participant pool makes this outcome rather unlikely.

Consequently, our primary focus in this study was on the qualitative data gathered from the semi-structured interview with each participant after testing. Interview questions were open-ended and were concerned with how to improve the multi-sided interaction experience. The interview was audio recorded, and the experimenter was taking notes during the interview. We provided participants with a pen and paper, and we encouraged them to sketch any design ideas or suggestions. Sketching was encouraged but was not mandatory; as a result, only 5 out of the 12 participants used the pen and the paper provided.

Chapter 5: Results

After reviewing the demographics of our participants, we analyzed the quantitative data to determine if there was a noticeable difference in responses between the low-fidelity and the high-fidelity testers. Since the focus of this study is on qualitative data and the number of participants is relatively small, we were not expecting a statistically significant difference. We used content analysis to analyze the qualitative data collected from interviews, and we distinguished different categories that emerged from the responses of participants. We discuss the differences in responses between the two groups and show the connection between fidelity and feedback. Lastly, we note the limitations of the study.

5.1 Participants

A total of 14 participants responded to our solicitation, but only 12 came to a study session and were randomly assigned to test a fidelity (so 6 to each level of fidelity). All 12 who began a session completed that session. Participants were required to be 19 years or older English-speaking adults who are comfortable using a smartphone device to interact with an app, respond to a survey, and answer interview questions.

The only demographic data collected was:

- Whether the participant had used a multi-sided booking app before (such as a car-sharing app). All of the participants but one answered in the affirmative.
- Whether the participant had previous experience in mobile app design. Five of the participants answered in the affirmative: Two participants had held an internship related to app design, two mentioned using Figma for designing interfaces or apps, and one of the participants was a product designer.

All but one of the participants mentioned previous use of multi-sided booking apps, such as car/taxi rental, grocery delivery, or event ticketing apps. One participant mentioned using a crowdsourcing app, and two participants mentioned using a property rental app. However, none of the participants mentioned using an item rental application similar to the app we are proposing.

5.2 Quantitative

The meaning of usability is summarized in [35] as the "general quality of the appropriateness to a purpose of any particular artefact." This paper also introduces the ten question System Usability Scale (SUS) questionnaire and provides formulas to compute a summative usability score based on the responses of those ten questions. In our study, we used the SUS survey but with a finer granulation Likert scale that has a range from 0 to 100 instead of the usual range from 1 to 5. After adaptation to our finer-grained scoring from 0 to 100, the formulas for the individual questions are:

$$odd\ question\ score = (1 + \left(\frac{(userReport - 0) * (5 - 1)}{(100 - 0)}\right)) - 1$$

$$even\ question\ scores = \ 5 - (1 + \left(\frac{(userReport - 0) * (5 - 1)}{(100 - 0)}\right))$$

Note that these formulas produce a score in the range 0 to 4. Different formulas for odd and even questions are used because the odd questions are stated in a form such that a higher participant response is more usable, while the even questions are stated in the opposite form (a lower response is more usable). After conversion by the formulas above, a higher "score" always represents more usable.

Figure 5.1 shows the average scores across participants for each question, separating the two fidelity groups (for detailed individual scores, see Appendix C). As we expected, given the range of responses among participants in each group, the relatively small differences between the averages of the two groups are clearly not statistically significant. Furthermore, there does not appear to be any pattern to the small differences that are apparent: Questions 1 and 4 show the screen version as more usable, questions 5, 7, and 10 show the paper version as more usable, and the scores for questions 2, 3, 6, 8 and 9 are essentially indistinguishable.

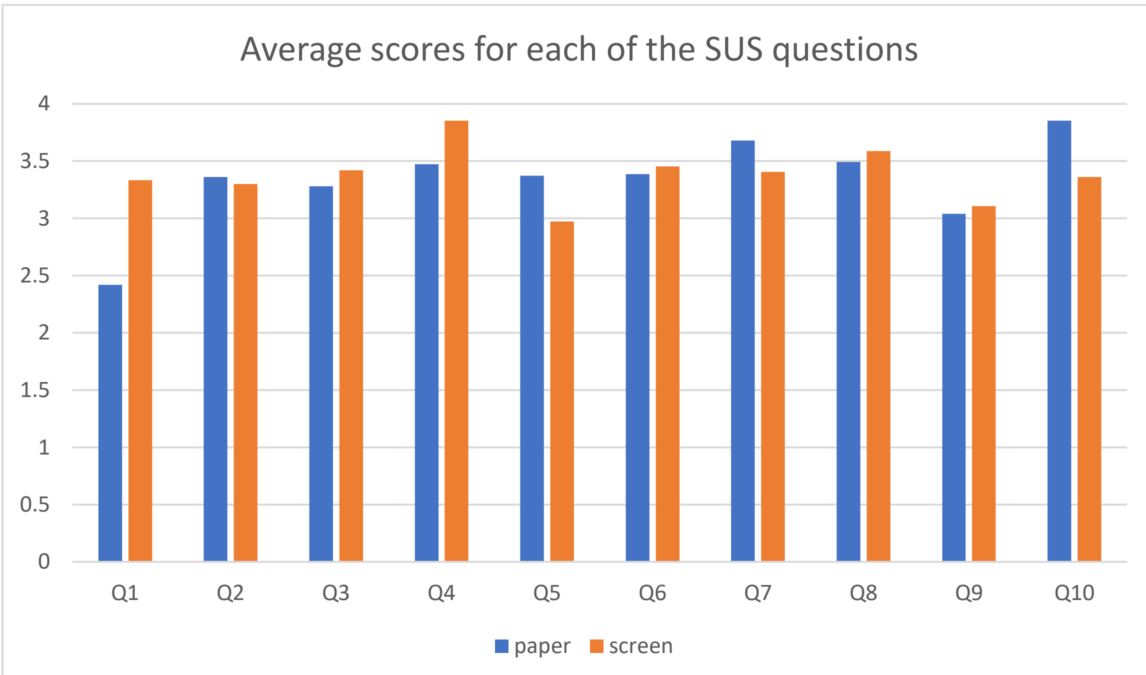


Figure 5.1 The means of scores by question between paper and screen participants

The usability score for each participant is the sum of the scores for the ten questions answered by that participant. Figures 5.2 and 5.3 show summative usability scores for each individual participant. The average summative score for the paper group is 33.36, and for the screen group is 33.79. The overall SUS score is obtained by multiplying the summative scores by 2.5, hence the SUS score equals **83.4** for the paper group and **84.5** for the screen group. The SUS score ranges from 0 to 100, and in a subsequent paper [36]the authors provide seven adjectives to describe different portions of this range from “worst imaginable” to “best imaginable”[36]. The range from 71.4 - 85.5 is considered “Good” (third highest rating) and the range from 85.5 – 90.9 is considered

“Excellent”.

