Tracking Mental Health:
A Study On Mental Wellness In Underprivileged Children

By Jackie Kuwabara, Sangyun Park, Christophorus William Wijaya,
And Nouzhan Vakili Dastjerd

Table of Contents

Description of the evaluation techniques, tasks and users involved in our study

---

As our study was divided into two sections: one for our mobile app and one for our physical device, this section is also divided into two for clarity.

Mobile Application:

Users:

In our mobile app portion of our study, our users were our fellow classmates as well as **Stanley**, our TA. This encompasses a group consisting of ~18-24 year old college students.
NOTE: This group in our study is considerably different from the target user group of BGCMA Staff Members.

Tasks:

Our mobile application, we had two main functionalities we wanted to test:

- **Function #1**: BGCMA staff views daily and long term analytics on children's feedback data. They will be able to filter the data by day of the week, weekly, monthly, yearly, or date range.
- **Function #2**: Staff receives a notification when certain stress-level was detected, and this settings can be configured.

We tested these functionalities using the following tasks:

Task 1: Following through with a notification:

- There is an alert on your lockscreen notifying you about a benchmark. Try and see if you can find out what the benchmark is referring to.
  - **Evaluation Technique:** Users were recorded with a stopwatch as soon as they began the task to the moment they completed it. Additionally, we recorded the number of clicks it took for users to progress through the system.

Task 2: Check Today's Stats:

- Consider this scenario: it seems like a rather good day at the BGCMA! Kids are smiling, and you don't seem to notice too many grumpy children. Can you try to check how many kids rated the saddest rating today?
  - **Evaluation Technique:** Users were recorded with a stopwatch as soon as they began the task to the moment they completed it. Additionally, we recorded the number of clicks it took for users to progress through the system.

Task 3: Viewing Analytics:

- From the main dashboard, try to view the analytics page.
  - **Evaluation Technique:** Users were recorded with a stopwatch as soon as they began the task to the moment they completed it. Additionally, we recorded the number of clicks it took for users to progress through the system.

Task 4: Filtering Analytics:

- Consider this scenario: it's November and you're noticing the club seems more positive as a whole! You start to think there might be signs of this upwards trend in the mental health app's analytics. To check if you're right, filter the analytics to only show a certain date range (09/06/2019 - 10/12/2019).
    - **Evaluation Technique:** Users were recorded with a stopwatch as soon as they began the task to the moment they completed it. Additionally, we recorded the number of clicks it took for users to progress through the system.

Physical Device:

Users:

In our physical device portion of our study, our users were also our fellow classmates as well as **Stanley**, our TA. This encompasses a group consisting of ~18-24 year old college students. NOTE: This group in our study is considerably different from the target user group of BGCMA attendees, a group ranging from young children to teenagers.

Tasks:

Our mobile application, we had a single functionality we wanted to test:

- **Function #1**: Children at the BGCMA are able to accurately select a button on the physical device that corresponds to their current mood.

We tested this functionality using the following tasks:

Task 1: Ordering the Faces

- Our researcher will now hand you four cut out faces. We ask you to arrange the faces from saddest to happiest and describe what each face means to you.
    - **Evaluation Technique:** Users were recorded with a stopwatch as soon as they began the task to the moment they completed it. Additionally, we recorded the order that our participants decided on.

Task 2: Choosing an Emotion

- For this exercise we will be giving you 4 different scenarios for the 2 children. After each scenario please press the face that most matches the child's mood.
    - **Evaluation Technique:** Users were recorded with a stopwatch as soon as they began the task to the moment they completed it. Additionally, we recorded the emotion our participants selected.

Task 3: Quick Matching

- We are going to give you a word and please react to the first emotion that pops into your head.

        - *Words*: bullying, candy, movies, recess, bad grades, divorce, poverty, lost jobs, birthdays, toys, puzzles, phones
    - **Evaluation Technique:** Users were recorded with a stopwatch as soon as they began the task to the moment they completed it. Additionally, we recorded the emotion our participants selected.

Justifying Our Particular Techniques

Mobile App

For our mobile application, all our tasks are centered around objective measurements -- that is, they are focused entirely on metrics. Primarily, we record the time and number of clicks it takes for participants to complete several of the main functions of our application. In doing so, we are able to obtain an observable metric by which we can then compare our system to others with similar functionalities. Considering our two functions, these tasks apply directly to them. Viewing and filtering the analytics corresponds with function #1. Similarly, seeing if a participant can follow through with a notification and find out what it refers to like we do in tasks one and two directly correlate to function #2. By evaluating the application in comparison with similar systems, we are able to see how well-designed, intuitive, and accessible our application is (according to our previously defined definitions of these terms).

- **Function #1**: BGCMA staff views daily and long term analytics on children's feedback data. They will be able to filter the data by day of the week, weekly, monthly, yearly, or date range.
- **Function #2**: Staff receives a notification when certain stress-level was detected, and this settings can be configured.

For the first task, the screen shows a notification and the user presses the notification to be taken to the screen containing pertinent information. All participants very easily clicked the notification, as it is the same as it would be for any smartphone application,and the use of one click shows us that the app will be **efficient** and **intuitive** - as this is our measure for efficiency. However, participants did not understand the meaning of the benchmark at first glance, indicating to us that while there is no difficulty in **usability**, there may be some in **learnability**, and we will have to do a good job of explaining what the project's intended use is in the future participant studies.

Our second task was for participants to use the dashboard to view the daily statistics, and tell us how many kids chose the saddest rating on our scale. Participants completed this task with about 3 clicks, and they correctly told us how many kids had rated the saddest rating on that day. However they did have a little trouble figuring out exactly what they were looking at. This indicated to us while the app is easy to use, perhaps the text and visualization the **aesthetics requirement**) is not as easy to understand, and we should be doing a better job of explaining the display.

Our third task is for participants to navigate to the analytics page from the main dashboard. Again during this task we saw that participants quickly and easily identified the screen, and upon doing so, we asked them to complete our fourth task, which is to filter the ratings data so that they can only see a certain age range. This fourth task was completed in six clicks by participants and took more time than the other tasks. While the fourth task is the most

complicated of the set, so the the clicks indicate efficiency are higher, we still found that participants were slow to understand what was going on immediately, but got the idea after trying more than once, which indicates to us that we need to be doing a better job of either explaining before the study or including **signifiers** in the app to increase **learnability**. Upon the completion of these tasks we have learnt that we need to spell out some words, include additional signifiers, and add additional filters to include both specified dates as well as frequencies. We also realized that the physical prototype must be the first one used in the study because we do not want the participant to gain a sense of where we rank the faces (see Task 2 of our physical prototype). FInally, we know that it needs to be made clear to participants when they are switching from the physical prototype to the mobile one that their perspective will be changing as well, from that of a child to that of an administrator or staff member.

