# Evaluation Paper 2

# HCI & IV 2022

## < Learning Platform for Sign Language >
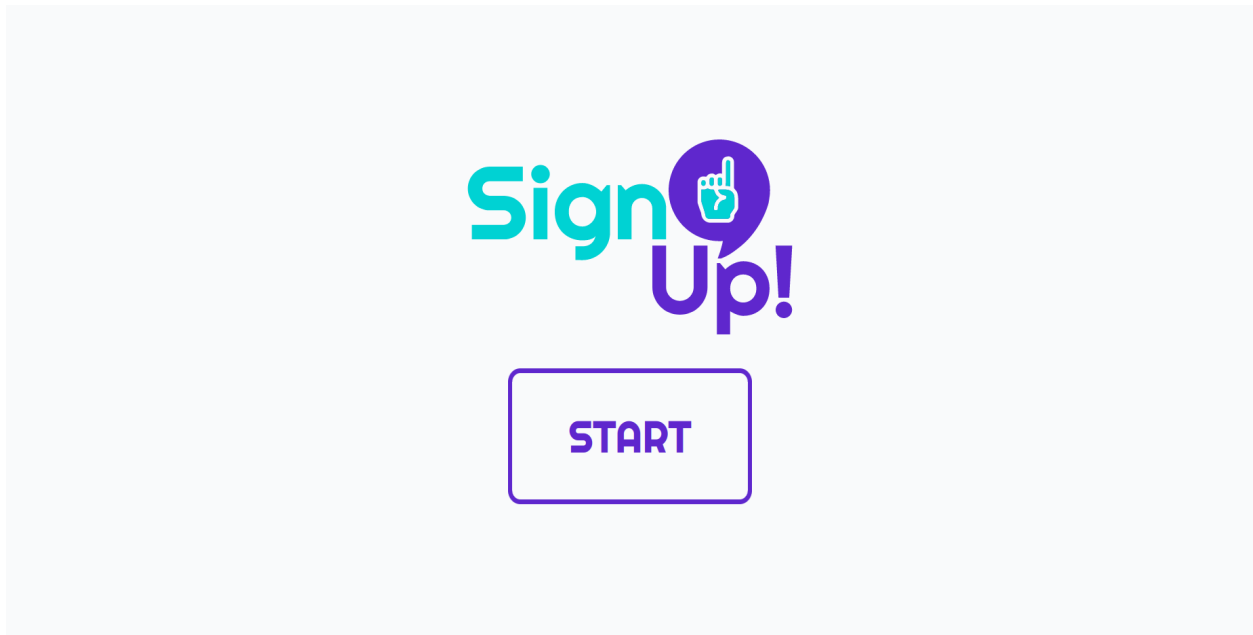
Team 31: Nayoung Jung & Yvo Hu

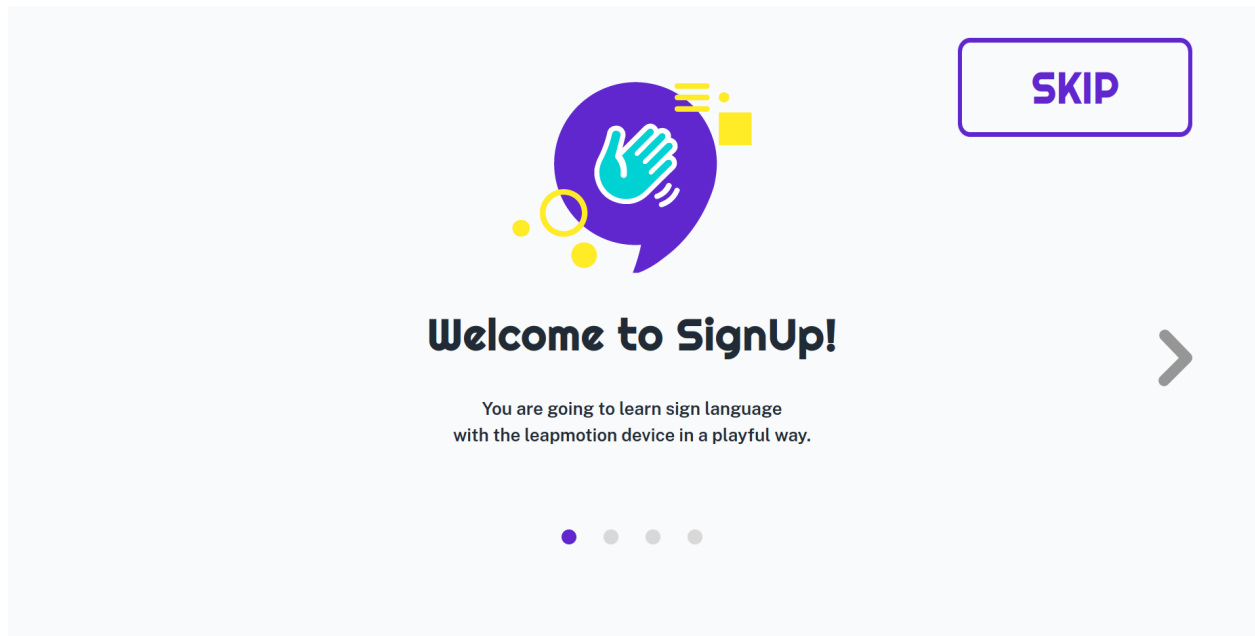Theme D New Interaction

## 1. Introduction

Our research question for this project is 'How can a leap motion be used to enhance the learning experience for sign language?'.