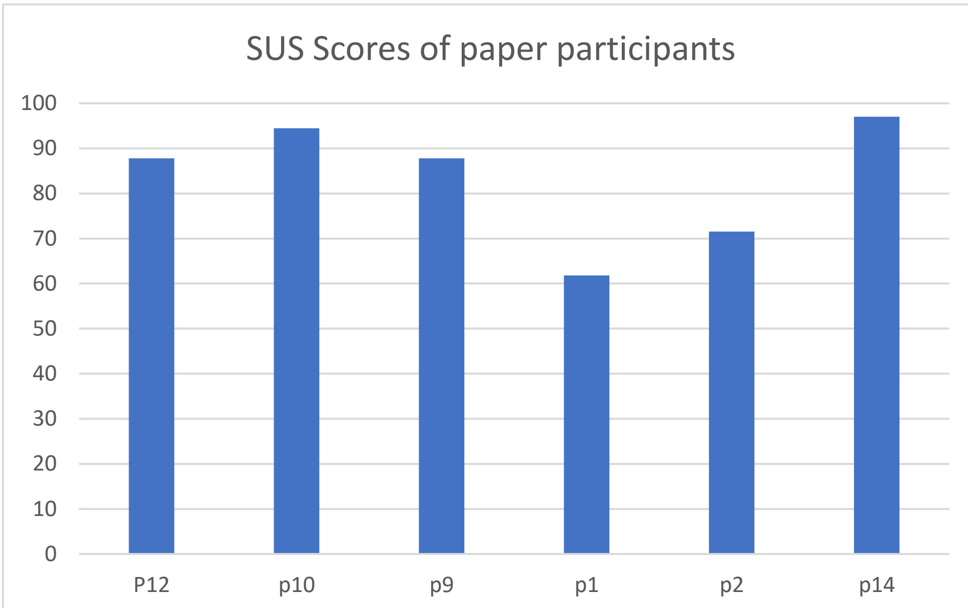


Figure 5.2 The SUS scores of paper participants

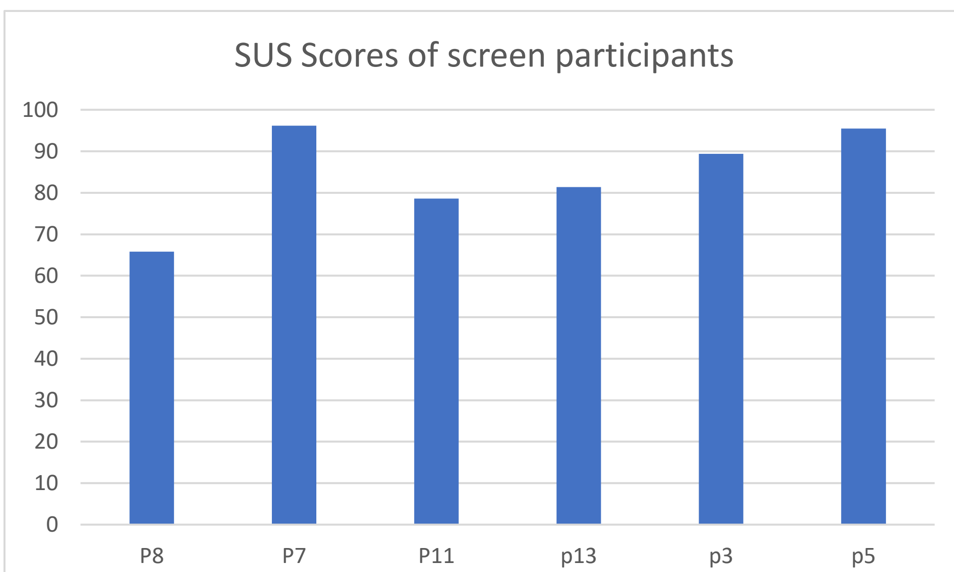


Figure 5.3 The SUS scores of screen participants

The survey showed that the application we proposed has a “Good” (nearly “Excellent”) usability rating for both fidelities, and there is no statistically significant difference between groups for either the overall score or any individual question. The most noticeable differences between scores (>0.35) are in questions 1, 4, 5, and 10 (see Figure 5.1). The following are the statements of the questions:

- Q1. “I think that I would like to use this system frequently” (screen participants gave scores indicating better usability).
- Q4. “I think that I would need the support of a technical person to be able to use this system” (screen).
- Q5. “I found the various functions in this system were well integrated” (paper).
- Q10. “I needed to learn a lot of things before I could get going with this system” (paper).

As mentioned, higher ratings of odd questions mean the app is more usable while it is the opposite for even questions. In terms of these larger differences between the groups, we hypothesize that:

- Screen participants gave scores indicating better usability for Q1 because they are familiar with using a high-fidelity implementation (most commercial apps on their own phones).
- Screen participants gave scores indicating better usability for Q4 because they were able to interact with the prototype largely independently, while for the paper participants the researcher (who could be considered a “technical person”) physically interacted at each step of the tasks (by moving the pieces of paper around after each button push).
- We are not sure why paper participants gave scores indicating better usability than screen participants for Q5.
- Paper participants gave scores indicating better usability for Q10 because they felt more supported already by the greater interaction with the researcher as described in our hypothesis for Q4.

However, we emphasize again that none of these differences are statistically significant and the general conclusion we draw from the quantitative data is that the influence of fidelity on quantitative usability rating is minimal.

5.3 Qualitative

To analyze the qualitative data from the semi-structured interviews, we used the conventional content analysis method [37]. In this method, the study starts with observation and recording of open-ended, free-text user feedback, and codes are derived and defined from the textual data. This method is used when no predetermined categories are set; instead, codes are discovered and allowed to emerge from collected data [38].

Our analysis was conducted in two rounds. In the first round, we started by formulating an interpretation of responses from the audio recordings and the notes taken during the interview of each participant. Then, we extracted comments from responses where each comment is one sentence that is concerned with one aspect of the prototype. Each comment was written down on a single sticky note page. Finally, we organized these notes into an affinity diagram and grouped them to form subcategories and categories from the bottom upwards. We conducted this analysis on each fidelity group separately and formed two affinity diagrams (See Appendix D for a photo of the original sticky notes and affinity diagrams).

In the second round of analysis, we formulated codes from the comments, and then modified the subcategories and the categories accordingly. Figures 5.4 and 5.5 show the final diagrams, in which each diagram contains four trees anchored at the category level, and each tree consists of three levels: a single category and then subcategories and codes. The categories that emerged from participant responses after this second round of analysis are the same in both diagrams. This feature of the diagrams is not surprising given that the two prototypes have the same purpose and structure, with the only difference being fidelity.

We carried out both rounds of analysis separately on the data from the two groups, allowing subcategories and categories to emerge from the data and to reflect the responses of participants; critically, categories were not predetermined for either group but a parallel set emerged from the data. The responses of participants fell into the following categories: User Experience, UI Elements, Content, and Suggestions. In the next subsections, we discuss our analysis for each group for each category, and then we discuss the similarities and the differences of responses.

