

Final Evaluation with Users

Study Description

8 users (n=8) from both the SENG310 laboratory sections and researcher's personal relationships were asked to perform six tasks on the medium fidelity mockup of our Financial Management application prototype. The InVision prototype was sent to the participants and interviews were conducted through Zoom, a remote communication application. If participants were unable to use InVision, interviews were conducted through Zoom's remote access feature; whereby participants took control of the researcher's computer and navigated the prototype remotely. Experiments were conducted in two parts: a pilot study with 2 users, followed by a formal study with 6 users. Both studies utilized a Co-Discovery and Conceptual Model Extraction usability test protocol, asking questions pertaining to the application's adherence to the principles of task-centered system design. Participants were interviewed individually.

Pilot studies were conducted asynchronously using the protocols defined above. Two issues arose following these studies, including the participants having to install InVision studio and participants being unable to clarify their opinions in a timely manner, resulting in delayed time performances. As a result, one small methodological change was made for the following interviews. This change was the timing of participants during task completion, as it allowed greater freedom for the participants and researchers to address concerns during task completion. Additionally, removal of this metric eliminated the limitations of Zoom's remote access features as the need for a formal baseline time for task completion. A baseline time would likely have to be established by navigating a comparable application as a gold standard and taking the average time for completing tasks. Therefore, establishing a 'good' or a 'bad' time was difficult without exceeding the time demands of the participant and breaking the implied consent agreement. Instead, a binary response of "Yes" or "No" was taken to account for any issues when navigating particular tasks as well as any details of potential issues encountered.