During the first evaluation phase, we could see how users interact with our initial prototype and how external factors -the prototype and the use of the leap motion device- can influence user experience. Based on our first valuable outcome, we expanded our second evaluation in another direction. Unlike the first attempt, we focused on the effectiveness of our content so that our collected data can be more relevant to answer our research question. There were some different setups we came up with for the second usability test, and those will be introduced in this article. The second evaluation was pretty similar to the first evaluation, but some conditions have been changed. The tests were held with five participants in one of the campus buildings. The test was held from December 19 to December 20, 2022.
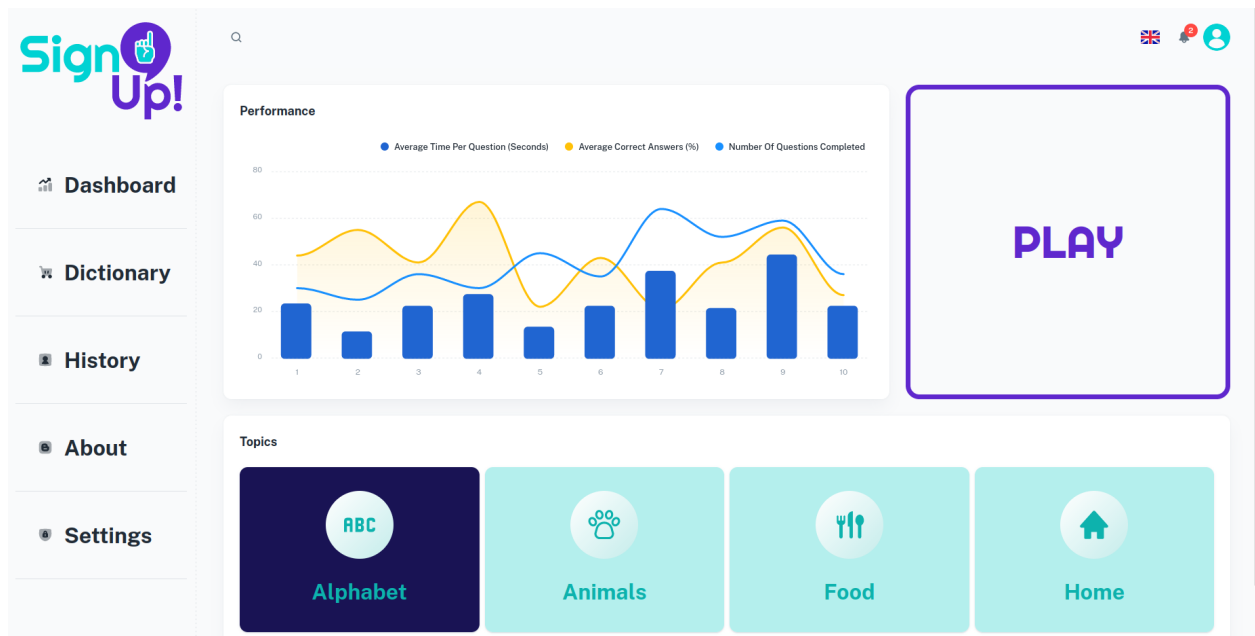
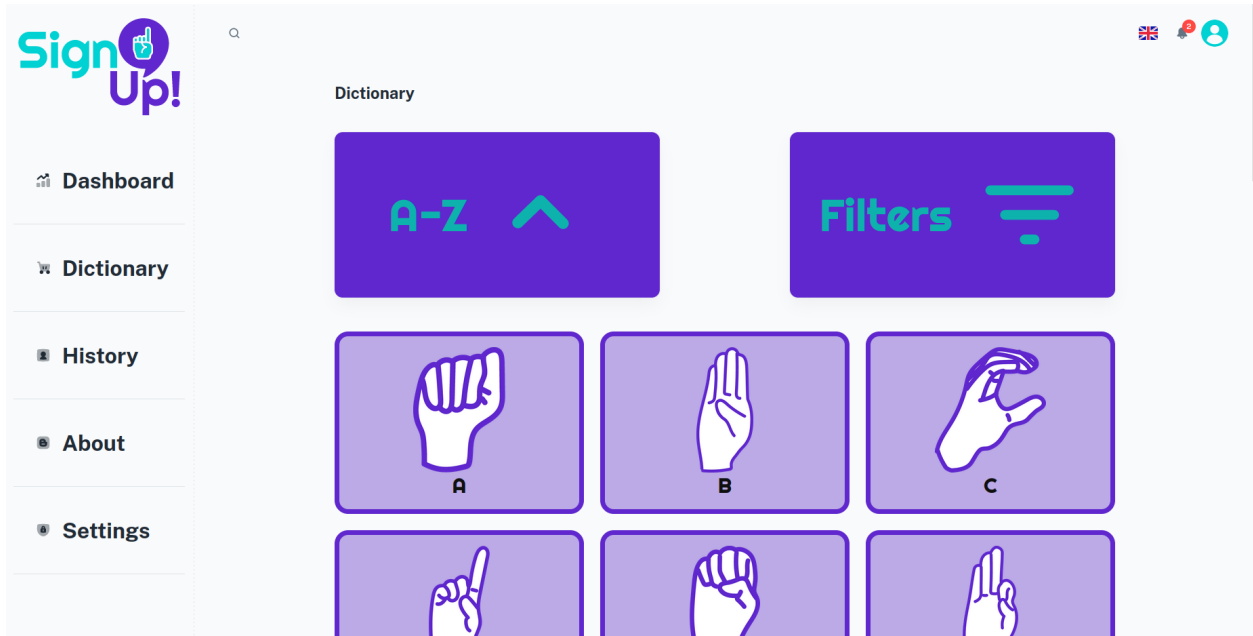The following are the screenshots of our second prototype.



< Splash screen >

# Welcome to SignUp!

You are going to learn sign language
with the leapmotion device in a playful way.

< Intro/Tutorials >

## Sign Up!

- 📊 **Dashboard**
- 📖 **Dictionary**
- 📋 **History**
- ℹ️ **About**
- ⚙️ **Settings**

**Performance**

- ● Average Time Per Question (Seconds)
- ● Average Correct Answers (%)
- ● Number Of Questions Completed