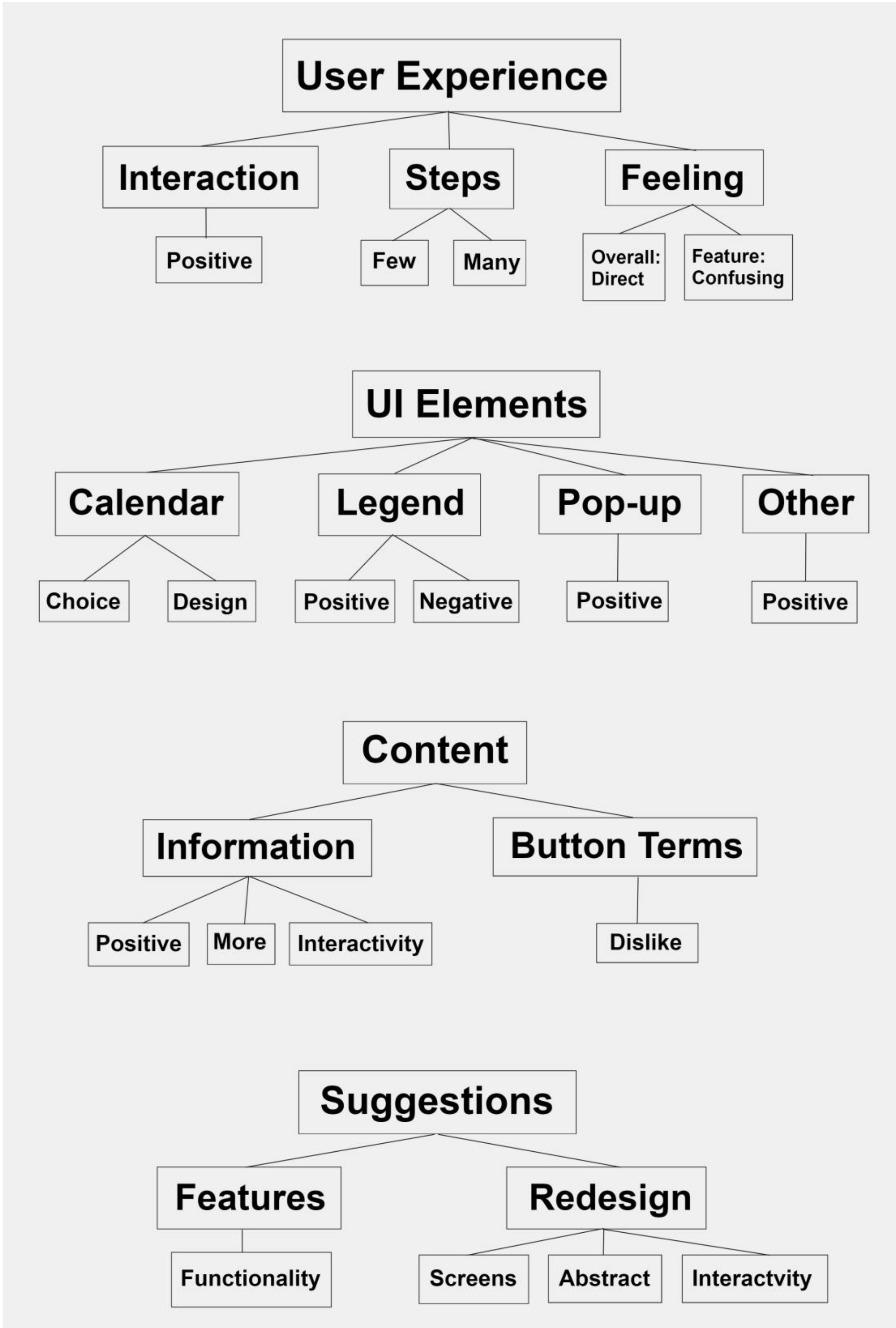


Figure 5.4 Affinity diagram of qualitative responses from "paper" participants

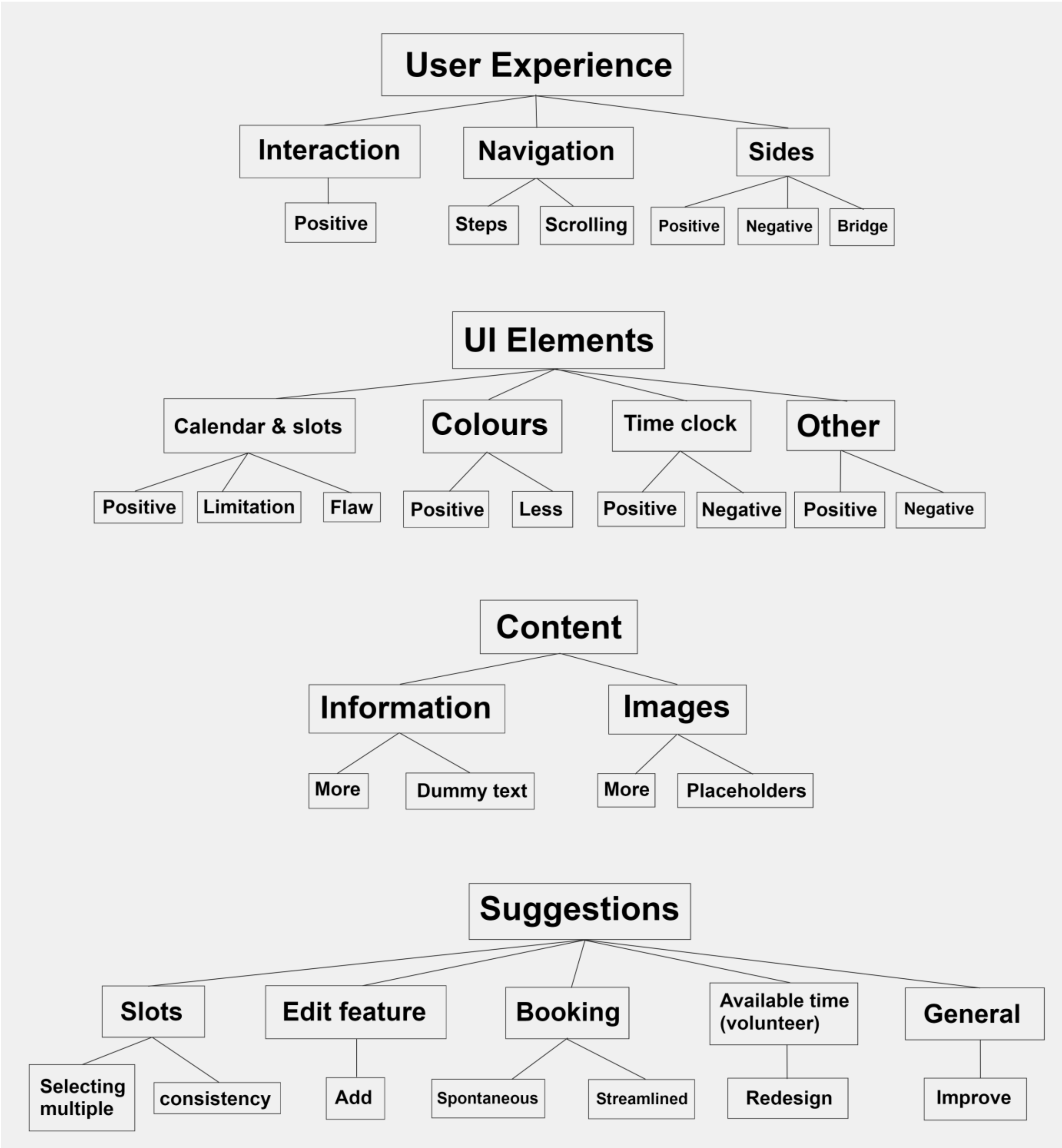
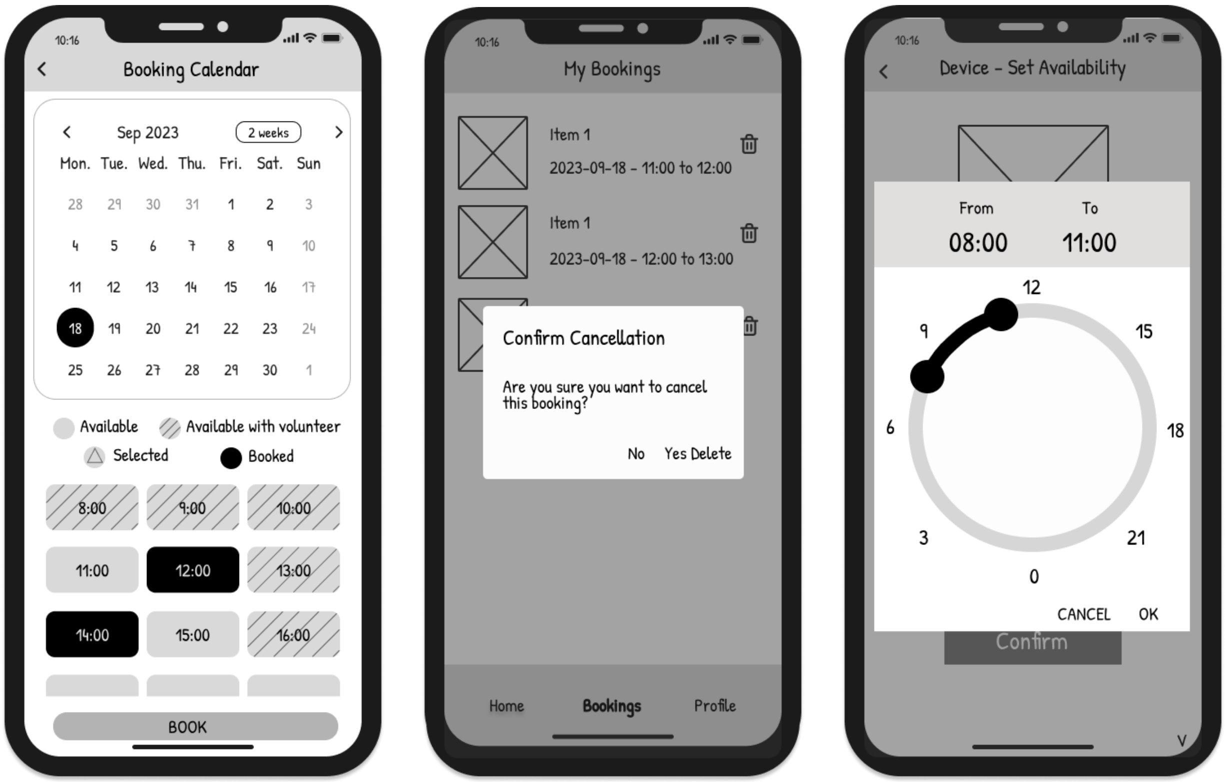


Figure 5.5 Affinity diagram of qualitative responses from "screen" participants

5.3.1 Low-fidelity (paper) results

User Experience. Most of the comments on user experience were concerned with the interaction with the prototype and with the number of steps taken to achieve the given tasks. However, there were comments related to the overall feeling of the booking process, which is described as direct; *"I disliked the direct feeling of booking" P14.* Another comment was about a feature in the volunteer interface (a time clock to set available time, Figure 5.6:b) which is described as "confusing". The comments on the time clock element varied in the pool of screen participants, and the mention of the "overall feeling" is noted in this pool only

UI Elements. The comments on the elements of the user interface were concerned with the calendar and the legend of the booking interface (Figure 5.6:a), and with a pop-up message that confirms the booking or the deletion of a device (Figure 5.6:b). The latter is found useful and "safe". Choosing a calendar for booking is found "suitable" and the comments on the legend varied from one participant to another. As shown in Figure 5.6:a, the legend design for the high-fidelity prototype is based on colour, so for the greyscale paper prototype it was replaced with different shades and shapes. Some participants described this shading approach as "clear" while others mentioned that it is confusing and not visible.



(a) **(b)** **(c)**

Figure 5.6 paper prototype, (a) device calendar page (b) pop-up message to confirm cancellation of a booking, (c) time clock to set available time of volunteer

Content. The comments on content were concerned with the dummy information provided for AMDs and with the terms used for buttons (see Figure 3.5). Participants disliked the terms "Book Now" and "Device Calendar" on two of the buttons. As for the text information that describes AMDs, participants found it "structured" but suggested that there should be more information about the AMDs and that this information should be more interactive, as the current information is static text data. Two participants stated that the low-fidelity prototype is "confusing" in general because of the lack of content, pictures, and colours.

Suggestions. Suggestions from this group were concerned with adding more features and with redesigning certain aspects of the app. The new features suggested were concerned with functionality, such as adding an "edit booked slot" feature, a filter to the home page, a search bar, and the ability to book multiple slots at once. The comments related to redesign are more abstract and suggest new ideas to be added to the user experience:

- "*The user should have the option to ask for a certain volunteer*" *P14.*
- *"The 'Bookings' screen should have two sections, one for passed bookings and for future ones" P9.*
- *"Setting available date and time of volunteers could be done without a calendar" P1.*
- *"There should be more distinction between clickable and non-clickable areas" P10.*