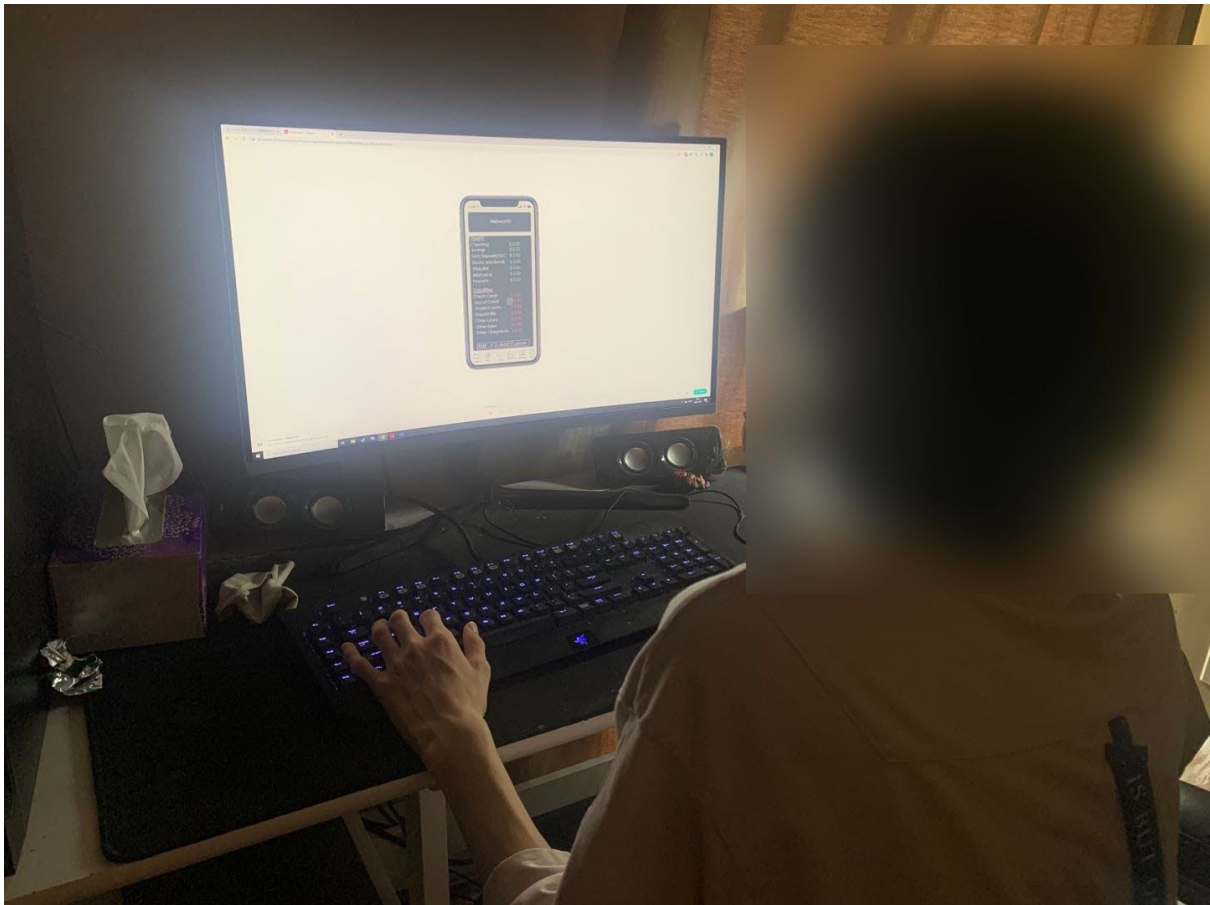


Figure 1. Participant Using the Prototype

The protocols employed involve the use of a set of questions asked to the participant as they complete two primary multi-step tasks as well as 4 minor single-step tasks (See Appendix for details). To complete the specific aspect of tasks, instructions were provided to the user; however, these instructions did not include details on how to navigate through the application. Ethnographic methodologies were utilized during this period as well, with documentation of user approaches being made during the task navigation process. Participant's questions or difficulties were also documented and assistance was provided when necessary. Questions asked during the completion of tasks are designed to obtain an understanding of the users intuition and thought processes as they navigate the application. Each step when navigating should be justified to ensure the application has maintained design fluidity and usability. If, from the participant's perspective, fluidity and usability was not maintained, these issues were documented in detail and noted as problems to be addressed. These issues would include visual bugs or inaccessible functions necessary for the completion of a task.

Following the completion of the tasks, an additional set of questions was asked of the Participant. These questions were designed to gain a holistic understanding of the users experience when navigating the application, drawing out any recommended changes, issues encountered, and overall opinions. Data taken from interviewed participants from both the pilot study and the main user interviews was divided according to individual tasks and post-test interviews. *Figure 1* shows a participant using the prototype. This picture was taken

with the consent of the participant. Key terms and errors were parsed and plotted on individual histograms to display any trends present in the data. Given the small sample size of the participants interviewed and high likelihood of Type II error skewing, formal statistical analysis was not applied. The primary goal of this study is to validate the prototype and gather as much information as possible to ensure a functional and beginner-friendly application is developed.

Results

Note: Similar to how applications are rated on the Apple iOS Store or Google Play Store, users were asked to rate the application on a scale of 1 to 5 stars. 1 being the lowest rating and 5 being the highest rating. Users and pilot testers were given followup questions upon completion of all tasks. The quantitative and qualitative results of all the participants is below in the respective categories.

Pilot study:

Aesthetics and Design:

Opinions of the applications design and aesthetic were mixed. With one participant reported liking the design of the application, but prefers more color. The other participant reported disliking the design due to the plain color scheme and minimalist UI.