Physical Device

Given the uncommon format of the physical device, the methods in which we evaluate its effectiveness through tasks 1-3 involved much deliberation and eventually brought us to the tasks we selected for the following reasons:

For the first task where participants reorder the scattered face images into a sequence of faces from the saddest to the happiest allows us to test whether our suggested emojis are consistent and will be able to represent the appropriate moods accurately. **The purpose of the first task is to validate the depiction of emojis related to the scale. If the participant has the same scale as we designed, it shows the intuitiveness and easy understandability we have created. Task 1 also helps us build our conceptual models and is the first exposure to the scale, eliminating any predisposed bias.**

For the second task, **we were attempting to see how different individuals evaluate different situations; in doing so, we would be able to tell whether or not our expected choices (the "gauges" we set for each button) were appropriate and accurate. One of the main concerns and disadvantages to our system is that there is no correct answer and everything is individual interpretation. Thus, task two allows us to gauge more knowledge of how people scale a very sad day versus just a sad day.**

For the third task, we wanted to further observe **how certain conditions could gauge people's emotions.** By listing out 12 life-related words that consist of "good" and "bad" words and asking for the first emotion that comes to mind, we are able to *fine tune* our physical device's buttons. It allows us to see if our emotions are regularly felt, appropriate, or even if we may need to add more emotional states.

Results of Our Study

Physical Prototype Study Results

The tables below show the objective data that we have collected from our physical prototype usability testing on the 19th and 20th of November 2019. Based on those quantitative data, we are going to analyze how consistent does our emoticons depict human emotion. On the first task, we took 2 types of feedback: completion time and the order of the emoticons from the saddest into the happiest. The feedback that we kept for the second task is the choice that our testing participants chose when they were given a few scenarios. For the third task, we collected their choices of emoticons that they think are the best to represent the words. Note that apart from the completion time on the first task, we use numbers to represent the emoticons **Note: (number 1 is the saddest and number 4 is the happiest).**

Task 1

Trial Number	**Time (seconds)**	**Result**
1	7	1→2→3→4
2	20	1→2→3→4
3	5	1→2→3→4
4	7	1→2→3→4
5	10.71	1→2→3→4
6	7.02	1→2→3→4
7	8.33	1→2→3→4
8	8	1→2→3→4
9	5	1→2→3→4
10	10.02	1→2→3→4
Average	**8.808**	**1→2→3→4**

Task 2

Trial Number	Scenario 1	Scenario 2	Scenario 3	Scenario 4
1	2	4	4	1
2	2	4	4	1
3	1	4	3	1
4	2	4	4	2
5	2	4	4	2
6	1	4	1	3
7	2	4	4	1
8	2	4	3	2
9	1	4	4	1
10	2	4	4	3
Average	**1.7**	**4**	**3.5**	**1.7**

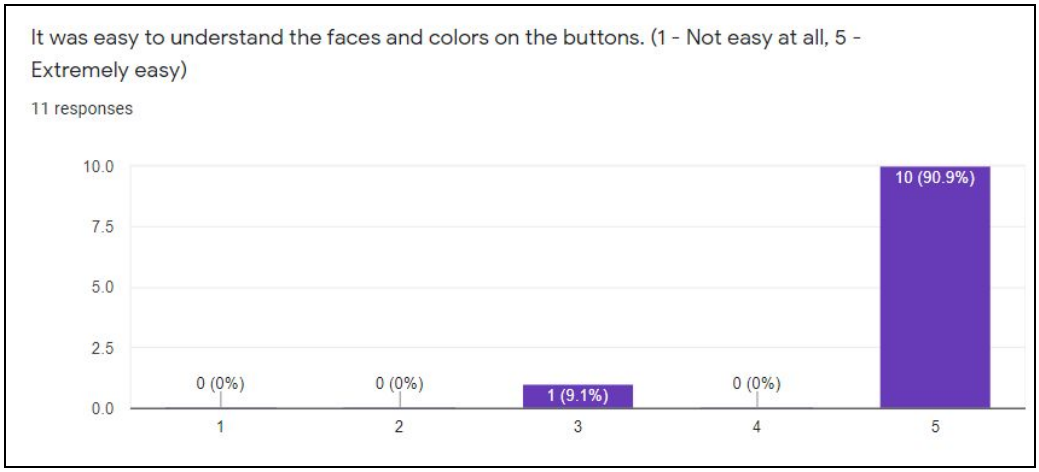
Task 3

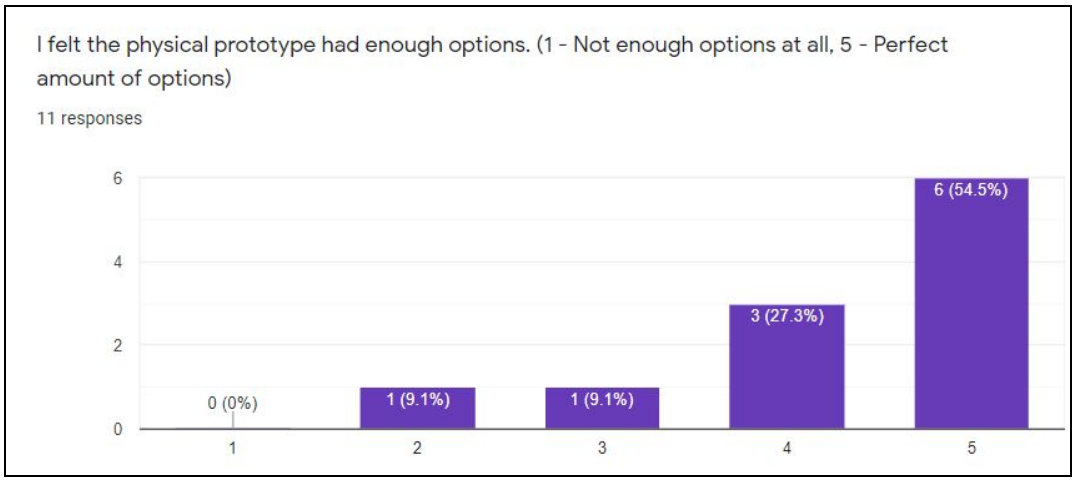
Trial Number	Bullying	Candy	Movies	Recess	Bad Grades	Divorce
1	1	4	4	4	2	2
2	1	4	4	4	2	1
3	2	3	4	4	2	1
4	1	4	4	4	1	1
5	1	4	4	3	2	1
6	1	4	3	4	2	1
7	1	4	3	4	1	1
8	1	4	4	3	1	1

