Part 4

Team 3

November 27, 2019

Table of Contents

Evaluation Techniques and Rationale **2**

Results of the Study **4**

Implications of the Results **11**

Revisions and Improvements **12**

Reflections **13**

Appendix A: Demographic Survey Results **14**

Appendix B. Team DLOGS **18**

Evaluation Techniques and Rationale

Task Descriptions

For report 3, we created a prototype mobile application which allows Boys & Girls Club volunteers to provide feedback to the club. Now that this is accomplished and tested by our team, we created a usability study to test the effectiveness and user experience our prototype provides users. To do this, we created scenarios that one would experience at a Boys & Girls Club and asked our participants to utilize the prototype to report these instances to the club. The first task accomplished with this study was providing general feedback to the Boys & Girls Club based on a scenario where the participant (a proxy volunteer) notices old fliers that need to be replaced. The second task our users performed is providing anonymous feedback to the club. The scenario related to this task is the user noticing two club members who seem to be neglected at home and the volunteer feels uncomfortable approaching staff to discuss this. The third and final task the users perform is using the club directory to find a staff member in order to provide feedback regarding this staff member. The corresponding scenario involves you finding the name of a club staff member that you spoke to in order to reference her in your feedback.

Since we do not have access to Boys & Girls Club volunteers, we used proxy volunteers to participate in our study. Eight of the ten users were in our class of CS 3750 while the other two were students at Georgia Tech. There are some drawbacks but also some advantages from this user base. Because these are college students, they are very in tune with technology. This is a good thing as the feedback we received from the usability study is extremely valuable to improving our prototype in both a technical sense and a human-centered design sense. This can also be seen as a negative, however, as the average person may not be as in tune with technology or as capable as the participants we used. Although we had a fairly diverse background in the ethnicity portion, the age and level of education statistics are highly concentrated in the college-aged and college-educated people. Because of this, we must use our results as a means to improving the technical and user interface of our design, but if we were to complete more testing, we would want to focus on the learnability from the point-of-view of a more diverse aged and educated sample.

We started our usability study with a demographic survey to gauge who the participants were. This is necessary information as we discussed earlier as it gives a very specific meaning to our results. From here we went on to usage scenarios where we asked participants to utilize our prototype according to scenarios we provided them. Finally, we had the users complete a post-test survey in order to receive their input on the experience. We had one group member working with the participant to navigate the prototype and answer any questions they may have.

We had two other group members evaluating the participant on their performance in a few measures including: speed, efficiency, and accuracy.

Justification

Using our evaluation techniques, we wanted to find quantitative and qualitative results to determine the overall usability of our prototype. We were able to test multiple different quantitative measurements through our study including the number of clicks for each task, time to complete each task, the number of errors made, and the number of questions asked. These values were recorded by each of the two evaluators to ensure the quality of the results, we also recorded the screen while the task was being performed in case the evaluators disagreed on any of the outcomes.

Through our usability study, we wanted to get a qualitative measure of the effectiveness and user experience of our prototype. To do this, once the test subject had completed our three tasks using the prototype, we had them fill out a survey that allowed them to qualify their experience.

Our testing was done mostly in a classroom setting where multiple other groups were also testing their prototypes. This created a loud and distracting environment for our study. To combat this, we had the group member working with the participant sit close and try to speak as loudly and clearly as possible to keep the participant focused. We also designed the experiment to be short so that the participant stayed engaged for their entire span of the study.

We conducted a Heuristic evaluation to help us to identify usability issues in our design. This allowed our users to evaluate each task on a set of heuristics. This method of evaluation was more ideal than the method of a cognitive walkthrough because we needed our whole system to be evaluated, not just one task. Initially, we tested our prototypes within the team under the supervision of our TA. Before our formal usability testing, we completed the formative assessment to address minor issues in our design and improve our prototype. Then, we completed the summative assessment using our participants to observe how easily the participant can utilize the prototype to complete the tasks. Using this approach of usability testing is cheap and fast, which is why we used this method rather than a field study. Although our participants do not represent the entire population, we had a sample of users that will provide valuable inputs regarding the usability of our design. To be consistent in our testing, we followed a script that detailed the responsibilities and objectives for the researcher.

Results of the Study

We randomly assigned a number to each participant. We had 10 participants and we have an equal distribution of gender. The age range is between 20 and 21. 50% of participants are Asians and the other 50% are either black, white or middle eastern. Other demographic information about participants can be found in Appendix A. Each participant also signed a consent (IRB) form that laid out the purpose of our study and the rights of the participants, such as the right to leave at any point of the study or the right to remain anonymous throughout the entire study.

Data Analysis

We measured the time one takes to finish multiple tasks individually to measure the learnability and to see how easy it is for one to learn the app and remember it when using it next time. We also measured the number of errors made to see the overall efficiency of our design. By considering the amount of time, the number of errors made, observing how comfortable they were to complete the tasks and doing post-questionnaire we measured the learnability, efficiency, user experience, utility, and accessibility of our design. We also have tasks to see how anonymously they can submit the feedback and how satisfied users are with this option. Our results show that 9 out of 10 people were highly likely or likely to use our design whereas the only one person said neutral, but he/she said it was easy to use and it was a simple design.

For task 1, which was to submit feedback with the option to choose the type of feedback and who the feedback concerns, the average time it took the user to complete the task was 39 seconds. The average number of clicks was 9 and only two people out of 10 people made an error.