**PLAY**

**Topics**

| Alphabet | Animals | Food | Home |

< Dashboard >

< Dictionary >



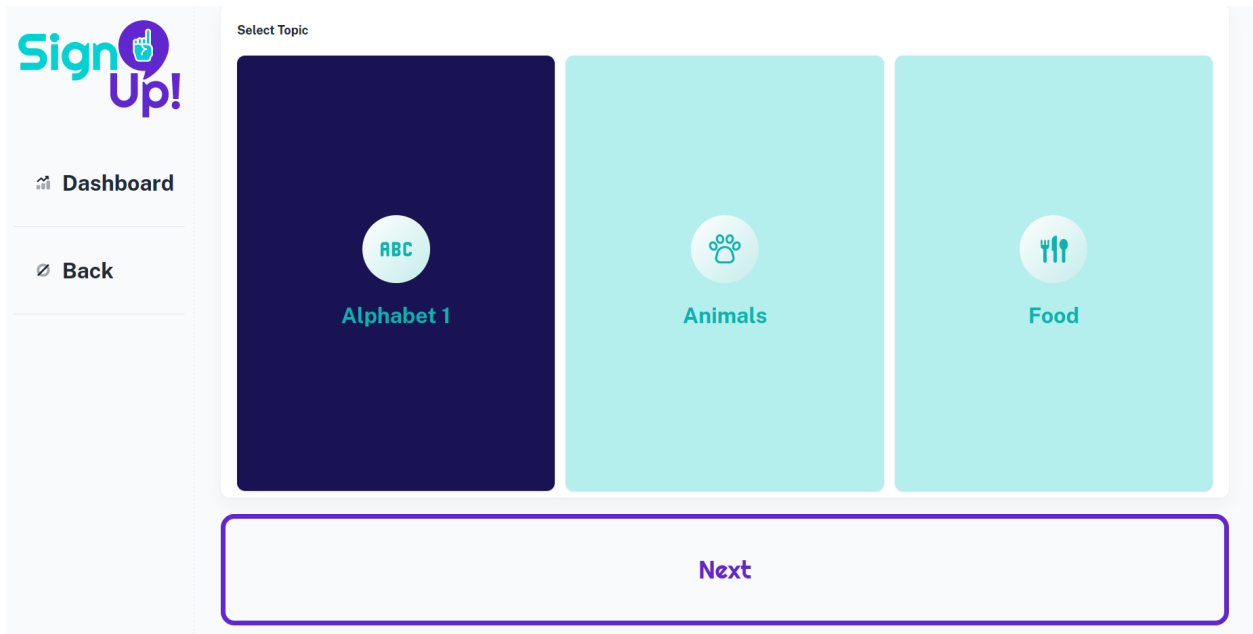< Select Topic screen (after pressing Play) >

< Game mode: Matching >
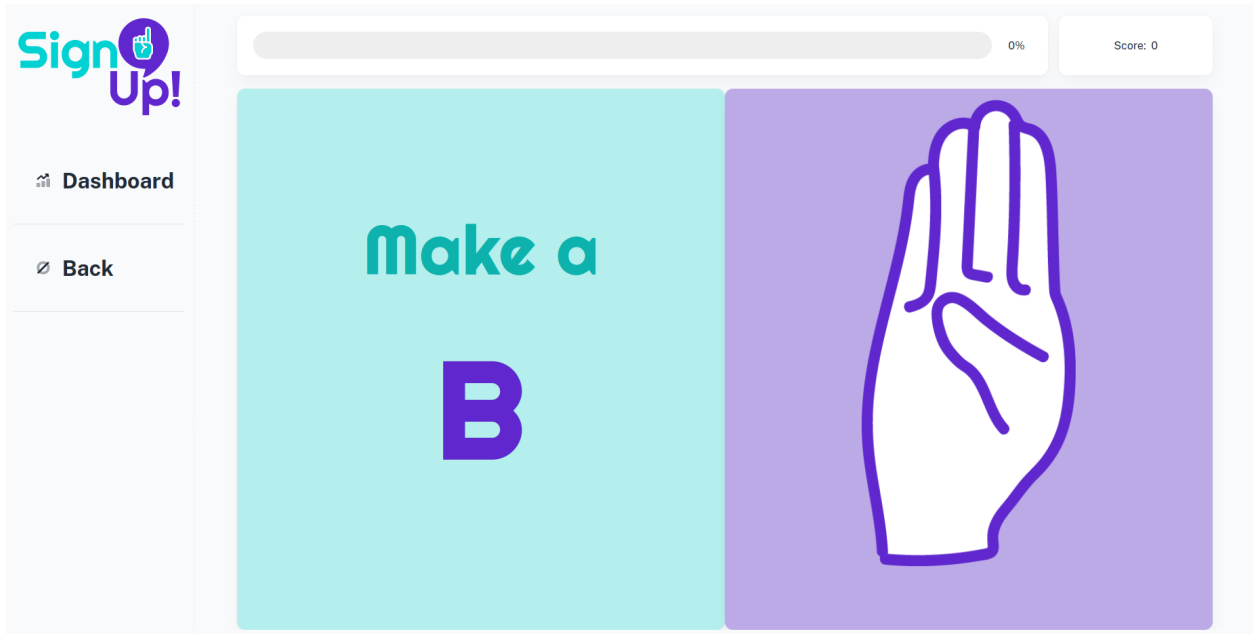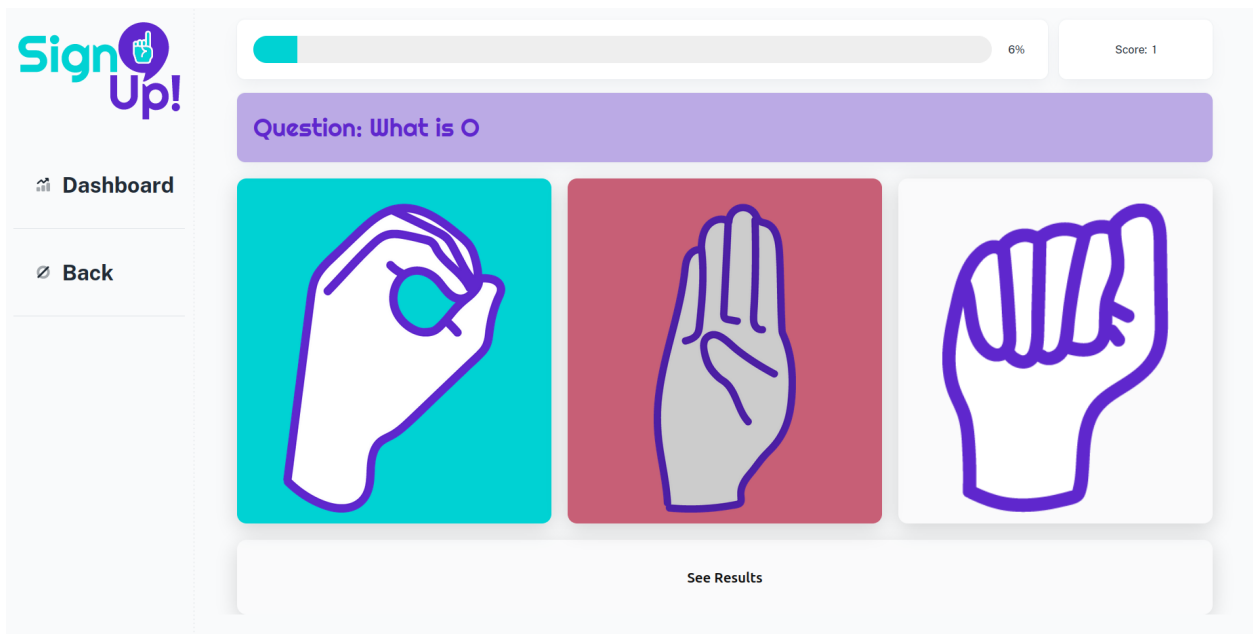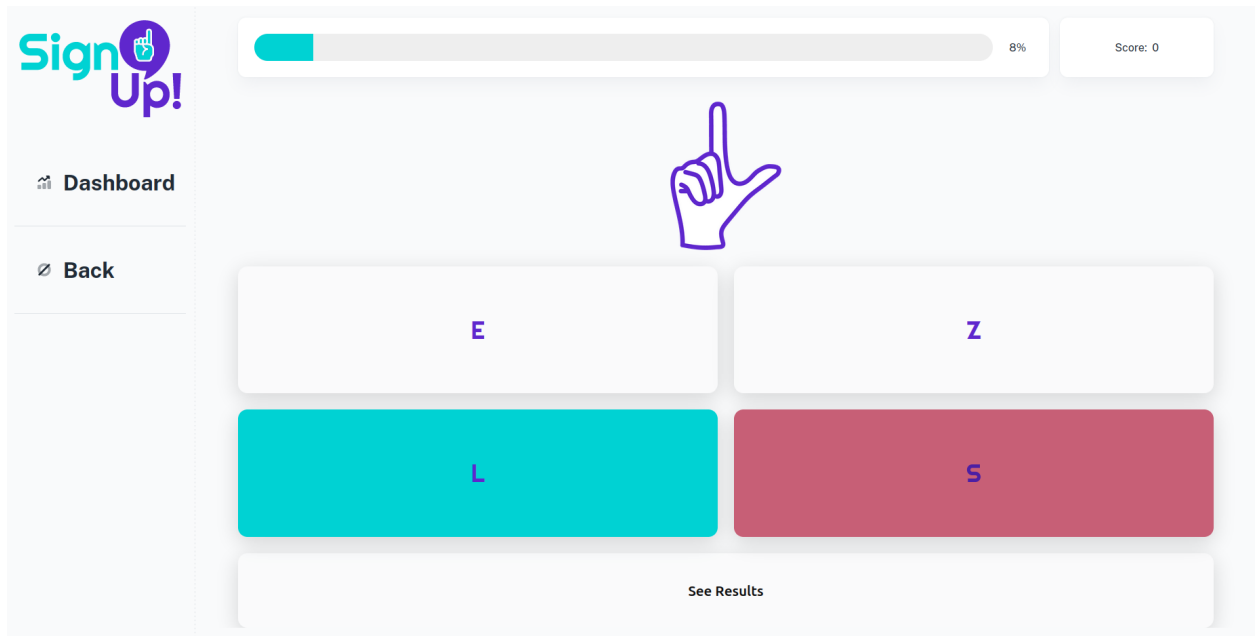