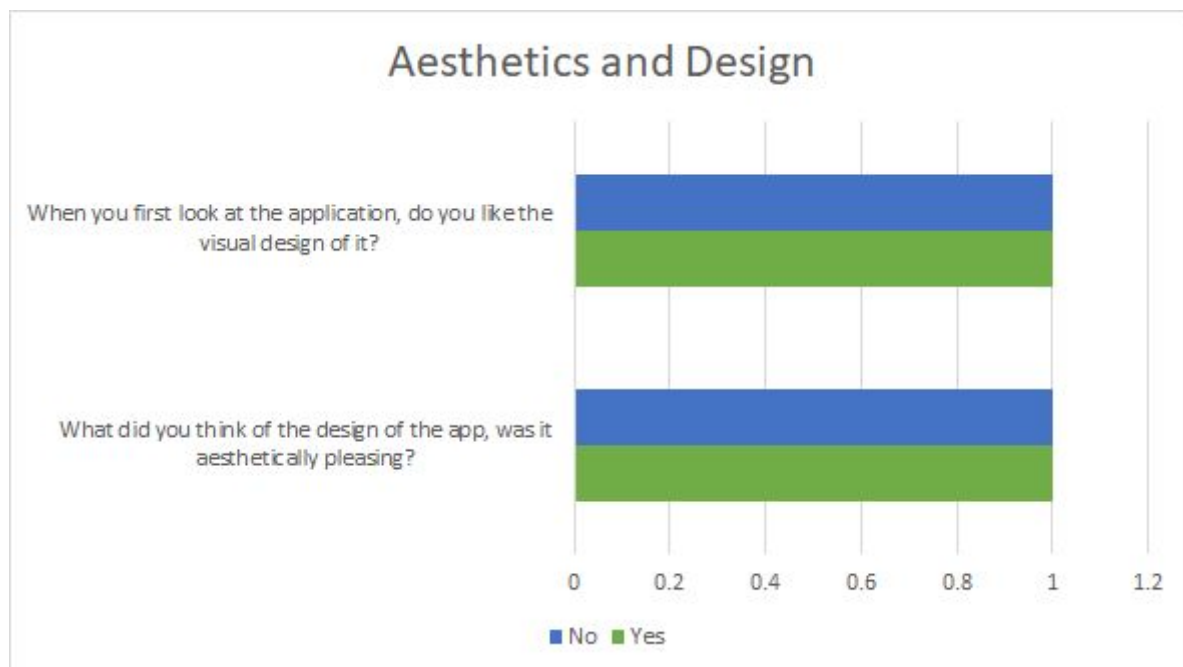


Figure 2. Pilot Session Opinions on Aesthetics and Design

Navigation and Usability:

Neither participant had trouble navigating the application overall, but one reported confusion with the number of income/expenses source given no change occurred. This is a limitation of

InVision. Both participants understood the description of the heading and content, but one reported issues navigating dates due to the lack of feedback from the date elements. Both participants think the features of the application would help them to manage their finances, with one participant specifically highlighting the benefit of the Net Worth feature.

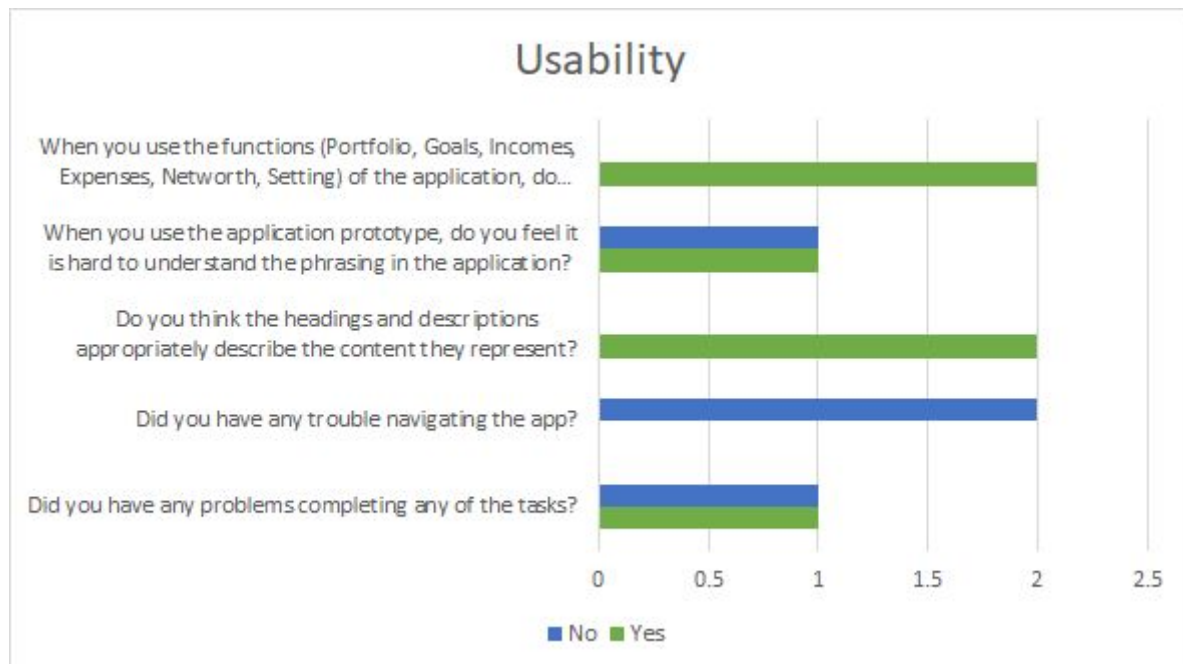


Figure 3. Pilot Session Opinions on Usability

As a result, the participants used more time on finishing the first and the second tasks, adding a new income source and editing an existing expense source, since they were figuring out if the changes have been implemented to the application. The number of the income and expenses did not change when the user submitted the changes. Additional, participants from the pilot session suggested improving the UI of the application, stating that the design was plain.