	AVG Time to Completion (1)	AVG Number of clicks (1)	Number of Errors (1)	Number of questions asked (1)
	45.70	14	0	0
	33.63	10	1	1
	28.19	10	0	0
	41.64	13	1	0
	52.65	6	0	0
	29.17	9	0	0
	43.79	5	0	0
	50.76	8	0	0
	29.14	9	0	0
Average	39.41	9		

Figure A. Average Time to completion, number of clicks, number of errors, and number of questions asked for Task 1

For task 2, Submit two feedback with an option to remain anonymous, the average time it took the user to complete the task was 40 seconds. Looking closely at the average time it is evident that the user took less time to finish the task. As in task 1, on average it took them 39 seconds to submit one form but this time they submitted the two forms in 40 seconds. It was one of our ways to measure the learnability. Thus, based on our data it is correct to say that while performing task 1, the user has learned how to use the app and remember it for the next time.

Furthermore, the average number of clicks was 15 which seems reasonable in comparison to task 1 as submitting the second feedback doesn't require login again which minimizes the number of clicks to login but it adds additional click to click the button to submit another feedback. However, this time three people made an error while finishing the task and one of the main reasons observed by our evaluators was users get confused when to submit the second feedback by navigating through clicking the button "Submit another feedback" as they forget the entire task and forget to submit the second feedback or get confused to submit two feedback separately for two children.

	AVG Time to Completion (2)	AVG Number of clicks (2)	Number of Errors (2)	Number of questions asked (2)
	54.83	20	0	0
	31.54	15	0	0
	29.07	16	0	0
	34.07	15	0	0
	50.64	10	0	0
	34.59	15	1	0
	59.21	13	2	2
	41.62	16	1	0
	30.35	16	0	0
Average	40.66	15		

Figure B. Average Time to completion, number of clicks, number of errors, and number of questions asked for Task 2

In the end, through our post questionnaire, we asked users how comfortable they feel submitting the feedback and as seen in Figure C, 80% said they were very confident submitting the anonymous feedback. Nobody said they didn't feel confident submitting it anonymously.

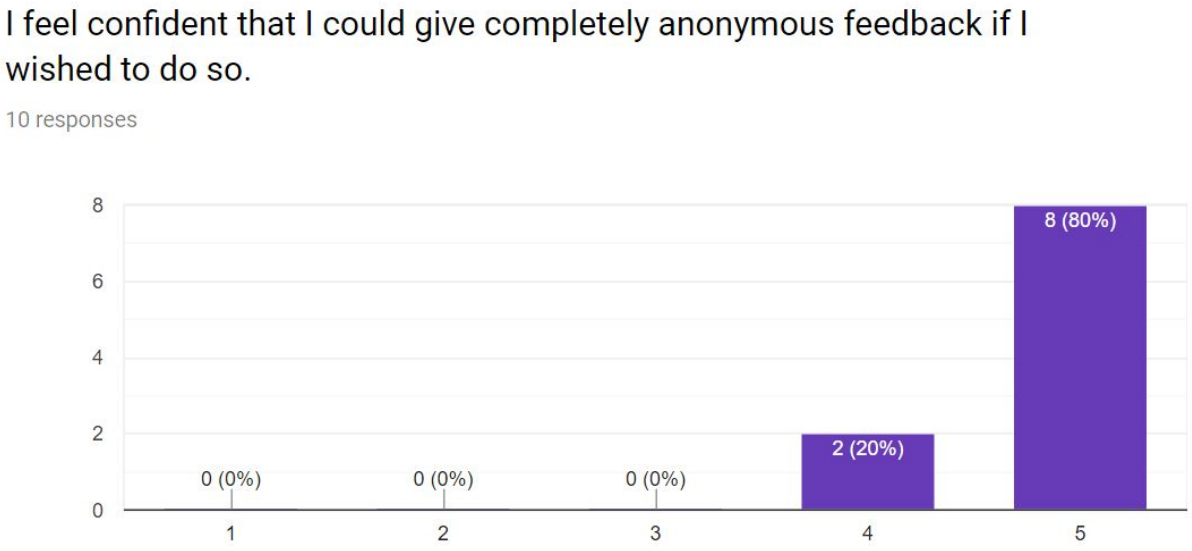


Figure C. Anonymity measurement during Post-Survey Evaluation

For task 3, Look up BGCMA staff and employees for reference when submitting feedback, the average time the user took was 47.99 seconds, approximately 48 seconds. We have expected the user to take the longest time to finish this task as it was one of the longest and it complicated task. It was complicated in a way as before submitting the feedback a user has to navigate to the directory menu and from there a user can search the staff and then submit the feedback. As expected it took more time. Again for this task, three people made the error while finishing the task and four people out of ten people asked the questions. Two of them asked two questions and the other two asked one question. Based on our evaluation most common mistake made or confusion one had while completing the task was navigating between different pages which increased the number of clicks. Also, users get a little confused about how to search for the staff when they were on the directory page. It may be because our prototype lacks the actual feature of typing in the search box.

	AVG Time to Completion (3)	AVG Number of clicks (3)	Number of Errors (3)	Number of questions asked (3)
	36.32	13	0	0
	44.10	13	0	2
	41.58	12	0	0
	21.69	12	0	0
	78.82	12	0	0
	37.89	12	0	0
	59.85	10	1	2
	44.95	12	1	1
	66.73	25	1	1
Average	47.99	13		

Figure D. Average Time to completion, number of clicks, number of errors, and number of questions asked for Task 3

Overall, based on our Post-Test questionnaire and evaluation forms, everyone liked our design, and we received feedback like Aesthetic, ease of use, simple transitions, easy to navigate (Figure E), like the option of being anonymous, and liked the directory feature.