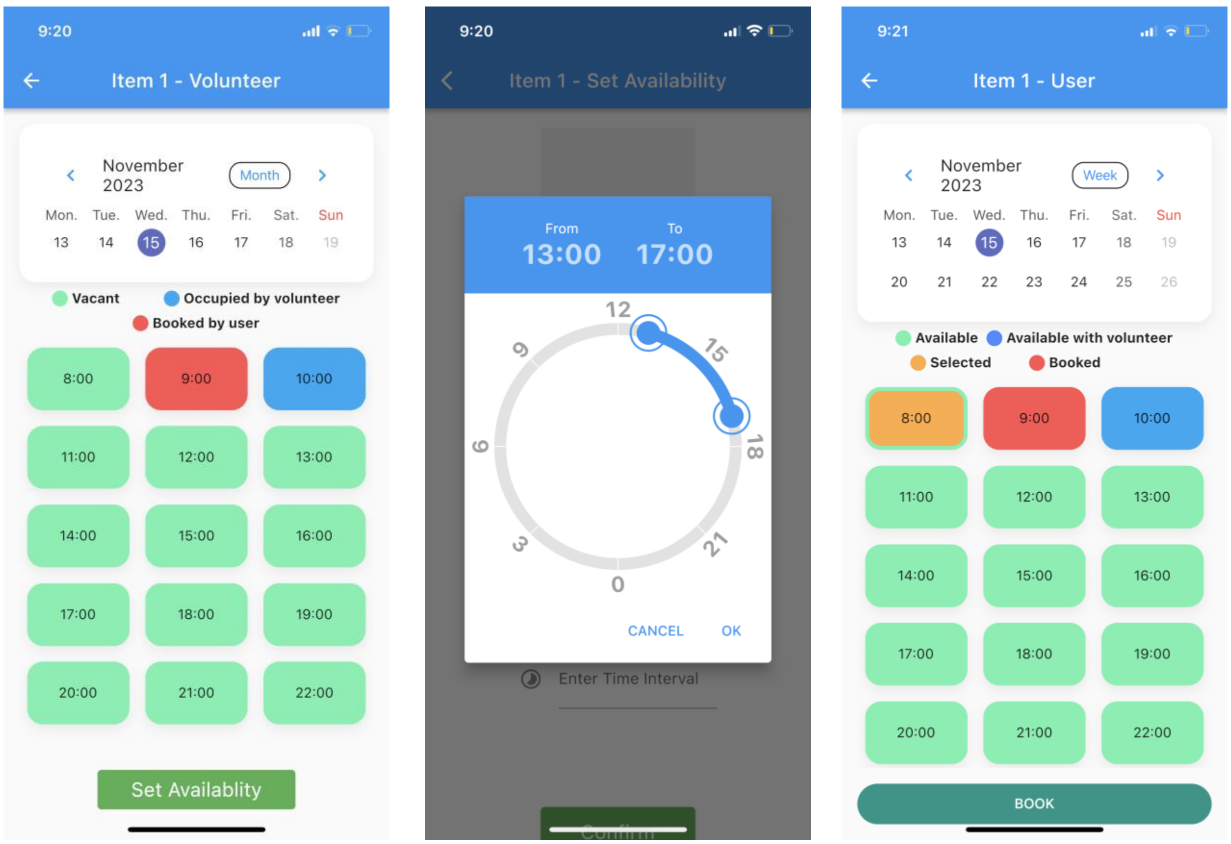
5.3.2 High-fidelity (screen) results

User Experience. The comments on user experience showed that participants found the interaction with the app is positive in general. The comments on navigation were not concerned with the steps of tasks only (as with the paper group) but with scrolling as well. The most notable from this group is the mention of sides and how *"the app would be a good bridge between users and volunteers" P5*. Participants of this group carried out sessions using two mobile devices where each device represented one side (user and volunteer). The two devices may have acted as an implicit reminder for participants that this is a multi-sided booking app, whereas the paper prototype used a different stack of papers for each task regardless of on which side that task fell.

UI Elements. The comments on UI elements varied between positive and negative and were concerned with several design components. In contrast to the paper prototype, the interactive nature of the calendar is present and the comments on colour are at the forefront (see Figure 5.7 for the corresponding screenshots):

- Two participants liked the colours.
- One participant described the colour coding of the legend and the slots as "simple".
- One participant mentioned that there are too many colours on the booking screen and that there should be a maximum of three.
- One participant mentioned that the colour coding of the legend and the slots is not obvious and that they should be limited to two.

Participants also noted a usability flaw in the calendar: the change view button has an in-line text that is inconsistent with the current calendar view (Figures 5.7:a and 5.7:c). Booking one slot at a time was mentioned as a limitation of the current design. Lastly, the comments on the time clock element (Figure 5.7:b) differed; some liked the element while others found it confusing, but one participant mentioned that it is better when holding the phone.



(a) **(b)** **(c)**

Figure 5.7 screen prototype, (a) the calendar interface from the volunteer side, (b) time clock to set available time of volunteer, (c)) the calendar interface from the user side

Content. The comments on content were concerned with the dummy text data and the images of devices (see Figure 3.3:a). Participants noticed the "lorem ipsum" text and the use of image placeholders instead of images of devices. Participants noted that there should be more information on the "Bookings" screen where the currently booked slots are displayed (Figure 3.3:c). Interestingly, participants from this group did not make any comments about the terms used for buttons.

Suggestions. Suggestions received from this group are varied and scattered across several aspects. One suggestion is to make selecting slots not restricted to selecting just one at a time (Figure 5.7:c), another is to add an "edit booked slot" feature to the "Bookings" screen (Figure 3.3:c). One participant noticed the inconsistency in selecting slots between the two interfaces, in which the participant in the user interface can click and select a slot, while in the volunteer interface they cannot (Figures 5.7:a and 5.7:c). The calendar of the volunteer interface (Figure 3.4:a) displays slots only, and the available time of volunteers is set through a date and time form (Figure 3.4:b). There were many comments about the process of setting available times on the volunteer side, and all the participants suggested redesigning this feature partially or fully. Participants suggested making the volunteer availability process the same as the user booking process, through a calendar and time slots. One participant suggested the volunteer process be "more customizable", and another participant suggested adding a monthly time frame of availabilities. There were comments on the booking process to make it "spontaneous" and "streamlined". The rest of the comments were general, such as suggesting adding more features, adding bulk bookings and cancellations, and improving the UI and the UX. One participant mentioned that the experience of the app will *"depend on how fast the app is going to be" P5.*

5.4 Discussion

Colour was mentioned significantly in the screen group as opposed to the paper one, especially since it is used to convey information in our design. Low-fidelity prototypes are usually paper prototypes, but the option of using colour is available in this medium. However, as low-fidelity prototypes are chosen for their rapid development, colour is often omitted in practice. In addition, prototypes are developed to emulate the final product (an app), hence the feedback on colour should be taken from the screen, not from paper. In our study, participants saw the greyscale of the paper prototype as a negative aspect and suggested that it should be more appealing; *"the low-*

fidelity prototype has no picture and no colour" P9. On the other hand, screen participants suggested reducing the number of colours used, especially for the calendar and slots screen; *"the number of colours used on the screen should be less or maximum of 3" P8.*

The comments on content occurred in both groups. The comments on text were more in the low-fidelity prototype, while comments on images were more in the high-fidelity prototype. We noticed that participants might be comparing the content of the prototype app (both versions) with the content of a market-ready app. When developing our app, the focus was on the interaction process and minimal attention was given to content beyond the structure through which the interaction process is accomplished. Participants suggested adding more content and that content should be more interactive and meaningful. As noted in the literature, pictures stimulate better suggestions of ideas [9]; consequently, the lack of realistic images is a downside of our prototypes and content should be taken into consideration when designing a prototype.

The effect of holding the phone while using the app appeared only once in the high-fidelity prototype group. However, this feedback is concerned with a late-in-development issue and is only expected to be addressed in a high-fidelity prototype.

As for sketching, only 5 participants used the pen and paper provided to them during the interviews, of which 3 were from the screen pool and 2 from the paper pool. Three of these sketches were scribbles, while the other two, both from a paper participant, were more informative. One sketch was a complete drawing of a mobile interface with a suggestion for a new design for the home page. This participant was a product designer. From these results we conclude that sketching design was not fully utilized in our pilot study but we have anecdotal evidence that paper participants are more likely to provide a detailed sketch than screen ones, and that the background of the participant has a significant effect on whether that participant is willing to go beyond verbal feedback during usability testing.

Suggestions between the two groups showed the most noticeable differences in feedback. We found the suggestions from the low-fidelity group to be creative and to provide new room for redesigning the user experience. On the other hand, we found the suggestions from the high-fidelity prototype to be more varied between participants and across features of the prototype. They were suggestions to enhance the current design and user experience. To be concrete about the relationship observed between fidelity and feedback, we outline the specific areas of focus in feedback from each fidelity in the lists below. Based on our pilot experiment, the following are the larger differences in feedback between groups:

Paper prototype:

- Comments on the terminology used for buttons.
- Comments on the overall feeling resulted from the interaction.
- Suggestions to redesign pages.
- Suggestions for new ideas to change and expand the current design.

Screen prototype:

- Comments on scrolling.
- Comments on design flaws related to interactivity.
- Comments on colours.
- Comments on images.
- Comments on the effect of holding the phone when interacting with a certain feature.
- Comments on the relation between different user role types.
- Suggestions for enhancing the user experience in general.

Finally, here are suggestions from the qualitative feedback for improvements to enhance the app. These suggestions were selected from the responses of participants based on their importance and applicability. We list them from those we feel are the most to the least valuable:

- *Booking multiple slots at once.* Participants suggested that the user should be able to book multiple time slots at once (Figure 5.7:c). This suggestion is important because the user is likely to book the device for more than an hour (one slot) consecutively in practice.
- *Redesigning the way volunteers set availabilities.* Participants suggested making the booking side of the user and the setting availabilities of the volunteer (Figure 3.4) identical.