Overall Rating and Market Value:

Both participants would like to use the application if it is fully developed, and both think it will help them with their financial management. However, one participant does not think it is a viable financial management application given the user has to manually import the income and expenses.

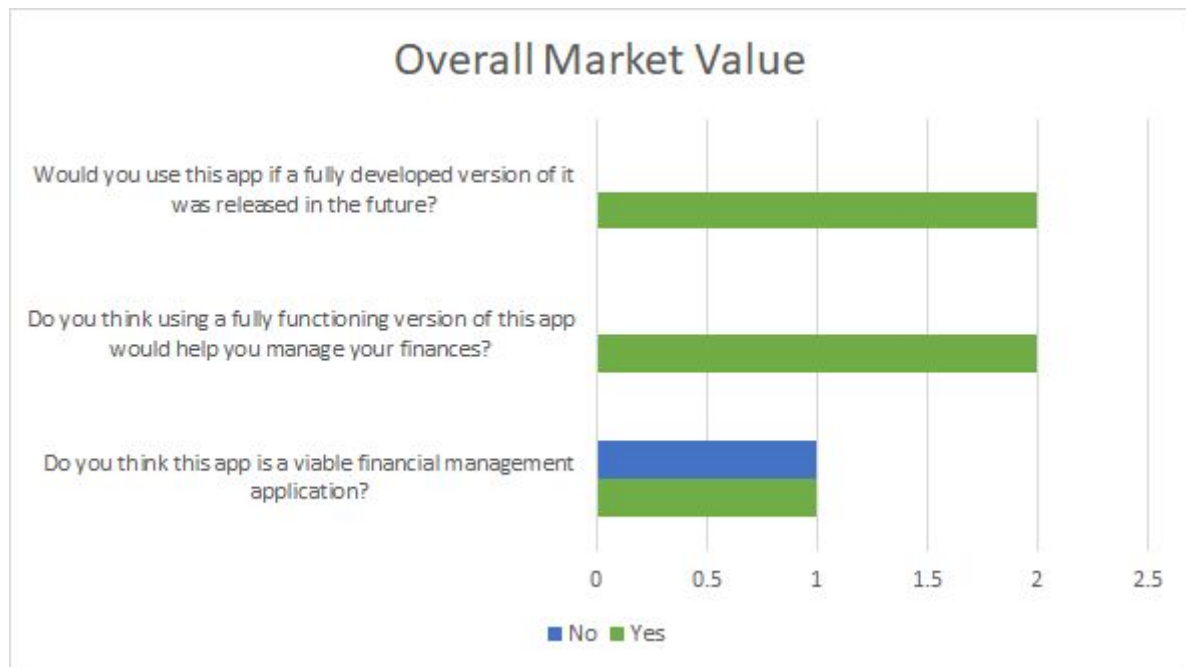


Figure 4. Pilot Session Opinions on Application Marketability

When rating the application, one participant rated it as 4, and the other a 3. The mean of the ratings is 3.5.