< Game mode: Recognition >

< Game mode: Translation >

2. **Instructions: What tasks did you give your users?**
From the previous test, we noticed that our prototype has room for improvement leading to a better user experience both in learning and in interacting with the device. The following list shows the main updates of our prototype.

- *Splash screen*: Previously, the layout was not done yet, and it gave an unprofessional impression to the participants. However, after the first test, we finalised the layout, which contains more information and formulates an impactful impression at the beginning of the user journey.

- *Colours to show a different state(UI design)*: The buttons in our prototype are designed to show the transition of two states- default and hover/click- by changing colours. In the previous design, the buttons shifted from purple/turquoise to dark purple. However, we noticed that this colour change does not make sense and does not seem to be natural, especially in the case of turquoise to dark purple. Therefore, we revised the colour scheme for the transition effect- from turquoise to dark navy.

- *Game screen*: During the previous user testing, we got feedback that it is difficult for them to read the text in the game. Since the screen shows the alphabet, the gesture to imitate and the current input, which are a lot of information to consume simultaneously, the layout of the game scene had to be changed to enhance readability and be more intuitive.

- *Leap Motion orientation*: During the presentation, one of the teaching assistants mentioned it might be better to change the orientation of the leap motion device to better accommodate how a user might naturally orient their hand in order to sign. We reflected on this comment and changed the position of the device accordingly.

The users were assigned 10 tasks, and those covered the same directions as the previous evaluation- 'navigation/the layout', 'user interaction with the device' and 'learning'. The guidelines below show the tasks for the second evaluation. The rest of the document was left unchanged.

1. First, try out the leap motion device and get used to its controls.
2. Navigate the application by clicking on a button.
3. Try to scroll up and down.
4. Find the 'Dictionary' & explore this feature.
5. Now return to the main dashboard
6. Find the play button, and click on it
7. Select the topic and confirm your selection.
8. Do the same for game mode
9. Play game
10. Return to the dashboard

The main difference compared to the first evaluation, is that the order of the tasks has been changed. Previously, the users had to explore the feature 'Dictionary' at the end of the test (task no.10). However, the current order asked them to explore it after the first encounter with the app (task no. 4). Through the update in the order, we aimed to make the users feel more familiar with the interface and ASL. Another reason behind this change is derived from feedback from our users, mentioning that *'finishing the test with just exploring did not make me feel that I'm done with it. It felt like there's more afterwards'*. Based on this quote, we formulated a clear starting point and a clear endpoint for the entire flow of the second evolution.

The following shows the questions we used for the first and second evaluations to collect more direct, unfiltered data from our participants.

1. How was the whole experience for you?
2. What did you like about this application? Why?
3. When did you feel '*ah, this could have been better*' while using this app? Why?
4. What do you think about the game mode? Why? What would make you feel more challenged/excited in this game mode?
5. What other things would you like to see more on this app?
6. Do you remember how to make your name in sign language?

As an extended approach, just like our previous usability testing, we asked the participants to fill out the questionnaire on Google Forms. The questionnaire (appendix b) is based on the 'System Usability Scale Questionnaire', which covers the effectiveness and efficiency of the system and user satisfaction. To define the success of this study, we set an ideal score for each question as follows. For the questions with a positive context, the final average values should get a higher score than the ideal score. On the other hand, the questions containing negative factors should get lower than the ideal score. These scores are based on our collected insights from the entire design process as well as the previous evaluation.

| Question | Ideal score |
|---|---|
| 1. I think that I'd like to use this system frequently | 4 |
| 2. I found the system unnecessarily complex | 2 |
| 3. I thought the system was easy to use | 4 |
| 4. I think that I'd need the support of a technical person to be able to use this system | 2 |
| 5. I found the various functions in this system were well integrated | 4 |
| 6. I thought there was too much inconsistency in this system | 1 |
| 7. I would imagine that most people would learn to use this system very quickly | 3 |
| 8. I found the system very cumbersome to use | 3 |
| 9. I felt very confident using this system | 4 |
| 10. I needed to learn a lot of things before I could get going with this system | 2 |

### 3. Test Setup

During the usability testing, we observed how users interacted with our application and measured how long it takes for each user to conduct a specific task by using the stopwatch feature on our smartphone. We have omitted some tasks because we have currently only been able to set up the application on one computer and because we examine the SUS questionnaire in a different section.