However, this is not feasible in the current design because we wanted the volunteer to provide a range of hours for their availability. This could be done if the previous suggestion of booking multiple slots at once is implemented. This change might also eliminate the time clock element (Figure 5.7:b) which some participants did not like.

- *Improve the "Bookings" screen.* Participants suggested making two sections in this screen (Figure 3.3:c), one for upcoming bookings and one for previous bookings. This could be applied by having a tab bar at the top. Additionally, participants suggested adding a way to edit the currently booked slot and adding more information about the booked slot.
- *Legend, slots, and colours.* Participants suggested making the legend of the slots to be more visible and use fewer colours. This change could be done by increasing the size of the legend and making both calendar screens (Figures 5.7:a and c) identical; the legend could contain just three elements with one colour for each: vacant, volunteer, and booked.
- *Content and usability flaws.* Participants suggested:
    - Adding more interactive content about AMDs.
    - Fixing the usability flaw of the calendar change view button (Figure 5.7:a and c).
    - Using better text on buttons.

5.5 Limitations

One noticeable limitation of this study is that we are using only one design to solicit feedback and ideas from participants. Presenting one design in usability testing makes users reluctant to provide criticism when compared to cases in which multiple design alternatives are presented [26]. Furthermore, in order to ensure that the test sessions were not too long only the booking and availability features are tested, and those are tested according to a detailed script. This approach limits the number of tasks that users can perform and their freedom in how those tasks can be performed, and hence the range of feedback likely to be received. The interviews were limited to five main questions (Appendix B.3) for the same reason.

A usability flaw of which we are aware is that we have used colour alone in our design to convey critical meaning: slots in the calendar have different colours representing the slot’s state. Accessibility design guidelines require adding text beside colour when colour is used to convey information [39]This shortcoming could be solved by adding text inside each slot in accessibility mode. In our study, the comments on colour were about how many colours are used rather than about using colour to convey information.

As for the prototype, the lack of pictures and meaningful content turned out to be a downside. We had incorrectly believed as we designed the prototype that having abstracted content would make participants focus on the interaction, not on content. It turned out that the content was critical because in practice participants consistently compared the prototype to a market-ready app in their feedback.

We used a population of participants sampled for convenience, so it does not represent the target audience of users and volunteers; furthermore, the number of participants is too small to identify statistically significant differences in the quantitative feedback between the treatment groups. These limitations were known in advance, so the focus of the study was always on qualitative data which might inform future study designs. That said, the analysis of the qualitative data was performed by only one researcher due to limitations of time (that researcher needs to graduate); therefore, the lack of triangulation potentially limits the repeatability of the results.

Chapter 6: Conclusion

This study aimed to determine whether and how the fidelity of a mobile app prototype would affect the feedback received from testers in usability testing. We developed a high-fidelity fully functional mobile app with a corresponding low-fidelity paper prototype to conduct an experiment utilizing quantitative and qualitative data collection methods on two independent groups which each used one of the prototypes.

The analysis of quantitative data showed that the level of fidelity did not affect usability rating from testers, as the scores from the usability survey between the two groups of testers (low-fidelity vs. high-fidelity) were proximate (83.4 vs. 84.5 respectively on a scale from 0 to 100).

The analysis of qualitative data showed differences in the nature of the feedback received from the two groups in line with reports from the literature. The low-fidelity group commented on the terminology used for buttons and the overall feeling of the interaction, and gave suggestions for adding new features or redesigning pages. The high-fidelity group commented on colours and images, scrolling and other design flaws with the interactivity, and their suggestions generally requested enhancements to the user experience rather than entirely new features or major redesigns.

The lack of meaningful content (images and text) in the prototype, the population of participants, and the lack of independent triangulation in the qualitative analysis were the main limitations of this study. In addition to fixing these weaknesses, a future study could likely elicit richer feedback using a multi-design multi-feature approach to test several prototypes with a different design for each.

References

[1] Y.-K. Lim and E. Stolterman, “The anatomy of prototypes: prototypes as filters, prototypes as manifestations of design ideas,” *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 15, no. 2, pp. 1–27, 2008.

[2] J. Rudd, K. Stern, and S. Isensee, “Low vs. high-fidelity prototyping debate,” *Interactions*, vol. 3, no. 1, pp. 76–85, 1996.

[3] M. Mccurdy, C. Connors, G. Pyrzak, B. Kanefsky, and A. Vera, “Breaking the fidelity barrier: an examination of our current characterization of prototypes and an example of a mixed-fidelity success,” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.*, pp. 1233–1242, 2006.

[4] E. Tiong *et al.*, “The economies and dimensionality of prototyping: value, time, cost and fidelity,” *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 51845, 2018.

[5] M. Walker, L. Takayama, and J. A. Landay, “High-fidelity or low-fidelity, paper or computer? Choosing attributes when testing web prototypes,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.*, vol. 46, no. 5, pp. 661–665, 2002.

[6] M. Rettig, “Prototyping for tiny fingers,” *Commun ACM*, vol. 37, no. 4, pp. 21–27, 1994.

[7] A. Harley, “Functional Fixedness Stops You From Having Innovative Ideas,” Nielsen Norman Group. Accessed: Apr. 29, 2023. [Online]. Available: https://www.nngroup.com/articles/functional-fixedness/

[8] B. Köhler, J. Haladjian, B. Simeonova, and D. Ismailović, “Feedback in low vs. high fidelity visuals for game prototypes,” *2012 Second International Workshop on Games and Software Engineering: Realizing User Engagement with Game Engineering Techniques (GAS)*, pp. 42–47, 2012.

[9] Y. Borgianni, L. Maccioni, L. Fiorineschi, and F. Rotini, “Forms of stimuli and their effects on idea generation in terms of creativity metrics and non-obviousness,” *International Journal of Design Creativity and Innovation*, vol. 8, no. 3, pp. 147–164, Jul. 2020.

[10] Y.-K. Lim, A. Pangam, S. Periyasami, and S. Aneja, “Comparative analysis of high-and low-fidelity prototypes for more valid usability evaluations of mobile devices,” *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles*, pp. 291–300, 2006.

[11] D. Rueda, R. Hoto, and A. Conejero, “Study of the influence of prototype aesthetic fidelity (a realism factor) in usability tests,” *Human Factors in Computing and Informatics: First International Conference, SouthCHI 2013*, pp. 122–136, 2013.

[12] G. Sim and B. Cassidy, “Investigating the fidelity effect when evaluating game prototypes with children,” *27th International BCS Human Computer Interaction Conference*, vol. 27, pp. 1–6, 2013.

[13] J. Sauer and A. Sonderegger, “The influence of prototype fidelity and aesthetics of design in usability tests: Effects on user behaviour, subjective evaluation and emotion,” *Appl Ergon*, vol. 40, no. 4, pp. 670–677, 2009.

[14] J. Sauer, K. Seibel, and B. Rüttinger, “The influence of user expertise and prototype fidelity in usability tests,” *Appl Ergon*, vol. 41, no. 1, pp. 130–140, 2010.

[15] A. Uebelbacher, A. Sonderegger, and J. Sauer, “Effects of perceived prototype fidelity in usability testing under different conditions of observer presence,” *Interact Comput*, vol. 25, no. 1, pp. 91–101, 2013.

[16] B. Kang, N. Crilly, W. Ning, and P. O. Kristensson, “Prototyping to elicit user requirements for product development: Using head-mounted augmented reality when designing interactive devices,” *Des Stud*, vol. 84, 2023.

[17] R. A. Virzi, J. L. Sokolov, and D. Karis, “Usability problem identification using both low-

and high-fidelity prototypes," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1996, pp. 236–243.

[18] L. Christoforakos and S. Diefenbach, "Idealization effects in UX evaluation at early concept stages: Challenges of low-fidelity prototyping," *Advances in Usability, User Experience and Assistive Technology: Proceedings of the AHFE 2018 International Conferences on Usability & User Experience and Human Factors and Assistive Technology*, pp. 3–14, 2019.

[19] C. Boothe, L. Strawderman, and E. Hosea, "The effects of prototype medium on usability testing," *Appl Ergon*, vol. 44, no. 6, pp. 1033–1038, 2013.

[20] T. Zhang, P.-L. Patrick Rau, G. Salvendy, and J. Zhou, "Comparing low and high-fidelity prototypes in mobile phone evaluation," *International Journal of Technology Diffusion (IJTD)*, vol. 3, no. 4, pp. 1–19, 2012.

[21] M. Deininger, S. R. Daly, J. C. Lee, C. M. Seifert, and K. H. Sienko, "Prototyping for context: exploring stakeholder feedback based on prototype type, stakeholder group and question type," *Res Eng Des*, vol. 30, no. 4, pp. 453–471, Oct. 2019.

[22] D. Pereira, "Multisided Platform Business Model," The Business Model Analyst. Accessed: May 01, 2023. [Online]. Available: https://businessmodelanalyst.com/multisided-platform-business-model/

[23] S. Silverthorne, "New Research Explores Multi-Sided Markets," HBS Working Knowledge. Accessed: May 01, 2023. [Online]. Available: https://hbswk.hbs.edu/item/new-research-explores-multi-sided-markets

[24] B. Tversky, "What do sketches say about thinking?," *2002 AAAI Spring Symposium, Sketch Understanding Workshop, Stanford University, AAAI Technical Report SS-02-08*, vol. 148, p. 151, 2002.