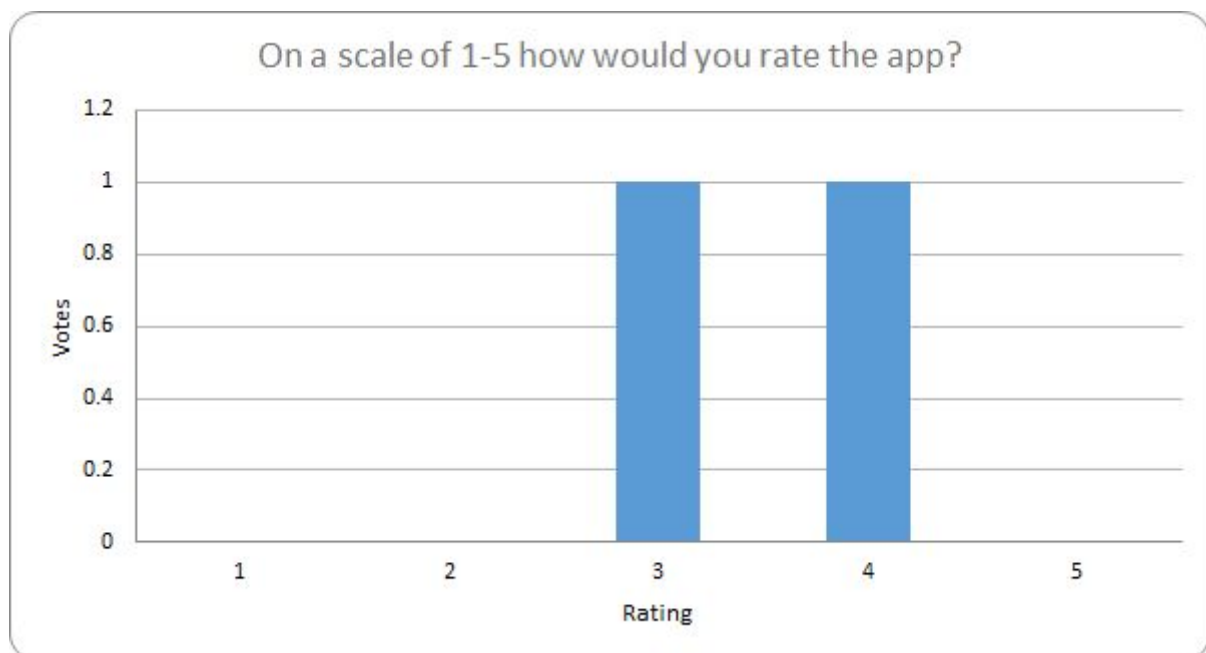


Figure 5. Pilot Session Application Rating on an App Store scale of 1-5

User evaluation session:

Aesthetics and Design:

Only 2 of 6 participants liked the visual design of the application. There was a general consensus among participants regarding the overall color palette, often saying it was 'plain' or 'dark'. This may have impacted participant opinion of the overall aesthetics and layout as opinions were evenly divided (3 of 6 'Yes' and 3 of 6 'No').