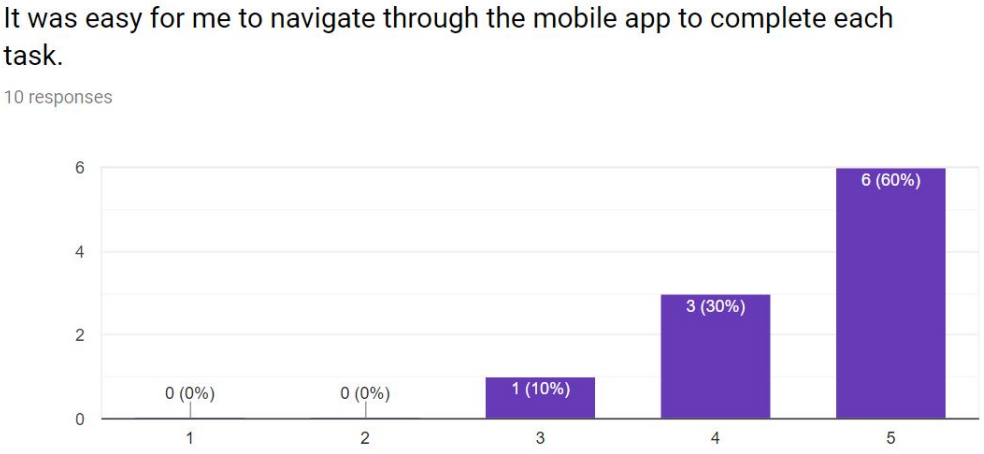


Figure E. User experience measurement during the Post-Survey Evaluation

It is evident from our data that users liked the app and it is something that can be implemented in the future with minor improvements. (Figure F)

Figure F. Participants' input on what they liked about the application

We also got some constructive feedback which we will take into account to improve our design. We received feedback like more guidance in the app, few users found it confusing submitting the feedback especially when one has to look up the person in the directory and it may be since in our prototype they can't actually type in the search field, so it was causing some confusions. Based on the feedback we realized we can reduce the number of clicks or time by adding three bars (main option) on every screen. Also, we learned that we should add some obvious signifiers when one is trying to lookup the staff member and submit the feedback to avoid confusion and so one can perform the task with fewer difficulties and efficiently with no frustration. (Figure G)

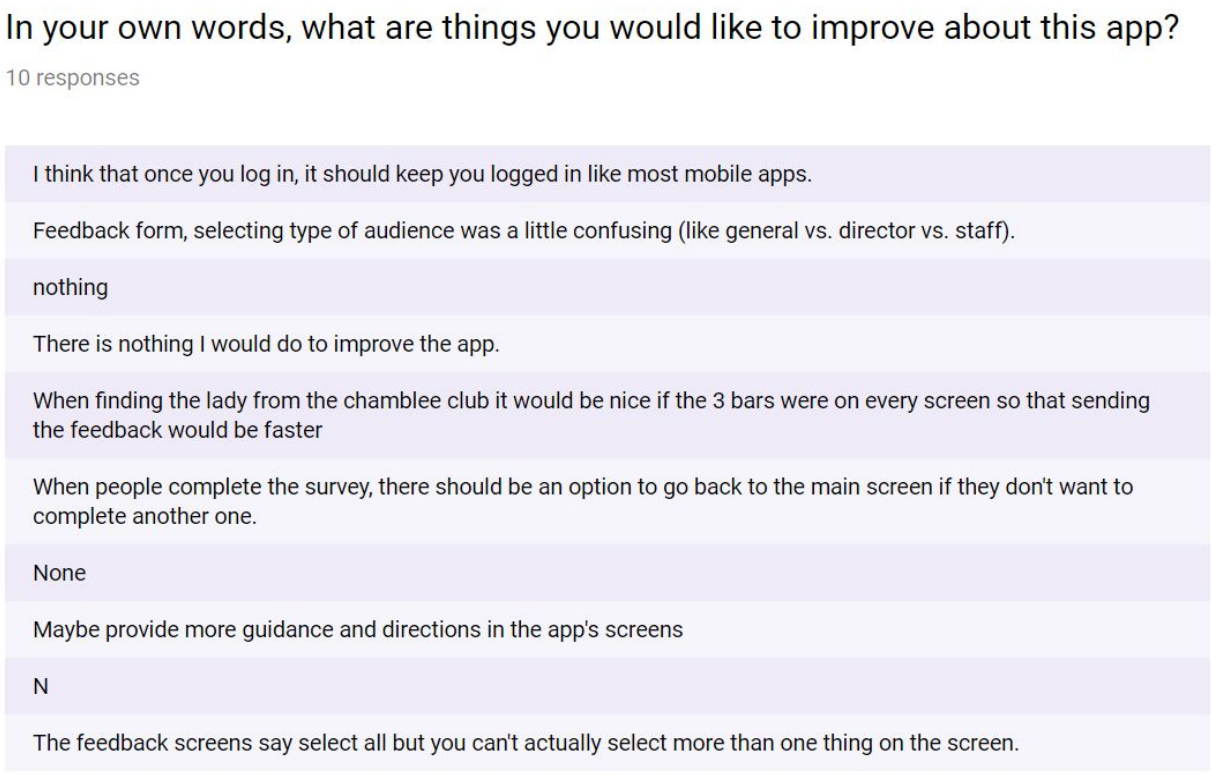


Figure G. Participants' input on their ideas for improvement

The following diagrams show our results and it is evident that users were satisfied using our app and they would consider using it.