[25] M. Tohidi, W. Buxton, R. Baecker, and A. Sellen, "User sketches: a quick, inexpensive,

and effective way to elicit more reflective user feedback," *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles.*, pp. 105–114, 2006.

[26] M. Tohidi, W. Buxton, R. Baecker, and A. Sellen, "Getting the right design and the design right: Testing many is better than one," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.*, pp. 1243–1252, 2006.

[27] "Top 10 Mobile App Development Frameworks in 2023 | Technostacks." Accessed: Jul. 21, 2023. [Online]. Available: https://technostacks.com/blog/mobile-app-development-frameworks/

[28] K. Herembourg, "8 Most Popular Mobile App Development Frameworks in 2023." Accessed: Jul. 21, 2023. [Online]. Available: https://www.purchasely.com/blog/mobile-app-frameworks

[29] A. Bizzotto, "Flutter vs. React Native - Which One to Choose? - Udemy Blog." Accessed: Jul. 21, 2023. [Online]. Available: https://blog.udemy.com/flutter-vs-react-native/

[30] "Flutter vs. React Native - A detailed evaluation · Sean Connolly." Accessed: Jul. 21, 2023. [Online]. Available: https://seanconnolly.dev/flutter-vs-react-native

[31] D. H. Sánchez, "Flutter Provider: What is it, what is it for, and how to use it? | Bancolombia Tech | Medium." Accessed: Jul. 22, 2023. [Online]. Available: https://medium.com/bancolombia-tech/flutter-provider-what-is-it-what-is-it-for-and-how-to-use-it-47d6941860d7

[32] C. Imoh, "A quick guide to Provider for Flutter state management - LogRocket Blog." Accessed: Jul. 22, 2023. [Online]. Available: https://blog.logrocket.com/quick-guide-provider-flutter-state-management/

[33] A. Blackler, "Applications of high and low fidelity prototypes in researching intuitive interaction," *Proceedings of DRS2008, Design Research Society Biennial Conference*, pp. 1–17, 2008.

[34] H. Smith, G. Fitzpatrick, and Y. Rogers, "Eliciting reactive and reflective feedback for a social communication tool: a multi-session approach," *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, pp. 39–48, 2004.

[35] J. Brooke, "SUS: A quick and dirty usability scale," *Usability Evaluation in Industry*, vol. 189, no. 194, pp. 4–7, 1996.

[36] A. Bangor, P. Kortum, and J. Miller, "Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale," *J Usability Stud*, vol. 4, pp. 114–123, 2009.

[37] H. F. Hsieh and S. E. Shannon, "Three approaches to qualitative content analysis," *Qual Health Res*, vol. 15, no. 9, pp. 1277–1288, Nov. 2005.

[38] J. Lazar, Jinjuan Heidi Feng, and H. Hochheiser, "Analyzing qualitative data," in *Research Methods in Human-Computer Interaction*, 2017, pp. 299–325.

[39] "Measuring and Interpreting System Usability Scale (SUS) - UIUX Trend." Accessed: Sep. 18, 2023. [Online]. Available: https://uiuxtrend.com/measuring-system-usability-scale-sus/#interpretation

Appendices

Appendix A Study Documents

A.1 Consent form

THE UNIVERSITY OF BRITISH COLUMBIA

Department of Computer Science
2366 Main Mall
Vancouver, B.C., V6T 1Z4

Consent Form

Exploring the influence of prototyping format on feedback and participatory design for a multi-sided booking mobile app

Principal Investigator: Dr. Ian M. Mitchell, Professor, Department of Computer Science, University of British Columbia

Co-Investigator: Yaman Sanobar, M.Sc. Student, Department of Computer Science, University of British Columbia

This project is a graduate student MSc thesis project contributing to the graduate degree of Yaman Sanobar.

Introduction: Thank you for participating in this study. This work is affiliated with the Collaborative Robotics Laboratory at UBC and is funded by NSERC. With your help, we seek to understand how different formats of mobile prototypes affect the feedback received from usability testing participants. The results might enrich the literature with insights about this topic and will help us in designing better interaction for the proposed mobile app.

Purpose: The overall purpose of this study is to examine the difference in feedback received from testing prototypes that use similar interaction designs but are implemented in different mediums. Each participant will test one format across two interfaces. We will analyze and compare the feedback received from different testing groups to identify any qualitative or quantitative differences in responses.

What you will be asked to do: After you have read this document, the experimenter will respond to any questions or concerns that you may have. Once you have signed this consent form, you will be asked to:

- **Fill in a pre-interaction survey.** Prior to using the prototype, you will be asked to complete a short demographic survey.

RAD-Online-Pilot, 2023. Consent Form Version 3 – August 23, 2023. Ethics ID: H23-01482 Page **1** of **3**

- **Complete three predefined tasks.** You will follow instructions to test and use the prototype and carry out three tasks through different scenarios in the booking app. The experimenter will observe and take notes.
- **Fill in a post-interaction survey.** At the end of testing, you will be asked to complete a brief usability testing survey about your experience with the prototype.
- **Participate in an interview.** After the survey, you will have an interview with the experimenter. The interview will be about your interaction with the prototype and the experimenter will ask you for your suggestions on how this interaction could be improved. You will be provided with a pen and a piece of paper as sketching is encouraged during the interview. Sketches do not need to be perfect or complete. The experimenter will be listening to your feedback during the interview and will be taking notes accordingly.

Time commitment: The study should take between 45 and 60 minutes and will be completed in 1 session.

How the data collected will be used: Data collected will be used for analysis and will be used by the student investigator to form the basis of thesis research which might be submitted as a research presentation and/or publication.

Audio recordings: Your session will be audio recorded. Audio recordings will be transcribed locally using Whisper, but audio excerpts will not be used in any reports, presentations, or publications. After the experiment, transcripts will be used by the experimenter (in combination with written notes) to formulate a better understanding of the responses of each participant.

Confidentiality: Your confidentiality will be respected. Any information that could identify you as a participant in this study will be kept confidential. All information that you provide will be stored in Canada. Your identity will not be revealed in reporting the study results. Original survey results and audio transcripts might be shared, but any information that could identify you will be omitted.

Data retention: Identifiable data and audio recordings will be stored securely in a locked metal cabinet or on an encrypted and password protected computer storage device. All data from individual participants will be coded so that their anonymity will be protected in any reports, research papers, thesis documents, and presentations that result from this work. All data will be destroyed or deleted after 5 years.

Compensation: You will receive monetary compensation of $20 for this session.

Contact for information about the study: If you have any questions, concerns, or desire for further information about the study before or during participation, you may contact Yaman Sanobar at

Contact of information about the rights of research subjects: If you have any concerns or complaints about your rights as a research participant and/or your experiences while participating in this study, contact the Research Participant Complaint Line in the UBC Office of Research Ethics at 604-822-8598 or if long distance e-mail RSIL@ors.ubc.ca or call toll free 1-877-822-8598. Ethics ID number: H23-01482.

Devices will be cleaned before the session and masks will be provided at the test site if needed.

Indicate your agreement to being audio recorded for this study by providing your **initials**:

- I consent to being audio recorded for this study. _______

I, ________________________________, have read the explanation about this study. I have been given the opportunity to discuss it, and my questions have been answered to my satisfaction. I hereby consent to take part in this study. However, I realize that my participation is entirely voluntary and that I am free to withdraw at any time.

____

Participant's Signature   Date

A.2 Call for Participants form

THE UNIVERSITY OF BRITISH COLUMBIA

Department of Computer Science
2366 Main Mall
Vancouver, B.C., V6T 1Z4

Call for Participants

Exploring the influence of prototyping format on feedback and participatory design for a multi-sided booking mobile app

Principal Investigator: Dr. Ian M. Mitchell, Professor, Department of Computer Science, University of British Columbia

Co-Investigator: Yaman Sanobar, M.Sc. Student, Department of Computer Science, University of British Columbia

You are invited to participate in a research study involving testing a multi-sided mobile app for booking adaptive devices conducted by the Collaborative Robotics Lab in the Department of Computer Science at the University of British Columbia.

Who can participate?

You must be 19 years of age or older, and be comfortable using English to respond to a survey, use a smartphone device to interact with an app, and answer interview questions that may involve basic sketching.

What is Involved?

- Test a booking mobile app through a pre-defined set of tasks for two user interfaces.
- Rate the usability of the interaction through a survey.
- Provide feedback on the interaction with the mobile app prototype through an interview.
- Suggest new design ideas or concepts ideally through sketching.

What is time commitment? The study should take between 45 to 60 minutes and will be completed in 1 session.

Is there compensation? You will be compensated with $20.

Interested in participating?