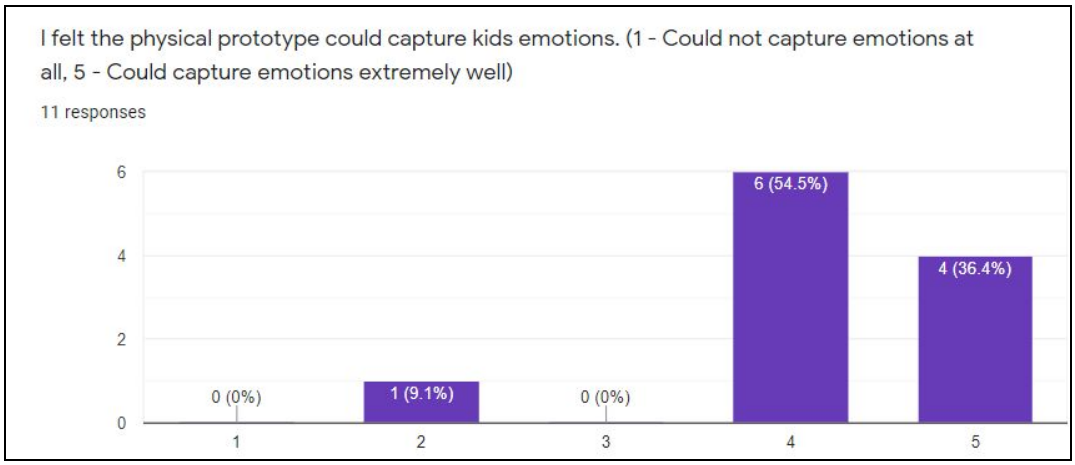
9	1	3	4	4	1	1
10	1	4	3	4	1	1
Average	**1.1**	**3.8**	**3.7**	**3.8**	**1.5**	**1.1**

Trial Number	**Poverty**	**Lost Jobs**	**Birthdays**	**Toys**	**Puzzles**	**Phone**
1	2	2	4	4	3	3
2	2	2	4	3	3	3
3	2	1	3	4	3	4
4	1	1	4	3	3	3
5	1	1	4	4	3	4
6	1	1	3	4	3	3
7	1	1	3	3	4	3
8	1	1	4	4	3	3
9	1	1	3	3	3	4
10	1	1	4	3	3	4
Average	**1.3**	**1.2**	**3.6**	**3.5**	**3.1**	**3.4**

Other than the objective data from the tasks, we also collected a few subjective data that were given by the participants throughout the testing and the post- testing survey. From that, we have succeeded in capturing the general trend of our prototype; such as whether our colorful emoticons is easily understandable or whether our emoticons options can fully represent children's emotion. We also gathered more insight toward the numbers of option we provided; whether we need more or less options for the button. Additionally, some of the participants were willing to tell use their struggle while doing the task and also give us their suggestions. All the results of the post-testing survey are attached below.







Q. What were your biggest pain points with the application?

- Just understanding initially what the app was representing and to who - staff vs individual student
- Charts
- Lagging when choosing dates

- I would reduce to 3 faces because the yellow face was barely used when i was doing the prototype
- Single range of emotions

Q. How could you improve the physical prototype

- More instructions on prototype
- More distinct emotions
- Maybe add one more emotion that is more neutral
- Maybe one more option for emotions not on that particular sad- happy scale
- You can give neutral face option

Mobile Prototype Study Results

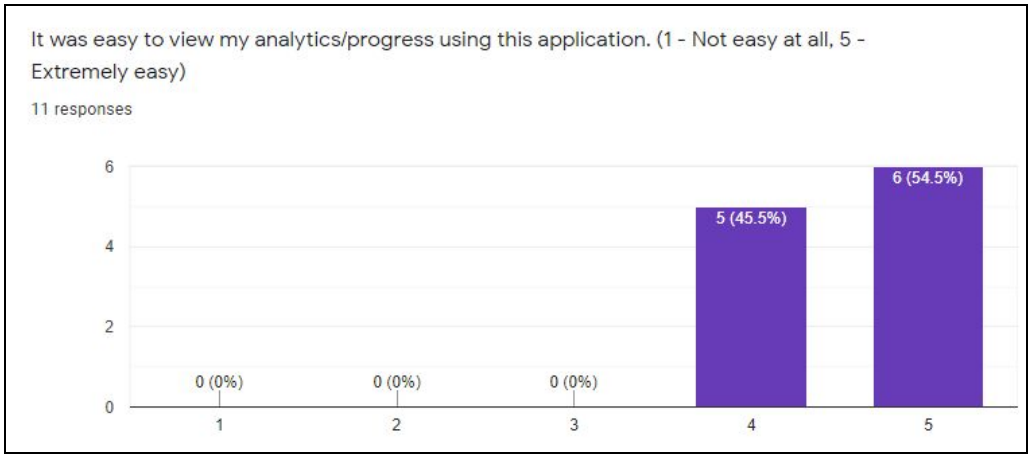
The results of the participant study focused on the mobile prototype is located below. The first table shows the time to complete (TTC)n and clicks to complete (CTC) for each of four tasks during each trial, as well as a final row calculating the average TTC* and CTC** for each task, while the second table shows the comments made by the study participant while the study was being conducted.