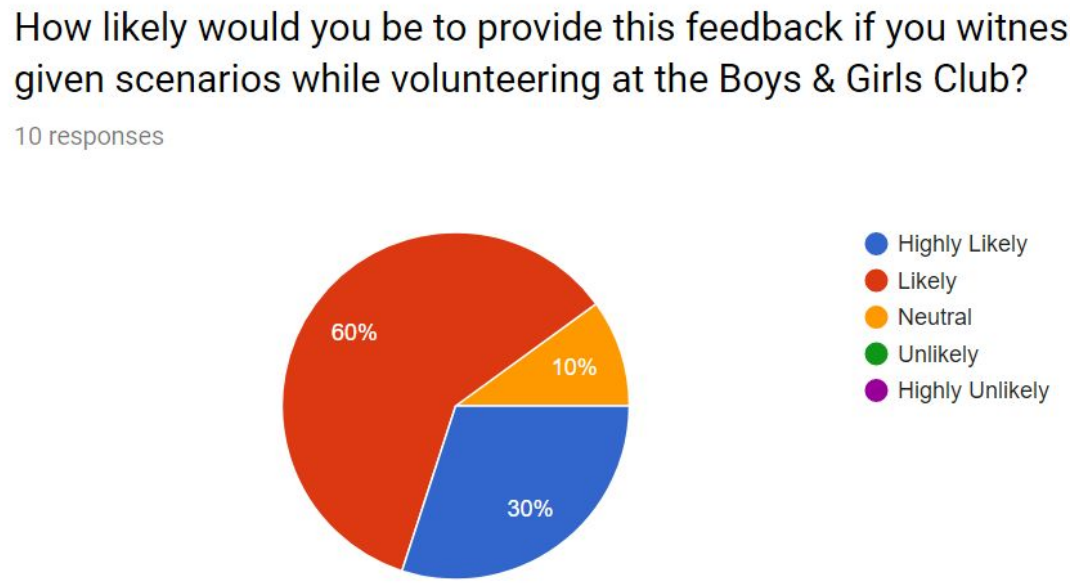


Figure H

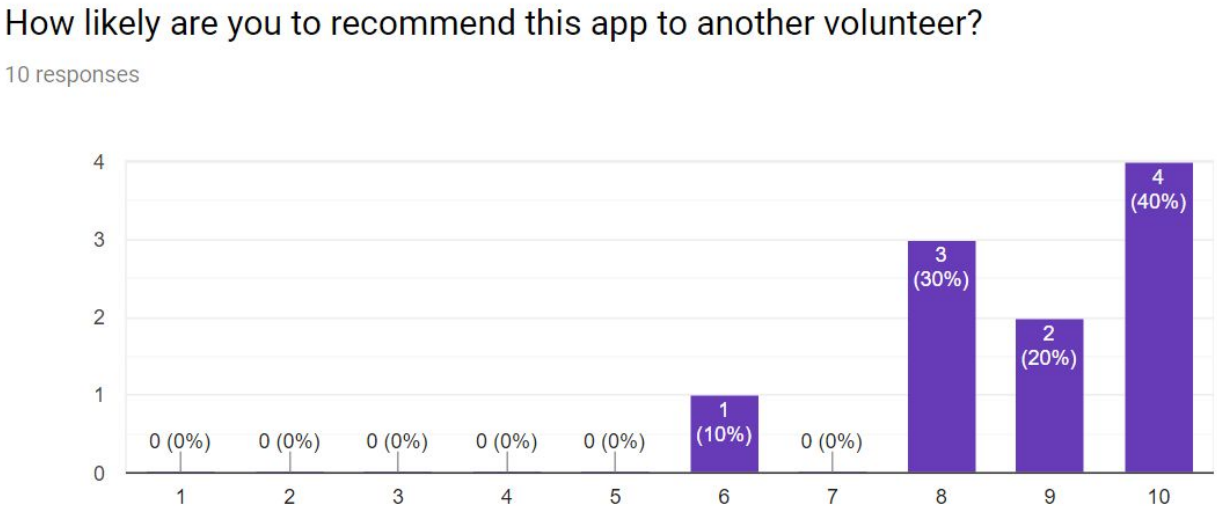


Figure I

Implications of the Results

Overall, the results we received from our usability testing gave us a good overview of how efficient our prototype was and what we did successfully to reach our goal. Each of our participants easily completed the three tasks we assigned to them in a reasonable amount of time. Some participants took longer than others, but almost everyone completed each task in under a minute. For example, there was one participant who took approximately 66 seconds to complete the third task, but this participant also asked a question during this task, so it correlates that the time it took for this participant to complete the task increased because of some kind of confusion that happened during the evaluation. Also, because of the short amount of time it took most of our participants to complete the tasks, we can conclude that the learnability of our product was pretty high in that it did not take long for the users to navigate through the application.

There were almost no errors that occurred during the testing, and the ones that did occur were only a result of confusion of the tasks themselves, not because of our prototype. A lot of our participants had to ask us who to direct the feedback to and who the feedback was regarding during the evaluation. Although we clearly labeled the different pages for who the feedback is about and who the feedback should be directed to, we also had some participants ask us questions about the pages on our application about which was which. This could have been a result of simple human error in that the participants could have skipped over the description of each page because they had the burden to complete the task while being timed. Because of the lack of error during the simulation, we can safely say that the efficiency of our application is superb in that it successfully allowed users to complete each task with no obstacle.