Please contact Yaman Sanobar at

RAD-Online-Pilot, 2023. Call for Participants Version 2 – August 10, 2023 Page **1** of **1**

Appendix B Study Session Documents

B.1 Tasks list

Tasks list

Guidelines

- The experimenter will tell you when to start performing each task.
- Each task should take no more than five minutes to complete.
- If five minutes have passed, a "time-out" will occur, and the task will be considered incomplete.
- If you cannot or do not wish to complete any task, you can tell the experimenter.
- Please do not ask questions of the experimenter. If you are not sure what to do next, take your best guess. We are trying to replicate the experience of using the app for the first time by yourself.
- The experimenter may intervene and return to a previous screen if you enter an unintended or unimplemented component of the app.
- There is no wrong action. We are testing the interface, not you!
- Tasks are found on the next pages. Steps should be done sequentially. When you complete a step, please put a mark inside the box located to the left of the step number.
- The app we are testing consists of two interfaces: one for the user who will book an outdoor wheelchair device, and the other for a volunteer who will provide their availability for the user.
- To help us in the audio transcript, please read the task name and the scenario aloud before you start doing the steps for that task.

If you have any questions, please tell the experimenter before starting.

RAD-Online-Pilot, 2023. Tasks List Version 2 – August 23, 2023 Page **1** of **3**

Tasks	Scenarios	Steps
Task 1: Book a device without the need for a volunteer	A user wants to book a device but doesn't require the help of a volunteer	☐ 1. As a user, go to the Home page and select "Item 1". ☐ 2. Go to the booking calendar of that device, then go to September 18th. Select the "available" time slot at "11:00", and book it. ☐ 3. Check that your booking is placed by finding it listed in the "My Bookings" screen.
Task 2: Set the availability of a volunteer	A volunteer wants to set their availability for a certain device	☐ 1. As a volunteer, go to the Home page and select "Item 1". ☐ 2. Go to the device calendar, then go to the 18th of September. Locate and identify the 3 consecutive "Vacant" time slots from 8:00 to 10:00. Please confirm by saying out loud. ☐ 3. Click on "Set Availability" button. Set the date on the 18th of September. Set the time interval between 8:00 to 11:00. Click "Confirm". ☐ 4. Check that your "availability" is confirmed by finding it listed in the "My Availability" screen.

Task 3: Book a device with the need for a volunteer	A user changed their mind about a previous booking and wants to book a device with the help of a volunteer	☐ 1. As a user, go to the "Bookings" page and delete the previously booked slot at 11:00. ☐ 2. Go to the "Home" page and select "Item 1". ☐ 3. Go to the booking calendar of that device, then go to the 18th of September, select an "available with volunteer" time slot at 9:00, and book it. ☐ 4. Check that your booking is placed by finding it in "My Bookings" screen.

B.2 Survey

Intro

Fields to be filled by the experimenter

Participant numeric code

Prototype tested

◯ Paper
◯ Screen

Demographics

Do you have previous experience in app design?

◯ Yes
◯ No

Have you used a multi-sided booking app before? a booking app where multiple types of users are involved? (Uber is a good example)

- ◯ Yes
- ◯ Maybe
- ◯ No

Please elaborate on your experience in app design

Please list any multi-sided booking app you have used. If possible, please tell us what did you like or dislike about them.

As a user, what are the most important factors for you when using a booking app?

Pause

Your answers have been recorded, Thank you.

Now we will start testing the prototype.

SUS

Thank you for participating in the testing session.

We want to evaluate the usability of the prototype through the following usability survey.

Please provide thoughtful and objective answers as much as possible. Thank you

Questions

I think that I would like to use this system frequently

	Strongly Disagree	Neither Disagree or Agree	Strongly Agree

0 100

select

I found the system unnecessarily complex

	Strongly Disagree	Neither Disagree or Agree	Strongly Agree

0 100

select

I thought the system was easy to use

	Strongly Disagree	Neither Disagree or Agree	Strongly Agree

0 100

select

I think that I would need the support of a technical person to be able to use this system

Strongly Disagree | Neither Disagree or Agree | Strongly Agree

0 — 100

select

I found the various functions in this system were well integrated

Strongly Disagree | Neither Disagree or Agree | Strongly Agree

0 — 100

select

I thought there was too much inconsistency in this system

Strongly Disagree | Neither Disagree or Agree | Strongly Agree

0 — 100

select

I would imagine that most people would learn to use this system very quickly

Strongly Disagree	Neither Disagree or Agree	Strongly Agree

0 100

select

I found the system very cumbersome to use

Strongly Disagree	Neither Disagree or Agree	Strongly Agree

0 100

select

I felt very confident using the system

Strongly Disagree	Neither Disagree or Agree	Strongly Agree

0 100

select

I needed to learn a lot of things before I could get going with this system

Strongly Disagree	Neither Disagree or Agree	Strongly Agree

0 100

select

Powered by Qualtrics

B.3 Interview Guide

Interview Guide

Hi! I would like first to thank you for your participation in this study. It is greatly appreciated. In this interview, you will be asked about your interaction with the prototype. We want to improve the multi-sided booking process and we want the participation and involvement of users in doing so. You can find a pen and a piece of paper here in front of you. Please use them to help illustrate any comments, design suggestions, or ideas you might have. Sketches don't need to be perfect or complete.

The interview will take up to 20 minutes. I will remind you that your participation in this interview is voluntary. You can stop the interview at any time if you no longer want to participate. You do not need to answer any questions that you do not wish to. During the interview I will be taking notes about your opinions and suggestions, but there are no right or wrong answers.

As the consent form indicated, this interview will be audio recorded. I will let you know when I will start recording. This recording will be transcribed to text prior to analysis. Identifying and confidential information will not be used when presenting findings. For example, your voice recording will not be replayed publicly, and if your name or other personally identifying information arises during in the interview it will be anonymized or redacted.

Do you have any questions for me before we start?

I am now going to start recording.

RAD-Online-Pilot, 2023. Interview Guide Version 1 – July 09, 2023 Page **1** of **2**

1. What do you think of the booking interaction proposed in the prototype?
    a. Prompt: What about the model of a calendar to select a day, and then time slots within the day?
    b. Prompt: In what ways did the prototype meet or not meet your expectations about a booking app?

2. Are there any aspects of the booking process that you particularly liked or disliked?
    a. Prompt: calendar? Slots convey different information?

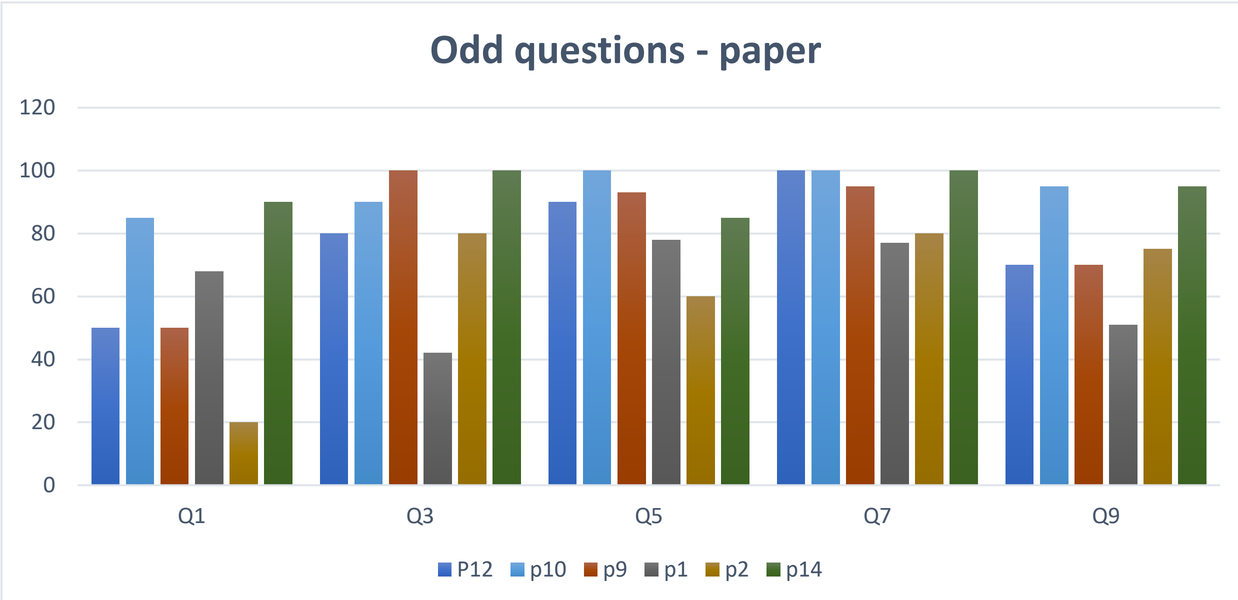
3. Are there any primary features that you think are missing in the current booking process? In other words, capabilities which would improve the booking process but are currently missing?
    a. Prompt: Editing a booked slot? Selecting multiple slots?

4. What improvements do you suggest to make the process more efficient or enjoyable? Feel free to sketch.
    a. Prompt: Perhaps redesign aspects of the current booking process?
    b. Prompt: Perhaps add or delete elements of the booking process?