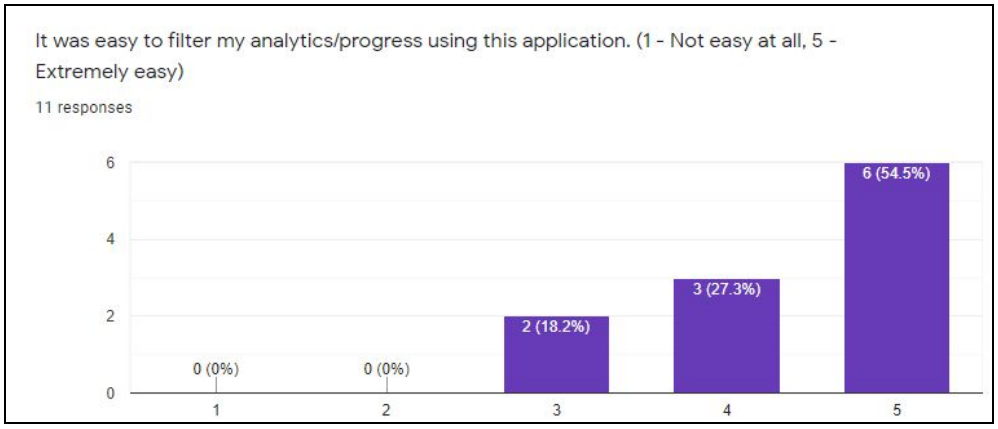
Trial Number	TTC Task 1	CTC Task 1	TTC Task 2	CTC Task 2	TTC Task 3	CTC Task 3	TTC Task 4	CTC Task 4
1	15	1	60	5	10	1	20	7
2	10	5	5	2	3	1	25	7
3	40	3	50	1	5	1	10	7
4	2	1	2	1	3	1	20	11
5	30	1	3.74	0	1.74	1	29.47	7
6	2.50	1	30.08	7	1.08	1	12.28	5
7	19.90	1	15.10	1	2.51	1	14.74	5
8	3	1	6	1	3	1	16	13
9	2	1	8	1	5	1	20	15
10	10	1	15	1	3	1	30	6
Average	**13.44**	**1.6**	**19.49**	**2**	**3.73**	**1**	**19.75**	**8.3**

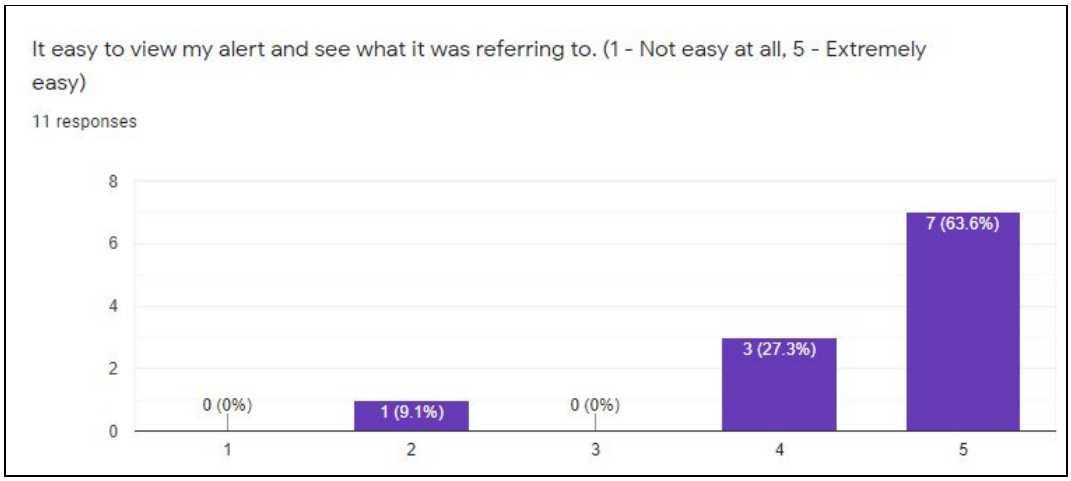
***(TTC = Time to Complete, all measurements recorded are in seconds)**
****(CTC) = Clicks to Complete**

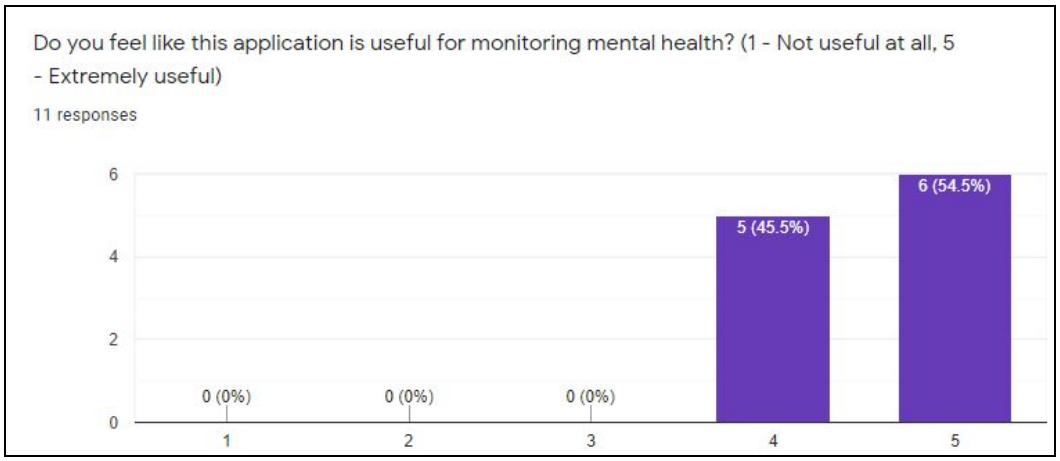
Task #	Comments
1	Notification Message is ignored*4, trying to click analytics page,aggregation about that day, “we did an activity that day”, aggregation about the day, red dots - sad very sad, threshold certain number of students, tagline get an alert
2	Did not look at dashboard, “Today’s Statistics” not standing out, looked at today’s stats, clicked on the sad face -> struggled to identify numbers and thought they were percentages.Clicked very sad, numbers too small
3	Clickable area is small, easily clicked, good symbol
4	Data range is not specific, “week” is unclear, did not see date range on first click*2, time, can see #1, #2 referencing the weeks, good work flow, date range too confusing, day of week option too confusing

In addition to our tasks and study, we also conducted a post-survey study participants could give us direct feedback on the app, and how they felt about the study as a whole. The responses to this survey are here:









Q. What were your biggest issues with the application?

- Sometimes what I was looking for wasn't super obvious, so I think a couple of areas needed clarification
- Charts
- Notification didn't necessarily show me anything other than the homepage it seemed. Once I knew what it was I understood though

- Clarity in bottom navigation menu. Hard to understand what/where the main page was
- Stuck a bit in the fourth task
- Analytics filtering is confusing

Q. Do you have any suggestions about how we can improve the application in any way?

- Make sure it's clear who the intended users are, and make the signifiers more noticable
- making the app more aesthetically pleasing
- Have notifications maybe go to different screen
- Maybe swipe between screens
- Make the filtering less confusing

Discussion of the results

Discussion of Physical Prototype Study Result

Based on the results from our usability testing in the previous section, we are able to make sure that our physical prototype is easily usable and understandable. By recording the completion time and the emoticons that our participant chose, we were able to ensure that our design of each button is easily understandable since every participant was able to distinguish each button correctly and choose the “correct” button in every task we have given them. Regardless, we noticed a slight difference in terms of which button that our participants chose. The reasoning of those differences will be discussed below.