Similarly, the subjective measures of user experience and the anonymity of our product provided a result that supported our goal. We asked users in the post-survey evaluation their personal opinions of the overall usage and experience of the application. Nine out of ten participants said that it was very easy to navigate through our application to complete each task. Also, because anonymity is one of the key aspects we wanted to implement into our prototype, we asked users how confident they felt that their input would remain anonymous. Eight out of ten users said that they felt completely confident in remaining anonymous when giving their feedback. Although these are subjective results and may vary depending on the user, our application overall successfully took into consideration our concerns of the users' experiences and anonymity.

Revisions and Improvements

All of our evaluation was solely focused on the user. However, our design for an app requires that BGCA staff members also use it since they are the ones to review the feedback and update the site with information (such as a change in leadership). In future implementations, we would like to focus on the faculty-facing portion of the app.

One way that we can improve the UX is to add a review page before the user finishes submitting the feedback. This would allow users to ensure that all of their answers are correct. Additionally, for feedback containing sensitive material, it's important that users have an opportunity to ensure that their wording and anonymity are both accurate. Another improvement would be to make the "Go back home" text larger on the successful submission screen. While the current font size maintains the convention of the previous pages, users didn't immediately look there to navigate back to the start. This issue could also be solved by putting a separate button that navigates home elsewhere on the screen. However, further iteration and testing would need to be conducted to determine the better solution.

During the feedback submission flow, back arrows oriented above the progress bar allow the user to navigate to the previous screen. Currently only the arrows are clickable, however, future iterations should make the entire area above the bar clickable. Because the evaluation was conducted on a web-based version of the mobile app, users with larger hands or dexterity issues didn't have any issues clicking the buttons. However, this may become an issue on an actual mobile device. Increasing the click area would mitigate this issue and increase the accessibility of the app. Additional improvements such as text-to-speech input would also better accessibility.

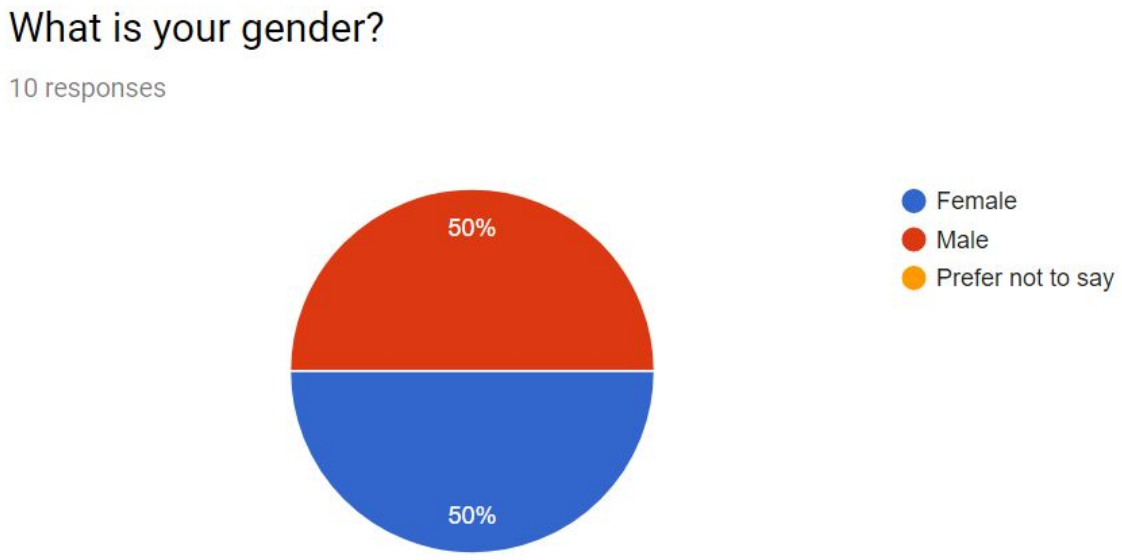
Reflections

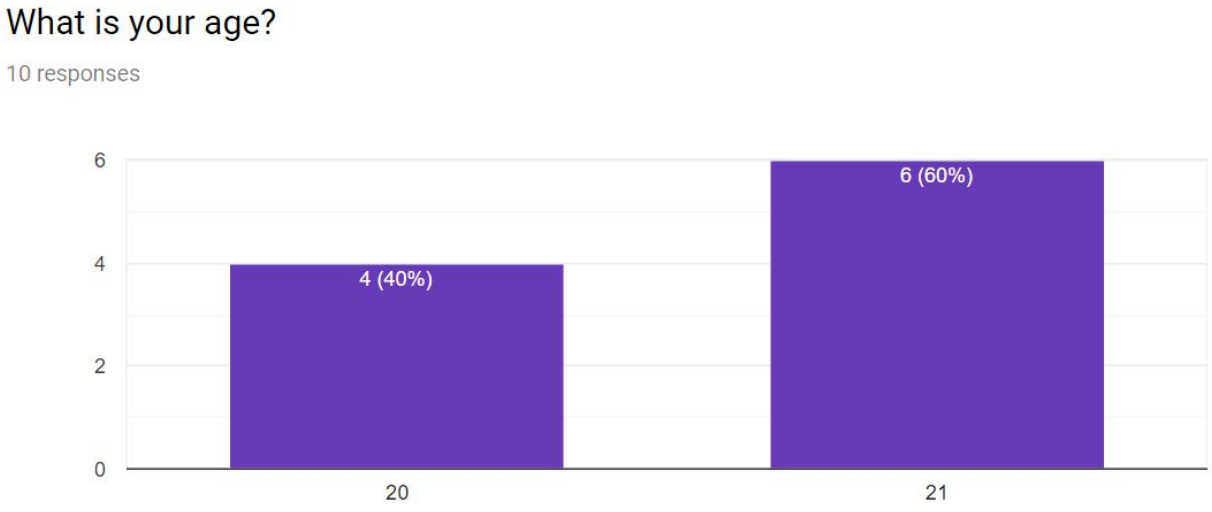
While working through the different stages of designing a capable feedback system for the BGCA, we realized how difficult it is to anticipate issues people run into while using interfaces like apps. As customers it is very easy to use an app or a website and criticize the developers, especially when a lot of the blame is placed on a developer's "lack" of coding knowledge. However, throughout this course, which did not utilize any coding whatsoever, we now understand that a lot of design issues have nothing to do with the actual code itself. Rather, trying to foresee all the design problems that users will encounter is very hard.