Usability specification table

| Task | Issue | Measured | Current | Worst | Planned | Best |
|---|---|---|---|---|---|---|
| Installation | Benchmark #1 | Time until successful launch of the application | TBD | 30s | 10s | 5s |
| Initial performance | Test out hand tracking | Time until user felt comfortable exiting tutorial | 104s | 180s | 120s | 120s |
| Initial performance | Launch game | Time until the launch of a valid game | 30s | 30s | 30s | 15s |
| Initial performance | Find a gesture in the dictionary | Time until gesture found | 30s | 30s | 10s | 5s |
| Learnability | Correct rate | Correct rate per game (out of 50 questions) | 42.8% | <= 25% | > 25% | 0 |
| Throughput | Completion of a gesture | Average amount of time spent to perform a correct gesture | 9.6s | 10s | 5s | 2s |
| Initial impression | Questionnaire | Average score [1-5] | TBD | 1 | 4 | 5 |

*Not included* (first row)

*Not included* (last row)

The previous table points out the issues of the tasks with the term 'Benchmark #'. The details of each 'Benchmark #' are shown below.

## Benchmark table

| Benchmark | Description |
| --- | --- |
| #1 | Run the script start.sh to start the application |
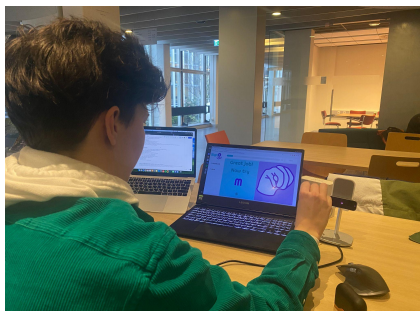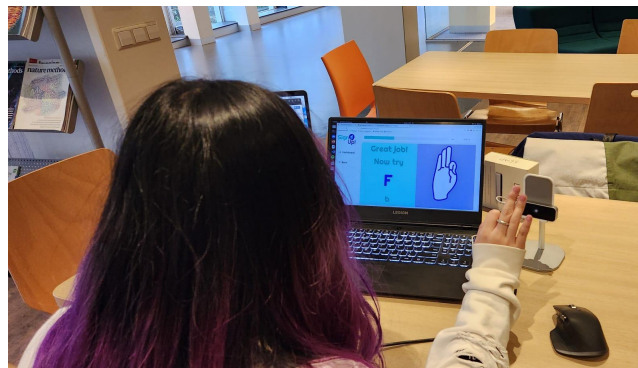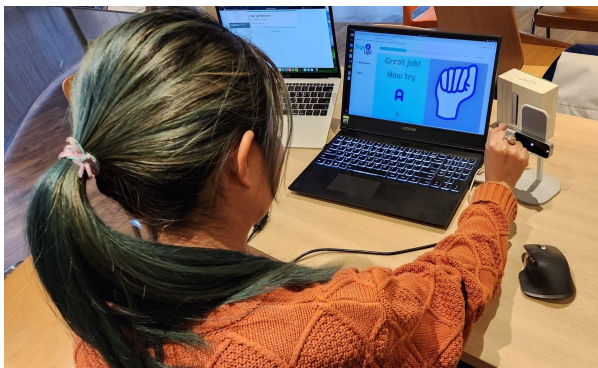


< Testing Scenes >

## 4. Results

The following is the data we collected from the second evaluation. The first figure is the collected data from each participant, and the second figure shows the comparison between the average user performance and the planned or the best value we mentioned in the usability specification table.

**Subject 1**
Gender: Male
Age: 22

| | |
|---|---|
| Time to test out hand tracking | 100 s |
| Time to pick game from dashboard | 40 s |
| Percentage correct quiz | 38% |
| Time to scroll to the bottom of the page in dictionary | 30 s |
| Time to perform 50 gestures | 410 s |

Notes:
The instructions weren't very consistent

**Subject 2**
Gender: Female
Age: 22

| | |
|---|---|
| Time to test out hand tracking | 120 s |
| Time to pick game from dashboard | 30 s |
| Percentage correct quiz | 46% |
| Time to scroll to the bottom of the page in dictionary | 25 s |
| Time to perform 50 gestures | 530 s |

Notes:

**Subject 3**
Gender: Female
Age: 23

| | |
|---|---|
| Time to test out hand tracking | 90 s |
| Time to pick game from dashboard | 40 s |
| Percentage correct quiz | 42% |
| Time to scroll to the bottom of the page in dictionary | 35 s |
| Time to perform 50 gestures | 490 s |

Notes:

**Subject 4,**
Gender: Male
Age: 21

| | |
|---|---|
| Time to test out hand tracking | 100 s |
| Time to pick game from dashboard | 30 s |
| Percentage correct quiz | 40% |
| Time to scroll to the bottom of the page in dictionary | 30 s |
| Time to perform 50 gestures | 450 s |

Notes:

**Subject 5,**
Gender: Female
Age: 23

| | |
|---|---|
| Time to test out hand tracking | 110 s |
| Time to pick game from dashboard | 40 s |
| Percentage correct quiz | 48% |
| Time to scroll to the bottom of the page in dictionary | 30 s |
| Time to perform 50 gestures | 520 s |

Notes:

The following is an overview of the collected data from the SUS questionnaire sent to the five participants after the user testing. The detailed data is in appendix c.

| Question<br>Scale: 1 (Disagree) - 5 (Agree) | Ideal score | Average (#1) | Average (#2) |
|---|---|---|---|
| 1. I think that I'd like to use this system frequently | 4 | 2.4 | 2.4 |
| 2. I found the system unnecessarily complex | 2 | 1.4 | 1.4 |
| 3. I thought the system was easy to use | 4 | 3.4 | 3.2 |
| 4. I think that I'd need the support of a technical person to be able to use this system | 2 | 2 | 2.4 |
| 5. I found the various functions in this system were well integrated | 4 | 4 | 5 |
| 6. I thought there was too much inconsistency in this system | 1 | 1 | 1.6 |
| 7. I would imagine that most people would learn to use this system very quickly | 3 | 2.8 | 3.6 |
| 8. I found the system very cumbersome to use | 3 | 2.2 | 2.6 |
| 9. I felt very confident using this system | 4 | 2.6 | 2.4 |
| 10. I needed to learn a lot of things before I could get going with this system | 2 | 1.6 | 1.2 |

The average values for each question were converted into textual feedback as follows for better understanding.