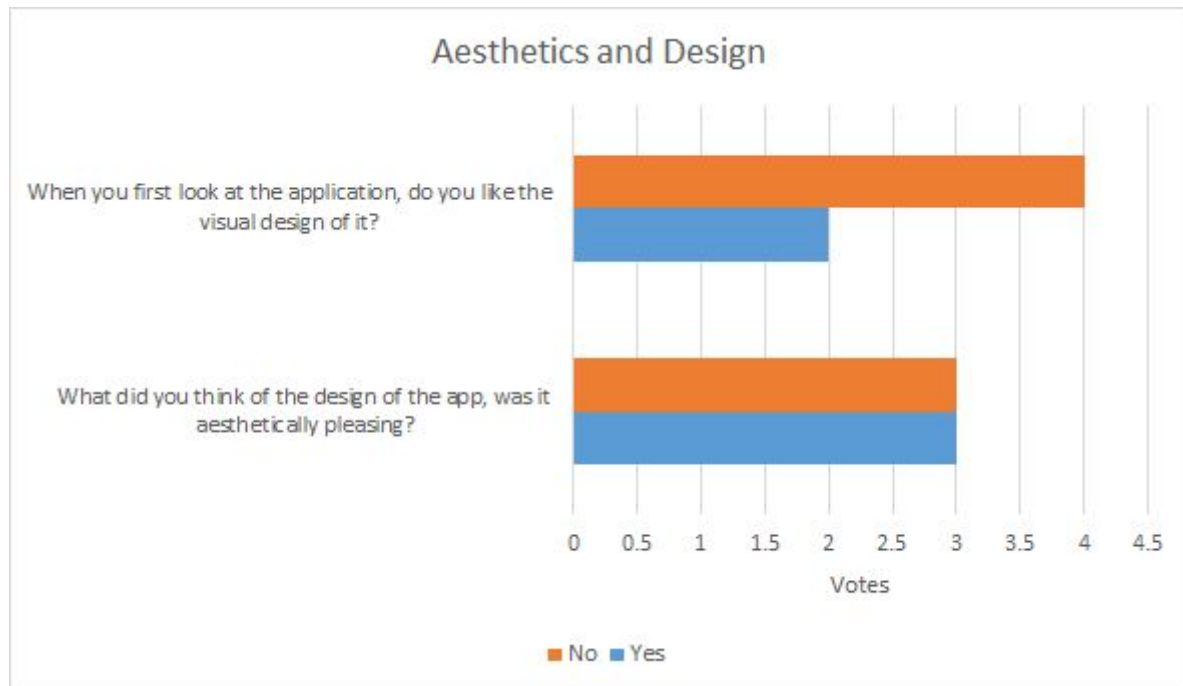


Figure 6. User Evaluation Aesthetics and Design Question Responses

Navigation and Usability:

3 of 6 participants had trouble completing any one task. 4 of 6 participants cited that they had trouble completing a specific task due to the applications's navigation. 5 of 6 participants felt that using the application's function would help them manage their finances. All users agreed that the bottom tab navigator and the current page headings were clear and adequately descriptive for what they represent.

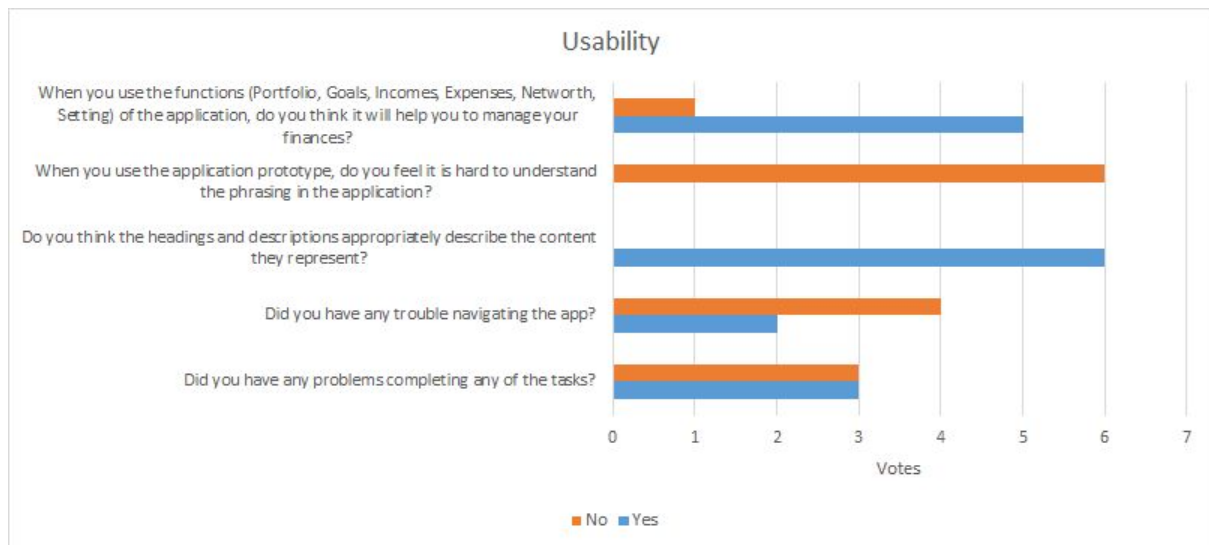


Figure 7. User Evaluation Usability Question Responses

Overall rating and market value:

Through averaging participant ratings, the application was given a 4.17 star rating (see Figure 8 below).