From task 1, we got an average completion time of 8.808 seconds with the fastest completion of 5 seconds and the slowest of 20 seconds. We noticed that almost every participant was able to reorder the emoticons correctly in under 10 seconds. Moreover, based on the post-testing survey, more than 90 percent of our participants agreed that our physical prototype can has a high readability and can be understood easily without prior knowledge. However, we are aware that there is still a chance that someone could not distinguish the button easily, such as in one of the case where the participant needed 20 seconds to finish the task. It might happen if the user did not notice the usage of different colors and the difference between each emoticon. On the bigger scope, users that have a disability such as color blindness might not be able to instantly distinguish the buttons correctly; thus, making their input less valid. From that, we realized that although the colors help people distinguish each button, our emoticons usage should have more distinction, other than just a slight angle difference in the mouth for each button.

Task 2 and task 3 are meant to make sure whether our prototype has enough options to represent children’s emotion. Through the objective data we have got, if we consider the average number of below 2 is for “sad” and above 2 is for “happy” situation or words, we can say that we have succeeded in creating a prototype that has enough options for user to choose based on their feelings as each of the results match our intended answers. However, we realized that there is a slight difference in each of the users’ answers; this might be affected by the fact that different people might perceive each situation differently. The other possibility is when the users perceive our physical button differently as some of them told us that the third emoticon means neutral feeling while the rest told us that they consider it as a happy feeling. Thus, those differences in perception between each person might affect the accuracy of our data. For that, we need to improve our choice of button to make it more distinct to each other; by doing that we might be able to remove the second possibility of the problem we encountered. To further improve our physical prototype, we can improve our numbers of options as around 45 percent of our participants were not fully satisfied with the amount of button options we offered. Some of them

suggested that we need more options for the prototype to fully represent the emotions of children, while others suggested we use less options for simplicity.

Discussion of Mobile Prototype Study Results

The results of the participant study focused on the mobile prototype is located below. We stressed simple, replicable, and quick measurements in order to ensure that our testing had high readability. In simply measuring the number of clicks and the time it took to complete the tasks, we found that the basic features we have in the app are easy to use and show no struggles among the study participants, although we did get some feedback which inspired changes to be made for the final prototype.

The results showed that the time to complete task 1 averaged 13.44 seconds, with the longest of 40 seconds and a quickest completion of 2 seconds. While this average is under 30 seconds, we noticed, through comments given, that many of the study participants glanced at the benchmark notification which popped up, or did not read it at all, which may explain the quick completion time. The task had an average of 1.6 clicks to complete, which indicates to us that it is simple and easy to use. The participants did notice the aggregation of data. Perhaps something to look into for the final prototype would be to call more attention to the notification.

Task 2 averaged a completion rate of 19.49 seconds, with a quickest completion time of 2 seconds and the highest that of 60 seconds. Task 2 also had an average of 2 clicks to complete the task, which again is a good indicator that the app is easy to follow and use. This follows, as our second task is more complicated than the first, but we also received feedback that the dashboard and the today's statistics pages may not have stuck out or caught enough attention for the participant to notice them at first. In addition, we were told that the numbers were difficult to identify and that it was not clear whether they represented percentages or some other unit. This means that this feature, the analytics display, needs some work for the final prototype.

Task 3 had an average completion time of 3.73 seconds, with a quickest completion time of 1.08 seconds and a longest time to complete of 10 seconds. This task took all users only one click to complete, meaning the task is very easy to complete, which is also reflected in that study participants noted that the clickable area is small, but the symbol present is good, and it is very easy to click and follow.

FInally, task 4, our most complicated task, takes 8.3 clicks on average to complete (the high being 15, the low being 5), while the average time to complete is 19.75 seconds, with the fastest time being 10 seconds and the longest time to complete being 30 seconds. This part of our app does not follow the criteria we had set for ourselves, as although the task is complicated in comparison to the rest, it should not take more than 15 seconds or 5 clicks to complete. The feedback we received from participants for the duration of this task also made it clear that it was not the most easy to understand. Multiple participants said that they could not see the date

range or that it was confusing and not clear, in particular the week and day of the week functionalities.

The Implications of Our Findings

Mobile

With our mobile prototype, the requirements we proposed in Part 3 included monitoring mental health of the children at the BGCMA and providing an intuitive user interface. The purpose of testing our mobile prototype was to determine whether users can finish designed tasks without crucial confusion or misunderstanding. Having analyzed the data during the user testing, we observed that most participants were able to complete the tasks within reasonable times and clicks. For example, they finished each task in less than 20 seconds on average. Also, most of them were able to easily analyze the data presented on the screen and navigate through notification, menu icons, and filtering system. It implies that our mobile prototype successfully delivers usability and understandability.

Meanwhile, we observed some drawbacks of our mobile prototype on different screens. For instance, few participants did not understand the data representation quickly. They took relatively longer time to complete a few tasks due to misunderstanding of the information. In specific, our second task was to analyze the data shown on the dashboard page. However, we noticed few participants navigate to different menus to complete the task, leading to confusion. We understand there is definitely a room for improvement in our design for better usability.

In earlier phases of our research, we discussed our concerns with privacy issues in our solution space. One of our greatest concerns was to ensure privacy for the kids, making sure that each feedback data is unidentifiable from the staff side. Having considered such circumstances, our design prototype only provides overall statistics to the users. In addition, our main goal in terms of sustainability was to help the BGCMA community assess stress levels amongst their children, so that the organization can respond to any noticeable feedback from the children. Throughout our design of choice, the users were able to monitor and assess the stress level of the community.

Physical

Emotional Scale:

1 - very sad	**2 - sad**	**3 - happy**	**4 - very happy**

Based on Part 3 we wanted to evaluate two different tasks through three different activities for our physical prototype. We wanted to evaluate our usability goals and also understand the symbol depiction across different user groups. We wanted the scale to be intuitive and accessible to everyone. Analyzing the results of Task 1 for the physical prototype,

we can assume that the depiction and scale was clear and understandable. All user subjects were able to complete the task in under 10 seconds and all got the correct order. In general we can call the depiction of emotions a success and can assume that kids will be able to detect the gradient of emotions. Looking at the comments from task 1, we have noted that two users have depicted the red face as angry. However, our intention was for the emotion to connote very sad. So even though the ordering was correct we believe the red yielded a reaction of the strongest emotion. One thing we can look into doing is before implementing the system, sit down with the children and explain the emotions. Looking ahead of the results of those two users, they did not deviate from the average, so we believe the impact is marginal and does not interfere with the overall effectiveness of the system.