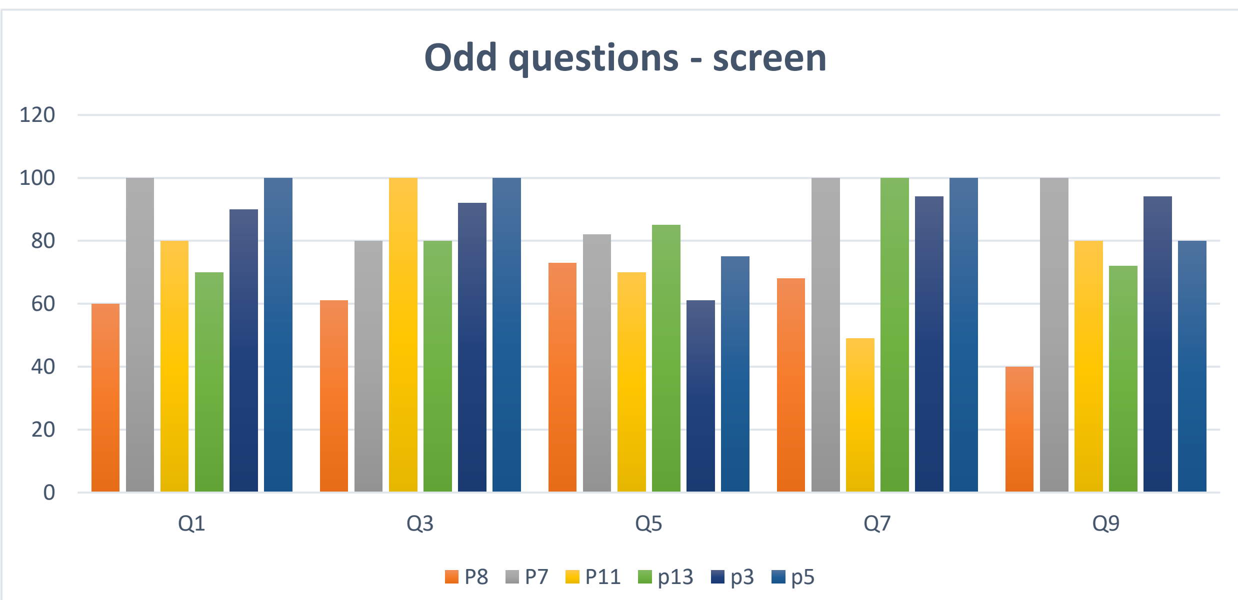
5. What if you had an entirely blank slate? Is there a better way to design the whole booking process? Feel free to sketch.
    a. Prompt: Are there other apps that perform a similar, booking-like task which use a novel interface or process?

Appendix C Quantitative Data

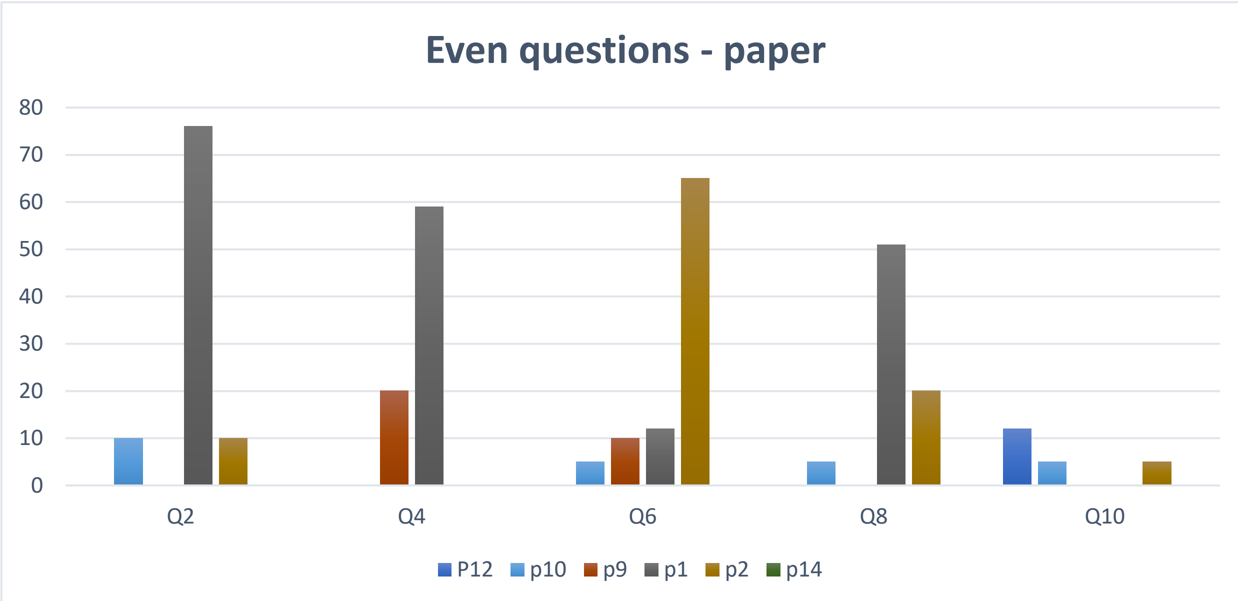
C.1 Paper participants' responses to SUS survey odd questions



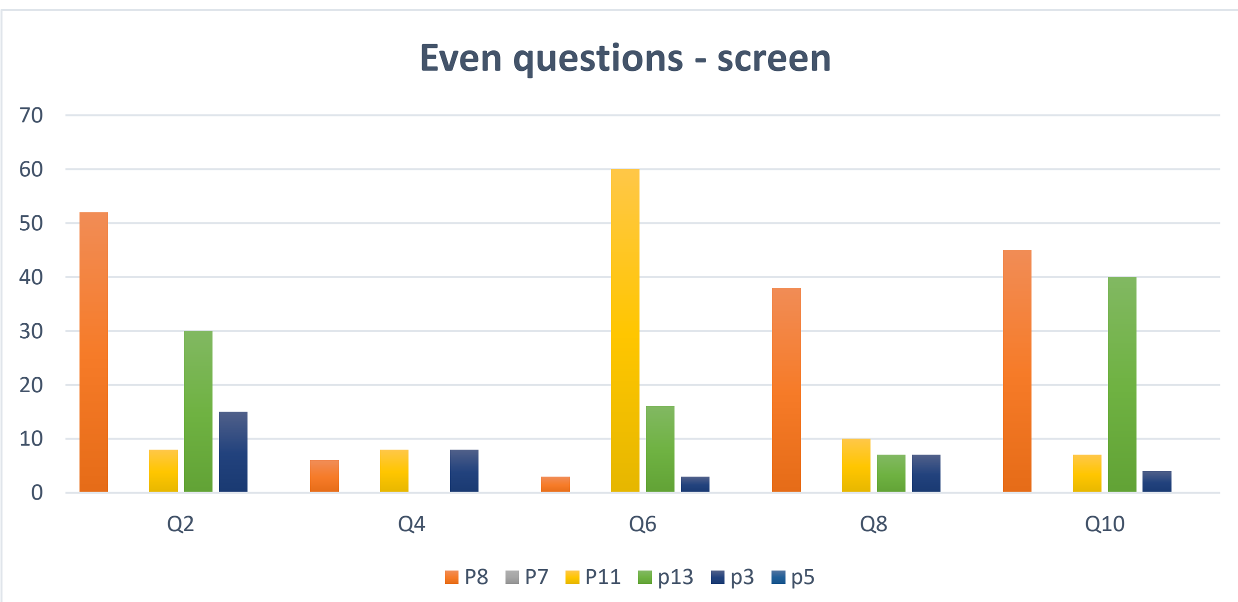
C.2 Screen participants' responses to SUS survey odd questions



C.3 Paper participants' responses to SUS survey even questions

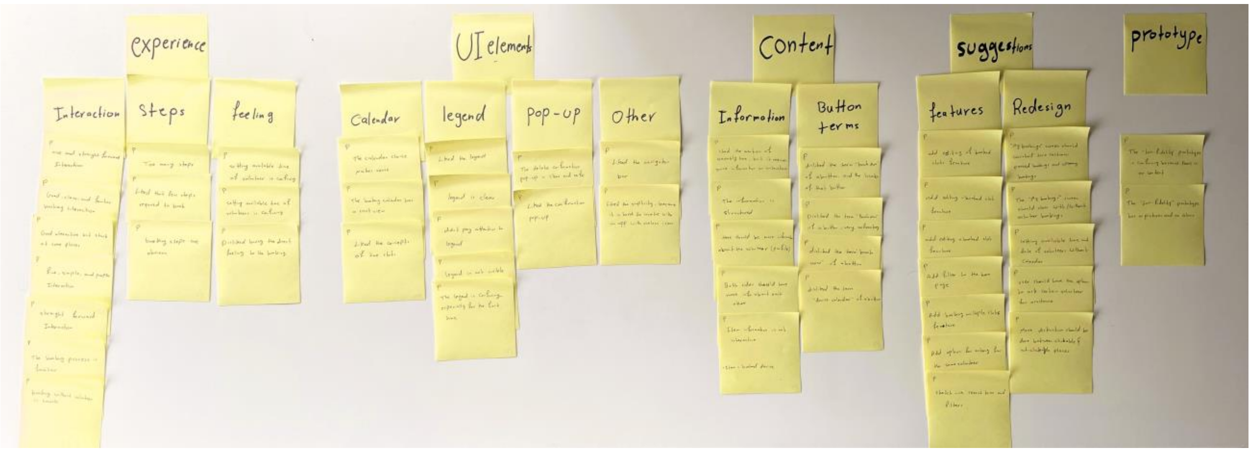


C.4 Screen participants' responses to SUS survey even questions



Appendix D Qualitative Data

D.1 Paper responses affinity diagram



D.2 Screen responses affinity diagram