As a team, it was interesting to discover the various perspectives from each of our group members. Most CS group projects are code oriented which results in members dividing up the code by classes and methods and discussed the most efficient way to implement the code. They are rarely concerned with the users who will utilize their final product. We each ended up realizing that each of us were so focused on the coding aspects of CS. However, this class taught us to always keep in mind the target audience of the final product.

If we were to start the project over, we would like to attend a BGCA event and discuss with the volunteers what their thoughts are on the current systems in place. At times it was difficult to envision ourselves in the place of a volunteer and try to design a feedback system that they would consistently use. We realized how valuable this insight would be after talking with the BGCA staff members about their expectations of a feedback system which allowed us to efficiently design the system to sort feedback to make it easier to classify the submitted forms. However, since we didn't have any of the volunteers' expectations, we had to imagine ourselves in that situation and design based on that.

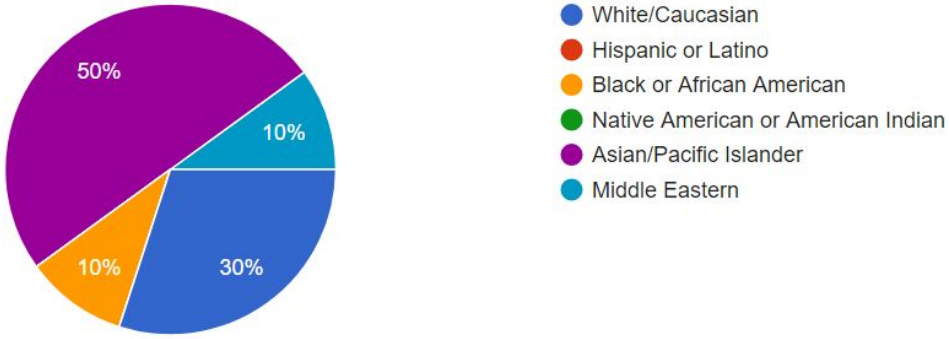
Appendix A: Demographic Survey Results





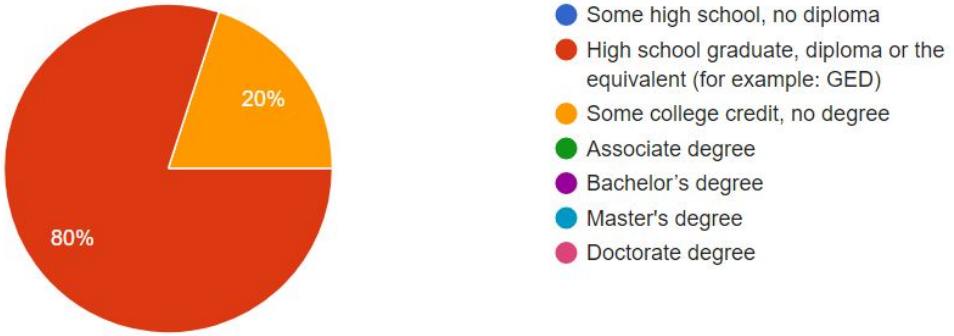
Please specify your ethnicity

10 responses



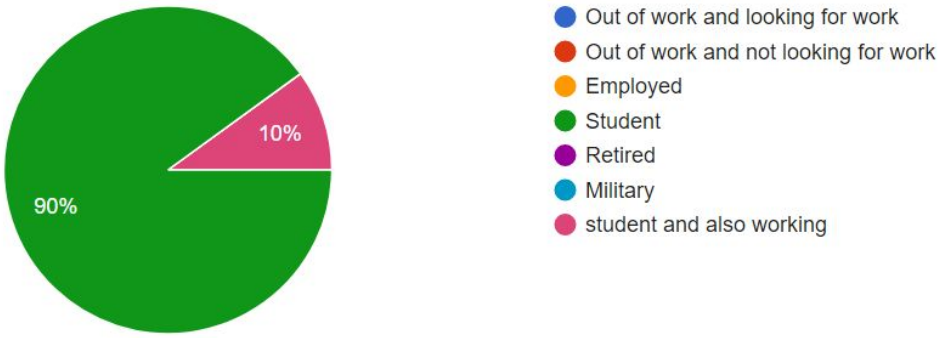
What is the highest degree or level of school you have completed? If currently enrolled, highest degree received.