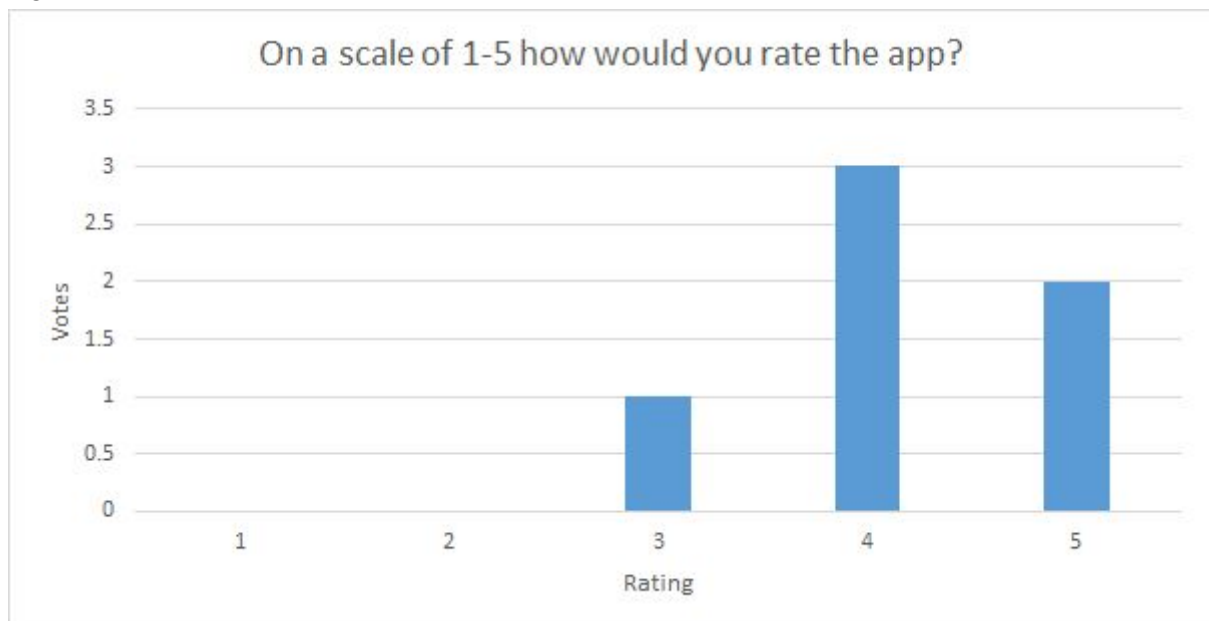


Figure 8. Overall Application Rating on a scale of 1-5 by 6 participants

4 of 6 participants thought that the application was a viable option for financial management and if the application was fully developed 5 of the 6 participants would use the application.

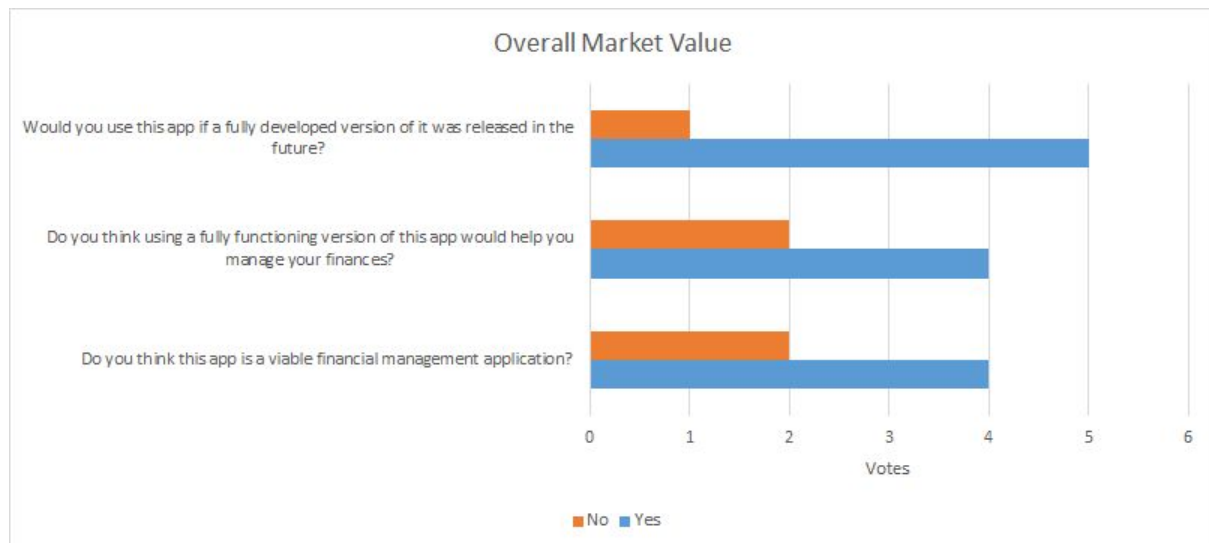


Figure 9. User Evaluation Opinions on Marketability of the Application

Discussion

One of the most agreed upon points brought up by the participants was a dislike of the design, with 4 out of 6 participants reporting that they disliked the visual design of the application (Figure 6). One of these 4 participants stated that they would like a brighter and more colorful interface, and mentioned that the “dark” theme of the application was not ideal for a financial application. Several participants mentioned that they felt the application had a lack of color in the design, claiming the pallet was limited to white, black, green, and red. Considering the large number of responses from the participants, we can infer that the neutral color tone of our application would not be suitable for long term application use. Users of the application would prefer to have a more vibrant color scheme when managing their finances.