In Part 2, we outlined our design assessment as a combination of efficiency, user experience, accessibility, and intuitiveness. The first task of testing was used to test the intuitiveness, while the next two tasks were used to test the other aspects of our design assessment. Our second task, "Choosing an Emotion", was used to help us evaluate the consistency of users pressing emotions to see if the data would be consistent. We had to simulate adults acting as children, and therefore there is bias involved because adults are more emotionally mature and cannot fully emulate a child's emotional state. When we design the scenarios we had written scenarios with a specific emotion in mind. Scenario 1, where Timmy drops his sandwich, we had designed for the emotion to be a 2, because Timmy is sad, but this is a one off experience where emotion 1 is reserved for persistent sadness. Our average was 1.7, which means all users were on the right side of the scale, but 3 users felt this yielded an emotional response of a 1. Since all of the users responded on the same emotional side of the scale, we can assume that the user experience is going to be consistent for children. The other three scenarios followed similar patterns. We wanted scenario 2 to yield an emotional response of 4 and the average was 4. We wanted scenario 3 to yield a response of 3 and it yielded a 3.5. And lastly, we wanted the last scenario to yield a response of a 1, but the data averaged at 1.7. Since scenarios 2 and 3 followed a similar pattern to scenario 1, we can assume the same reactions and deductions. However, scenario 4 got the correct side of the scale but leaned heavily to emotional response of 2 instead of 1. Scenario 4 was the following:

Over the past couple of months, Sally has received some terrible news. Her dad recently lost his job and can no longer afford Sally's dance lessons. Sally's joyous personality has slowly started to become more melancholy. Sally has started to isolate herself and is spending more of her time at recess alonel. Today, when boarding the bus, she sits in the second row and looks down at her feet until she gets to the Club. How is Sally feeling upon her arrival today?

We believed this was a designed scenario that would yield a 1, because of the persistent upsetting observations taking place. We believed her father losing her job and her inability to take dance classes anymore would make Sally very sad. However, 5 users did not answer with a 1: three users answered with a 2 and two users answered with a 3. In task 3 however we asked for immediate responses for words. One of the words we tests was "lost jobs". What we observed is that all 5 users that did not respond with a 1 for Sally's scenario, responded with a 1 when asked what emotion "lost jobs" specifically yields. Therefore, there is still some inconsistency in the

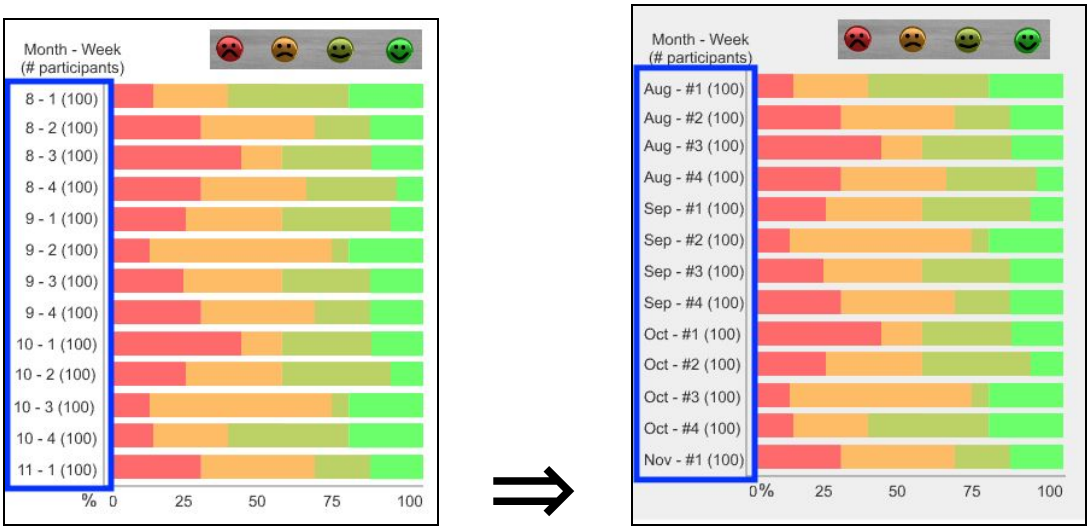
testing, and requires further analysis. What these inconsistencies demonstrate is the usefulness of the product is not fully developed. To be expected, people process experiences with different emotions. Therefore, the system might produce some noise for the sad vs very sad. What is promising is having only four emotions, forcing users to lean happy or sad, and that seemed to work in task 3.

Changes In UI & Future Directions

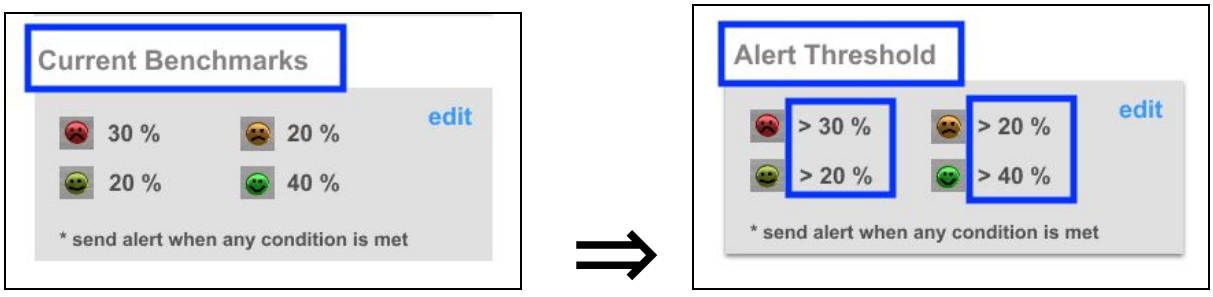
Mobile

Description of UI changes

We have made small changes in our UI concerning important feedbacks we received during the demo day. First, we changed the display format of the time duration in our analytic view as shown below (marked with a blue rectangle).



As another minor change, we changed the word “Current Benchmark” to “Alert Threshold”. The word represents the alert threshold in which the application sends a notification to the BGCMA staff when a certain level is met. Our participants during the first demo day mentioned that the word does not deliver the meaning of the purpose.