10 responses



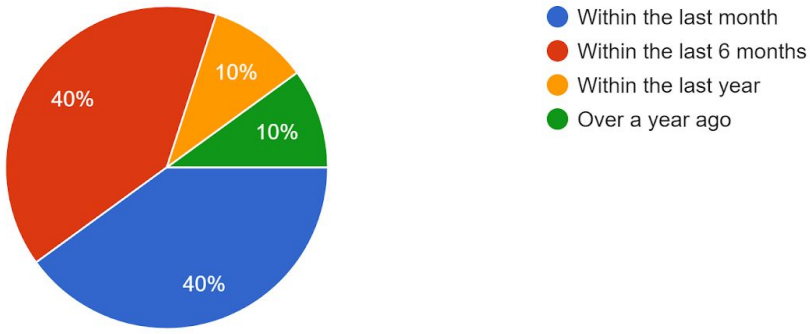
Employment Status

10 responses



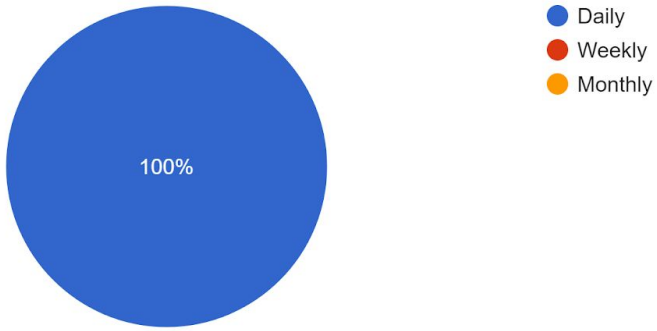
When was the last time you participated in volunteer/philanthropy work

10 responses



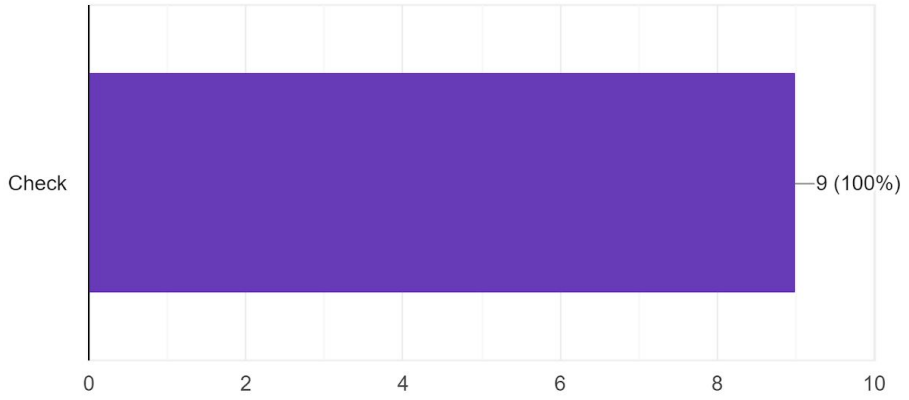
How often do you use your mobile phone for things besides texting/calling?

10 responses

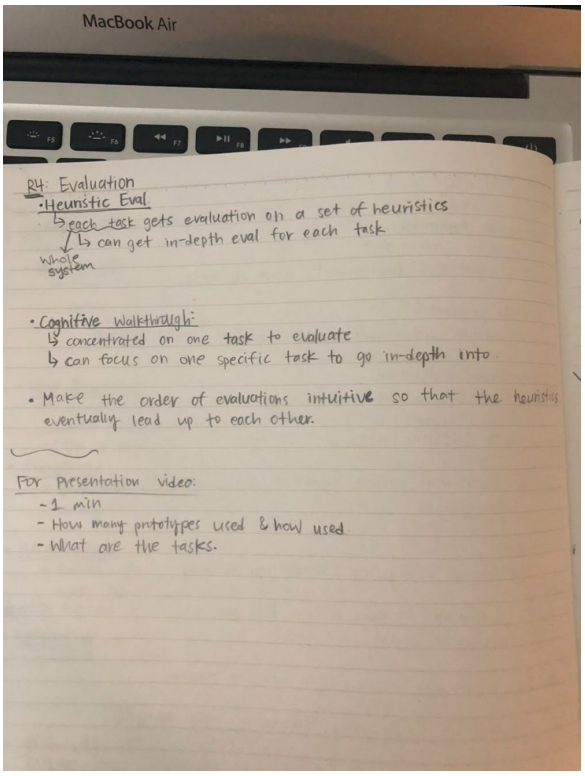


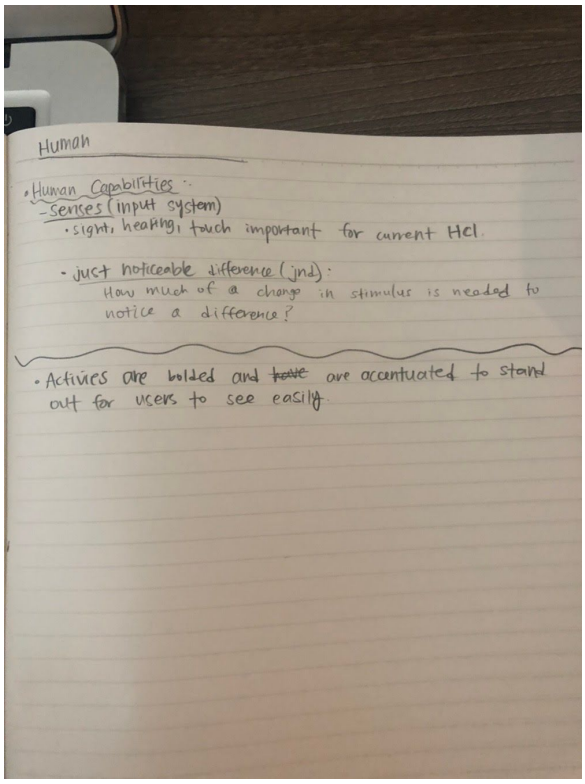
By checking the box below, I agree to the terms of the consent form and to the fact that I was given information about the form.