1. Users are still uncertain about using the app frequently.

2. **Success:** The system is far from complex.

3. The system is easy to use although the system is a bit challenging yet.

4. Sometimes tech support might be needed, but it is not necessary.

5. **Success:** Features on the system work together smoothly and efficiently.

6. It is hard to find inconsistency in design unless you dive into a specific feature, which is not part of the given tasks.

7. **Success:** Users think that most people would be able to learn to use this system quickly as a consequence of their own personal experience

8. **Success:** Sometimes there are hassles while using the system, but it's okay to use.

9. Users are not entirely confident in using the system.

10. **Success:** As a user, there are not many requirements to prepare for this app to use.

The results of our study showed that users' task performance and their perceived experience were only able to meet half of our expectations. The comparison between the usability specification table and the table showing the average values of the task performance of all the test participants revealed that the users were able to pick up the various interactions using the leap motion device rather quickly. Whilst we have only implemented a couple of interactions as of now, we think this could be expanded upon to encompass a large diverse range of interactions to easily navigate the application using the device.

Furthermore, in reference to the outcomes from the SUS questionnaire, the perceived user experience could achieve half of the success. Just as in the previous test, the users showed positive reactions towards the general functionality and the layout. However, there were more unsatisfying responses due to the ineffective interaction between the system and the user through the Leap Motion device, such as the lagging responsiveness of the system, and this led users to feel less competent with using the system.

## 5. Reflection

In this chapter, we would like to reflect on our process in three ways in bullet points- things went well, things didn't go well and suggestions.

### Things went well

- *The finalised design*: Unlike the previous time, our test was able to start within a proper setup, which has often led to the users formulating a positive initial impression. The participants gave positive feedback on the interface, such as 'it looks super cool', 'I like the colours and looks professional' and 'how did you make the design?'. Compared to the previous design, we made a minor adjustment to the colours of the buttons in the final design. This helped the users to clearly know where their gestural inputs through the device are aimed at or are heading. Some minor parts were missing, but those were out of our project scope.

- *Improved navigation*: As was outlined earlier in this paper, we improved upon our previous prototype, by integrating a dedicated hand-tracking tutorial near the beginning of the application. Although this did increase the number of time users were spending on getting comfortable with the hand tracking, it also improved their response times in other parts of the application significantly.

- *Learning outcome*: The correct answer rates of the game were around 38% to 48%, we can also calculate a confidence interval for this metric, as was similarly done in the first evaluation. The 95% confidence interval of the current evaluation is [39.5, 46]. Which is slightly better than the performance in the first evaluation which was [36.5, 47.5]. This is logical, as we did not change the mechanics of the applications in any significant way, but we did improve some minor parts of the aesthetic. During the interview, we noticed that some participants could remember some alphabets they learned while testing the application. Although it was a temporary learning moment, it is expected to see positive outcomes in the long run.

### Things didn't go well

- *Interaction with Leap Motion*: It was still difficult for users to interact with the device even when we tried to give sufficient guidelines on how to use it and changed the order of the tasks to help them to be more familiar with it. The revised setup also made it hard for the device to detect the user inputs, and this significantly broke immersion and the flow of the experience. As was outlined in an earlier section, users spent more time executing tasks at the beginning e.g. testing out the hand tracking, scrolling,  and just getting comfortable in general with the device. Furthermore, during the interview with the participants, one said '*it was hard for me to control my fingers and the system seems to require specific actions which I'm not confident with*'. Also, there was another user who commented on a similar aspect- '*It (controlling) doesn't seem to be that hard but it is really difficult*'. We attribute this to the limitations of the hardware and not the software. We can therefore conclude that enhancing the user interaction with the Leap Motion device still has some obstacles ahead of it.

- *Unclear instructions*: One of the participants clearly mentioned during the test that the given instructions were not so well defined and that it made him feel a bit uncomfortable/pressured. We planned and prepared the evaluations beforehand and although the instructions were clear for us to understand, thinking from the users' perspective was missing.

- *Matching the real life & the system:* The illustrations we used for the prototype remained the same as the previous prototype. However, as mentioned, we changed the placement of the Leap Motion device, and we noticed that some illustrations were hard to imitate while the device was in this new orientation e.g the letter 'O'. Similarly, the device had problems detecting the gestures in the new setup. This of course, made it harder for the users to complete the gesture. The letters 'M' and 'N' were the most difficult ones, whilst the letter 'J' was the easiest according to the participants. Though, we may be able to improve on this by retraining the machine learning model.

## 6. Conclusion and Suggestion

Throughout the past couple of weeks, we conducted two usability tests with the target audience to assess whether our outcomes have any user experience related issues or any strengths to emphasise for further steps and, eventually, can answer the research question- 'How can leap motion be used to enhance the learning experience for sign language?'.

Our research data revealed that the Leap Motion device has the potential to enhance the learning experience. However, we noticed that several factors should be considered simultaneously to guarantee an improved user experience in learning ASL. The first aspect is the device itself. During our evaluation, the setup of the device had an impact on the task performance of the participants. While the device was facing the user, it was not able to properly detect the user's gestural inputs and some glitches occurred, which led to an unsatisfying user experience. This made our study conclude that the users performed better in the previous setup where the device was placed in a tabletop orientation. Although this study does not consider this diversion from the setup to be particularly problematic, it guides further studies to consider improving the machine learning model to more accurately capture user inputs in this front-facing orientation. Another factor is the content of the system. Sign language is not commonly used among general users. However, from the interview, some users said it was an interesting experience and are now interested in learning more about the subject, such as gestures that are actually used in daily

conversations. This increased interest consequently shows us that there is potential in our application and the field of ASL. Therefore, in order to maintain the users' interest through the learning experience in this application, we found that there should be more content related to ASL instead of only the alphabet, and more gamification elements should be implemented to motivate users to use the system.

The final paper will contain the final overview of our research data and the conclusion of this project in more detail.

## Appendix

Appendix a. Protocol

# Structure: Test #2

**Overview**
- **20-30 min in total**: Task performance & interview
- Offer 2 things to the participant: the instruction slides (PDF) & the SUS questionnaire (Google Forms)
- Since this is our first testing, we'd like to see the core flow- getting used to the leap motion & performing tasks on the website with the device.
- These are the main priority. Once we finalise this, we'll come up with additional features to support the engaging learning experience.