We have made two changes to the feedbacks, and our team restarted to have a new set of ten participants in our usability testing after the changes.

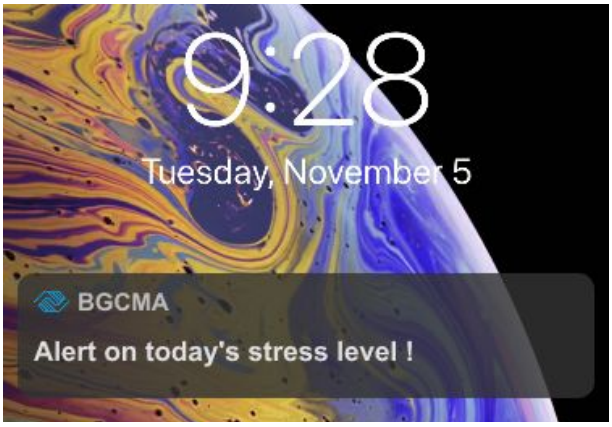
Future Additions

- Heuristic Evaluation by human-computer interaction experts
- Tasks to experiments
    - Graph types
    - Different messages

There are several changes we would implement concerning the materials we learned after the completion of our evaluation. First of all, we could have integrated the concept of heuristic evaluation. With help from different human-computer interaction experts, we would evaluate our system in a more accurate and advanced manner. We could assess whether our prototype provides appropriate languages, delivers a fluent workflow between different views, and alerts the user with valid error messages.

Secondly, we could have changed a few tasks into experiments. By having experiments in addition to the standard usability testing, we would have discovered new knowledge on the user interface that delivers better usability. As an experiment, we would have different types of graphs to display the feedback data. In the current system, we have a bar-type graph as shown below (Current Bar-Type Graph).

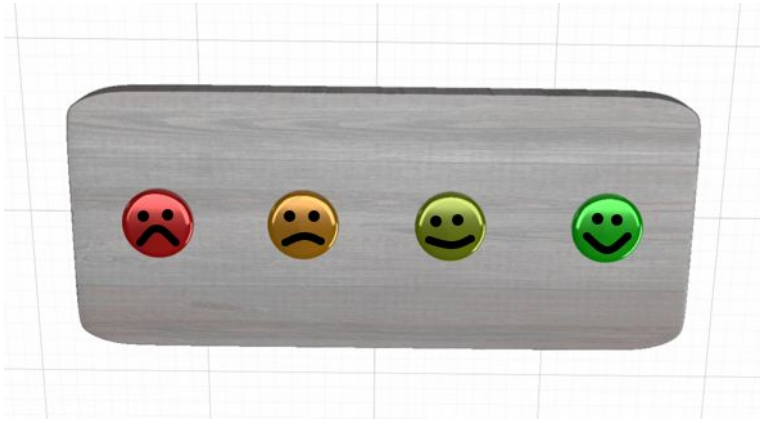
Current Bar-Type Graph



Current Notification on Mobile Device

This type of graph is good for comparing different blocks in each bar. However, we would want to experiment if different types of graphs deliver a better understanding. Another factor we would include in the experiment is an amount of information included in our notification. The current notification only includes the message, “Alert on today’s stress level!”. The purpose of the notification is to alert the BGCMA staff that the number of children who pressed a certain button exceeded the threshold. As an experiment, we would test if including more or less information helps users to comprehend the alert faster.

Physical



Description of UI changes

During testing we did not make any changes to the UI itself, but we did make changes to the protocol after our demo with our TA and one other test subject. First, we had originally designed our script to allow users to be tested starting with either the physical or mobile application tasks. We did this to allow us to optimize our time with user subjects. However, we realized if users start testing on the mobile application they would be exposed to the scale of emotions and therefore it would skew Task 1 of the physical prototype, which is to order the emotions to see if our scale is consistent among users.

In our script:

Original: **Final Survey**

If the user has completed the demonstration of the physical device already, please read 1, if not read 2.

1. [team member] - "Now that you have completed both of these demonstrations, please fill out this post-survey giving us your final thoughts."
2. [team member] - "Now that you have completed the mobile app segment, please continue with us to the next demonstration."

Modified: **Final Survey**

3. [team member] - "Now that you have completed both of these demonstrations, please fill out this post-survey giving us your final thoughts."

Also during our demo, we were noted by the test subjects that for task 1 the cutouts of the faces were hard to see and needed to be blown up. Therefore, before testing on new users, we blew up the sizes of the cutouts by about 4x. This allows us to get a better understanding of the speed the ordering can happen in. We believed this is linked to the quickness we saw in task 1 when testing our physical prototype.

Our last task was designed for quick responses to a variety of words. While running our demo, we realized that our original order was grouped poorly with a lot of the positive words being at the beginning and a lot of the negative words being at the end. We decided to change up

the order for our final protocol. This way we can mix up the emotions and could give us a more truthful response, since we read quick responses and are often decided on predisposed notions. Therefore, if we gave you 3 very negative words, no matter the connotation of the next word you would have a more negative attitude.

Original Order: bullying, candy, movies, recess, bad grades, divorce, poverty, lost jobs, birthdays, toys, puzzles, phones

Modified Order: bullying, candy, movies, bad grades, recess, divorce, toys, poverty, puzzles, birthdays, lost jobs, phones.

Description of changes that you would make if you had more time or other information

With more time we could create a much more extensive testing policy. One major change or addition we would implement is to create a prototype that allowed for a clickable interaction. We did not have time to build a 3-D model, therefore, the users interacted with the conceptual model on a piece of paper. If we presented users with actual live feedback, they could give us reactions with the entire system.

Another change we would make would be the user group we used to test on. The majority of our test subjects were other students in our class. This was due to the convenience of our sample. During our class demo, we were able to get through 7 different users and therefore did not need many other users for this project. And due to the time constraint, team members tested friends and peers so we could have the results in a short amount of time.

We also only collected data through user studies. We controlled the environment and were able to seek information and feedback based on what we wanted to know. With more time we would look into creating a field study at the BGCMA. This would also give us data about how children interact with the system, which due to the lack of background checks we could not conduct research on. Also with more time we would follow the entire heuristic evaluation. We were able to gather the information and evaluate the results, but we have a time constraint. Organizing, debriefing, and assigning severity to the issues does not fit in our time window. These steps would be vital when productizing the interface to allow for the system to best suit our design assessment.

Also with more time we would conduct the same tasks on the same subjects to establish a scale. This would allow us to see if users react to the same experiences and words in the same way to help us make the system more consistent. This is important in gathering effective data that the club can use and start to integrate into their everyday system and classrooms.