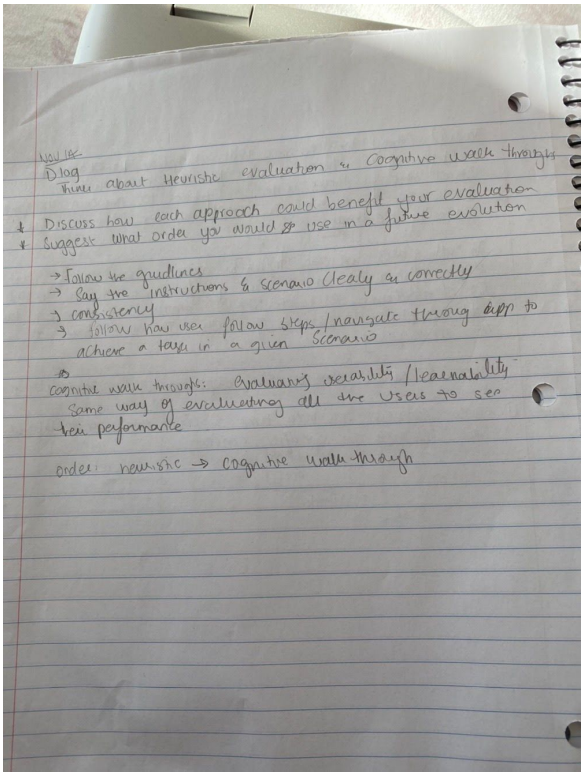
9 responses

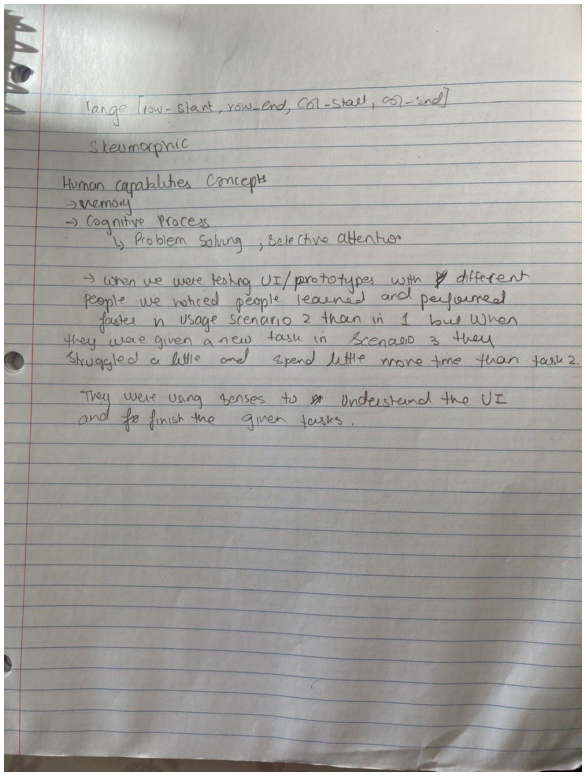


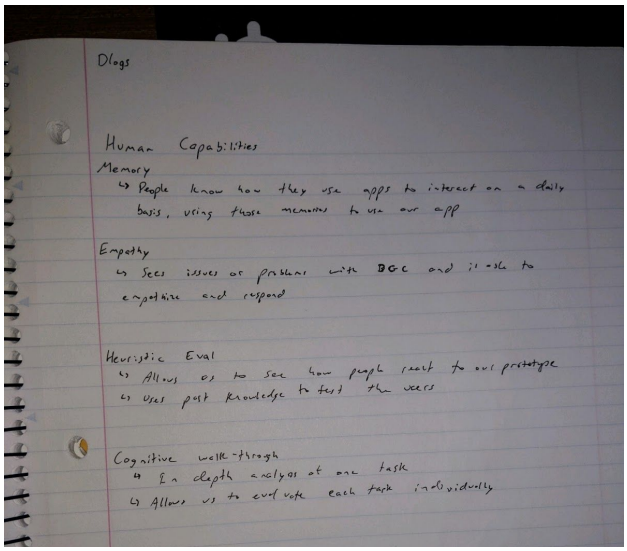
Appendix B. Team DLOGS











Dlogs

Human Capabilities

Memory
↳ People know how they use apps to interact on a daily basis, using those memories to use our app

Empathy
↳ Sees issues or problems with DGC and is able to empathize and respond

Heuristic Eval
↳ Allows us to see how people react to our prototype
↳ Uses past knowledge to test the users

Cognitive walk-through
↳ In depth analysis of one task
↳ Allows us to evaluate each task individually

Heuristic evaluation: overview of the entire system; often performed by those outside the design process

Cognitive Walkthrough: zooms in on a specific flow or task and examines its ability to meet the goal

Human Capabilities

Memory: people use existing mental models and experiences to interact with the app
↳ ppl also use problem solving and selective attention to become more efficient

Empathy: people respond quicker and with more vigor to BUGA issues because they mean something