**Preparation**
- The team launches the app. Because the test participants are not familiar with programming.
- Simultaneously, the team guides the user to read the document- showing a brief instruction and a background information of the project.
- After users are done with reading, the testing will begin.

**Users will follow these steps.**
In the meantime, the users are asked to 'say-out-loud' while conducting tasks- say whatever comes to your mind while testing

| | | | | | | |
|---|---|---|---|---|---|---|
| **1** | First encounter | Getting used to the leap motion control | Hover over the dashboard | Click to go to a certain page | Scroll up-Down | Go back to the main screen |
| **2** | Explore (Dictionary) | Go to 'Dictionary' & Explore | After exploring 'Dictionary', go back to the main screen | | | |
| **3** | Learn #1 (Navigate) | Find 'Play' | Click 'Play' | | | |
| **4** | Learn#1 (Play) | Select the topic 'alphabet' | Learn the basics | | | |
| **5** | Game mode (Navigation) | Select the topic | Select the gamemode | | | |
| **6** | Game mode (Play) | Do as much as you want | Press 'stop' if you'd like to quit | | | |

**After the user is done with the tasks, a brief interview session will be followed.**
Disclaimer: The interview will be take in an informal way. Try to have a casual conversation with the users.
However, the following topics should be covered during the whole interview.

1. How was the whole experience to you?
2. What did you like about this application? Why?
3. When did you feel '*ah, this could have been better*' while using this app? Why?
4. How do you think about the game mode? Why? What would make you feel more challenged/exciting in this game mode?
5. What other things would you like to see more on this app?
6. Do you remember how to make your name in sign language?

| | |
|---|---|
| Changes in the order | add a detailed task 'learn'<br><br>to answer our RQ |

Appendix b. System Usability Scale Questionnaire

Its Google Form format is available in this link.
https://docs.google.com/forms/d/e/1FAIpQLSdcDQHa05Auo6BnC5TGxUwszf6rit8ESSsnc1n-OO5e8BPzug/viewform?usp=sharing

## Sign Up Testing #2

Thank you for joining our testing session. In this form, feel free to tell us how you experienced our first prototype!

🚫 **nayoung.jung.design@gmail.com** (not shared)
Switch account
\* Required

---

1. I think that I'd like to use this system frequently *

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Strongly disagree | ○ | ○ | ○ | ○ | ○ | Strongly agree |

---

2. I found the system unnecessarily complex *

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Strongly disagree | ○ | ○ | ○ | ○ | ○ | Strongly agree |

---

3. I thought the system was easy to use *

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Strongly disagree | ○ | ○ | ○ | ○ | ○ | Strongly agree |

---

4. I think that I'd need the support of a technical person to be able to use this system *

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Strongly disagree | ○ | ○ | ○ | ○ | ○ | Strongly agree |

---

5. I found the various functions in this system were well integrated *

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Strongly disagree | ○ | ○ | ○ | ○ | ○ | Strongly agree |

**6. I thought there was too much inconsistency in this system** *

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Strongly disagree | ○ | ○ | ○ | ○ | ○ | Strongly agree |

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

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Strongly disagree | ○ | ○ | ○ | ○ | ○ | Strongly agree |

**8. I found the system very cumbersome to use** *

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Strongly disagree | ○ | ○ | ○ | ○ | ○ | Strongly agree |

**9. I felt very confident using the system** *

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Strongly disagree | ○ | ○ | ○ | ○ | ○ | Strongly agree |

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

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Strongly disagree | ○ | ○ | ○ | ○ | ○ | Strongly agree |

Submit                                                      Clear form

# Appendix c. Responses: System Usability Scale Questionnaire

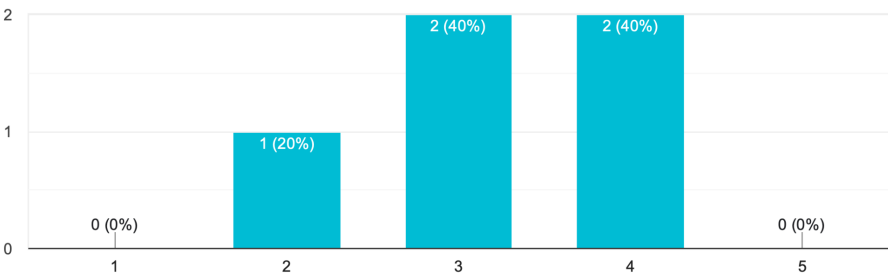## 1. I think that I'd like to use this system frequently
5 responses



| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 0 (0%) | 2 (40%) | 3 (60%) | 0 (0%) | 0 (0%) |
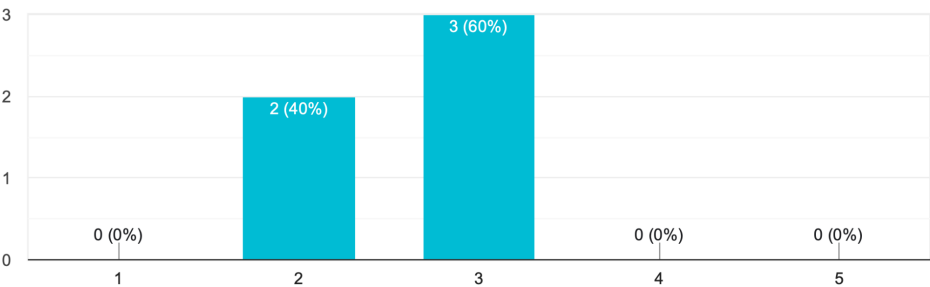
## 2. I found the system unnecessarily complex
5 responses



| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 3 (60%) | 2 (40%) | 0 (0%) | 0 (0%) | 0 (0%) |
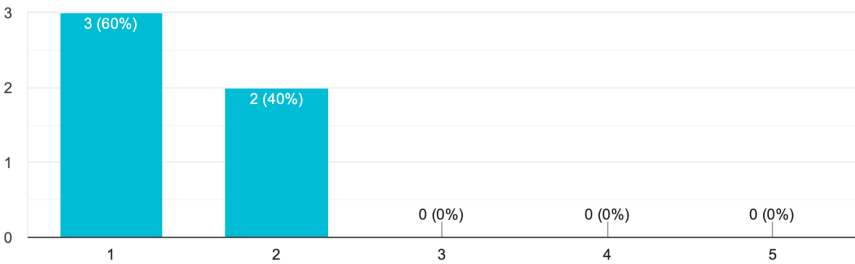
## 3. I thought the system was easy to use
5 responses