Three of the participants had complaints when changing the months from July to August during the completion of task 3. One participant commented: “I was confused why clicking July 2020 income went to a calendar. It would be more obvious if there was a calendar icon to click on.” The mapping of the month element may be a source of confusion for the user due to it not being clearly defined as an interactable object. The user interface when switching between months is unintuitive and does not fall in line with the user’s conceptual model. Instead users suggested the addition of a “calendar icon”. An additional issue documented was confusion regarding where to add or modify existing income and expenses. Initial attempts users tried adding or modifying parameters through the settings menu. Another point of contention that was brought up by many participants was the necessity of manual input. Participants repeatedly cited manual additions are “too time consuming” or “inconvenient”. Some participants were uninterested in using the application because of the duration of time it would take to properly manage their account. They would be more interested if there was a way to automatically import their bank information directly to their account.

Overall it can be inferred that there is a general market for the application and the niche it covers. 4 of the six participants consider the application to be a viable financial management application and think using a fully functioning version of the application would help manage their finances. 5 of the 6 participants expressed interest in using a fully developed version of the application if one was released (Figure 9). The overall rating of the application on a scale of 1 to 5 is 4.2 with the median and mode of the rating being 4 (Figure 8). The application has a positive reception among the participants, but there is plenty of room for improvement.

Study Limitations and Reflections

Once all of our data had been compiled and analyzed, flaws in the study became more apparent. The sample size of the participants was quite small, with a total of 6. The lack of variance could make the data misleading since 6 participants does not accurately represent the average user, especially when the users are our friends and peers. Having our friends test our medium fidelity prototype may result in information gathered to be favorable towards the interviewer since friends may be less critical than strangers. When considering our data from Figure 8, there is an uncertainty in how realistic it would be if friends weren't rating our application. Since a "backend" couldn't be implemented without additional software (like Craft), our medium fidelity prototype lacked functional elements on the interface and couldn't be used to manipulate data. The InVision model's limited functionality made it difficult to conduct any research regarding the design of the input and edit menus, besides collecting feedback on the aesthetics. Some participants found the wording of the research questions misleading or confusing. In particular, participants were confused by questions one and two as well as questions five and ten, thinking they were asking the same thing. As Software engineering students, we sometimes use jargon that carries slightly different meanings for us than it would for those not in the field. This can affect the participant's understanding of some questions and they approach questions with a different interpretation than intended. The lack of interviewees having access to InVision (or the willingness to download InVision), severely decayed the user experience. 3 of 6 interviews had participants control the application through Zoom's remote control session, and participants did not have the opportunity to fully explore the application at their own pace.

Reflecting on the results collected and the limitations outlined above, there are a few things we would change about our application if we were given the chance. Firstly, we would change the color profile of the interface to make it feel brighter and more inviting. We might have also looked at implementing different color options or "themes" like dark theme or light theme. This was already partially addressed in our 'Settings' page, where a button to select custom themes is available but not implemented. Secondly, we would emphasize greater user feedback by making the distinction between text that is clickable (like the dates) and text that is not clickable more apparent. This could be done with the addition of colors to distinguish clickable text, or by adding underlined text. Thirdly, many participants disliked the fact that data had to be entered manually, despite this contradicting the core premise of our initial design (privacy). However, if user demands require us to do so, we may have to integrate methods of collecting user data automatically.

If we had a longer time period to conduct this study, we would change several aspects to our approach. First, if we increased the sample size of participants interviewed, the data we collect would be more reliable and we may discover more issues with our application. Having participants that were not associated with the researchers would also be preferable, as their observations would be more objective. Increased sample size would also allow for the integration of formal statistical analysis, and allow us to make claims more confidently. Second, making sure all the participants had InVision downloaded on their computers would have made for a more realistic testing environment. Having the users complete the tasks through remote access may have impacted our data quality. Finally, we would be more aware of the wording regarding some of the research questions. We would avoid terms that the participants might not understand or confuse the participants.