Reflections

Team Happy's Semester Reflection

Throughout this project, as well as the class itself, we gained an incredible amount of insight about the design process, UI design, and design in general. However, more than other principles or concepts, designing our system most expressed to us how design is truly an iterative process. In terms of the obvious example embedded into our project, the division of our design into four separate portions, or reports, we had clear guidelines that made our application improve iteratively. However, more than this, we see the iterative nature of design present in the work done between each of these aforementioned reports. Our team went through a plethora of different ideation changes, including the many shifts in our initial goals, prototypes, tasks, functionalities, and more. From the numerous lectures, interactions with the BGCMA, and Norman readings, we found so many different reasons to fine-tune and change our direction with the system we have designed.

Furthermore, the entire experience showed us the intricate dynamics of working in a team composed of completely random individuals, longterm, all with a single, deliberate, and planned goal. As Dr. Arriaga stated at the beginning of our project, the random nature of our team designations was to give us the experience a real design team within a company would have -- "You don't choose your employees!" Consequently, we were thrown for a loop when the first few weeks and the majority of Part 0 and Part 1 were spent learning how one another functions. After we were able to get into the flow of acknowledging, respecting, and adjusting to one another's strengths and weaknesses, our team finally flourished.

One of our biggest issues with our system we designed was the issue with ensuring the button presses were valid. While we have, unfortunately, not solved the issue completely, we have worked extensively to limit the possibilities of the inaccuracies in the reported button presses. Our primary methods to combat inaccuracies (mainly from buttons being pressed multiple times by the children), was through the placement of the button. The system is designed for the button to be placed in the entrance of the BGCMA. This placement affords us two benefits: first, the children will be asked to press it as they check in to the BGCMA, giving us the ability to record a press of the button system from every child entering the BGCMA. Second, by placing the button at the check in area of the BGCMA, BGCMA staff will be able to monitor children attempting to make inappropriate presses and intervene. Although the placement of the button mitigates most of the worries with the button's validity, we also plan to implement a cooldown timer. This timer would not allow the buttons to be pressed after having just been pressed for a short time -- most likely a five second timer. Such cooldown timers have shown their efficacy in many systems, both physical and electronic alike. All in all, while still a large issue, we believe the aforementioned approaches help deter most problems with the validity of the button presses.

In regards to Part 4 itself, we do have some qualms with the way we assembled the final report presented to you here. The many setbacks, obstacles, and problems we faced in creating our evaluations were eventually ironed out, but more became apparent in attempting to finalize,

analyze, and learn from our collected data. The biggest issue of all was finding a way to organize, digitize, and then systematically go through our evaluation forms. If we were able to start again, there is no doubt that we would immediately record all data electronically. Since we used paper evaluation forms for convenience, we ended up having to painstakingly scan, upload, and digitize the results. This process, especially the fact that we went about it in different ways, meant we had to restart many parts of our report several times before getting it to the point presented to you today.

Moreover, had we had more time, we would have recreated our evaluations electronically to showcase the improvements we could have created with a better designed system. However, the most beneficial reason for this electronic conversion would have been to re-record our data with different participants to not only save us more time (we would no longer have to digitize the results), but also allow us to eliminate any errors we may have made during the digitizing process. Nevertheless, the many troubles faced along the way to completing our P4 only made it stronger at the end, and in doing so, led us to the report we are proud to present here.

Individual Reflections

Jackie

- This semester has taught me a lot not just about creating a full and thorough design assessment, but also working on a team and how I conduct myself. Dr. Arriaga highlighted the importance of having a holistic team, and during the whole semester I felt our team came together and worked well. There were times when certain team members did more work on one part, however, the next part those same people stepped up to take more work. As far as a design approach, I learned the importance of iteration. At the beginning of the semester, I believe I was very one dimensional in my thinking, but was quickly proven wrong by the direction our project wanted. For example, going into the poster session, I believed the physical prototype was the option for the goals we set out to do, however, every person gravitated towards that system, which forced me to relook at my own evaluation. I realized that I was focused on the wrong parts of the system and was thinking what I would enjoy the most, and was forced to go into the steps of the client. I am very excited and proud of the work our team completed this semester.

SangYun

- The project during the semester gave me valuable lessons. In addition to the course material, our team was given the opportunity to solve a real world problem by applying the four steps: requirement gathering, alternative designs, prototype, and evaluation. In each step, our team successfully went through the details and requirements. During our evaluation phase, we wish we could have known more about experiments. If we knew, we could have modified our plan to apply multiple variables, such as different colors on the buttons and types of graphs on the analytics page. Also, having considered the lessons we learned from Human Capability lectures, we might have planned to incorporate pressure level on the buttons. Yet, I am very satisfied with the process we followed along each report with my team, and I believe we all learned valuable lessons in terms of human-computer interaction.

Sanjana

- I believe this project has changed not only the way I think about projects, but the way I work in groups as well. I feel that at the beginning of the class I was a little confused about the scope of the project and it made me feel like I could not contribute as much to the team, but as things got going I felt that my team members were very helpful in explaining what was going on and it allowed me to contribute much more to this project. I've realized the importance of communication and making sure everyone is on the same page when moving forward because of this class, and I enjoyed working with this team for the course of the semester. In terms of design, I think the most important thing I have learned is to approach things with an iterative thought process. I believe that going through the project and narrowing our scope as well as our functionalities after receiving feedback on our prototype helped bring our project together in an amazing way. To expand, we moved from the concept of having a physical submission box and a single button to press to indicate stress to having a panel and submission box, as well as a website or app when we presented R2. Upon receiving feedback on those prototypes, we made changes to our physical and mobile prototypes that have received excellent feedback in the studies we did for R4. I am extremely proud of how our group's final prototype has come out and the work we did.

William

- Based on our work toward this project, I realized that through brainstorming our different ideas toward a certain problem, we can bring up the best solution that we would not have thought before. I also learnt the importance of user's feedback and how it really helps us create an intuitive and usable interface. From that, if I could start this project over again, I would like to get more feedback from different users and get more insight towards the possible problems and solutions for our prototypes.

Nouzhan

- By the end of this project, I take a look at things I create and work on in my life completely differently. I approach everything I do *iteratively* now, and start projects knowing they are not initially in their best and final forms, and that I now have the capabilities to get them there. Without too much conjecture, I can genuinely say that the concepts we learned this semester have changed my outlook and approach to just about everything.