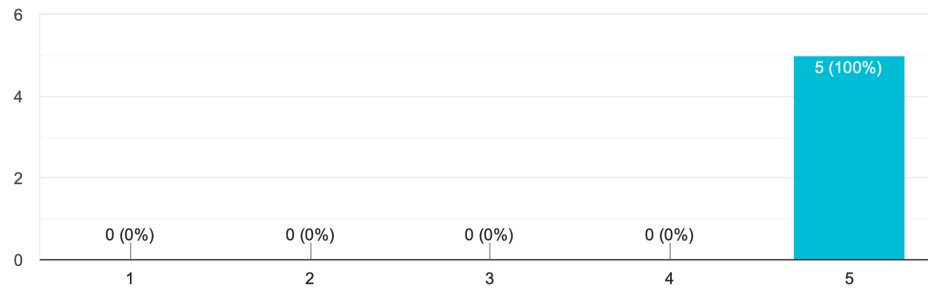| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 0 (0%) | 1 (20%) | 2 (40%) | 2 (40%) | 0 (0%) |

## 4. I think that I'd need the support of a technical person to be able to use this system
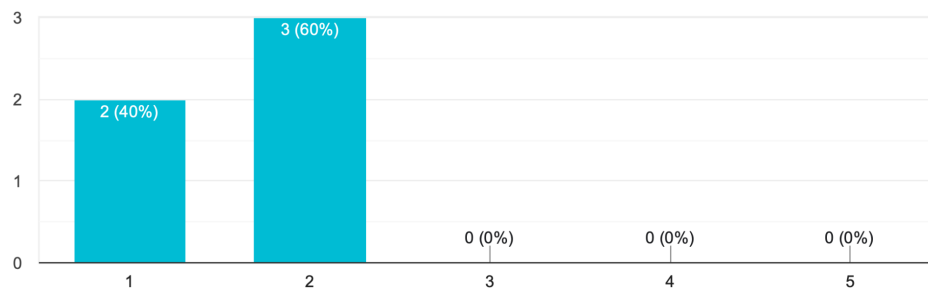5 responses
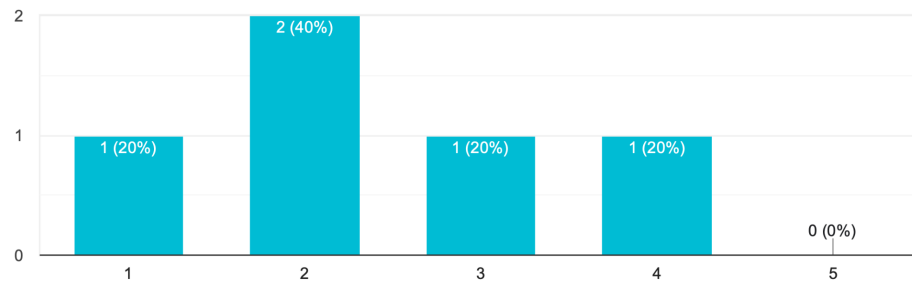


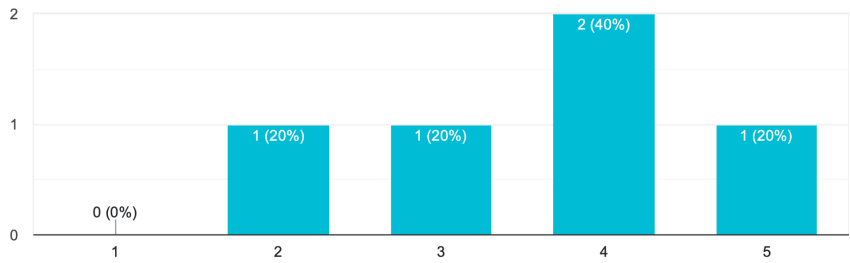## 5. I found the various functions in this system were well integrated
5 responses



## 6. I thought there was too much inconsistency in this system
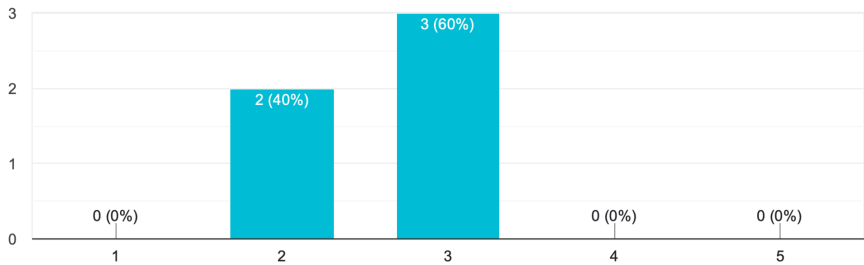5 responses

## 7. I would imagine that most people would learn to use this system very quickly
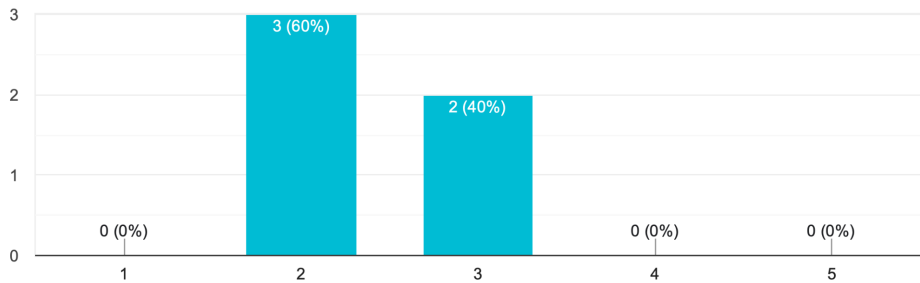
5 responses

Bar chart showing responses on a scale of 1 to 5:
- 1: 0 (0%)
- 2: 1 (20%)
- 3: 1 (20%)
- 4: 2 (40%)
- 5: 1 (20%)

## 8. I found the system very cumbersome to use

5 responses

Bar chart showing responses on a scale of 1 to 5:
- 1: 0 (0%)
- 2: 2 (40%)
- 3: 3 (60%)
- 4: 0 (0%)
- 5: 0 (0%)

## 9. I felt very confident using the system

5 responses

Bar chart showing responses on a scale of 1 to 5:
- 1: 0 (0%)
- 2: 3 (60%)
- 3: 2 (40%)
- 4: 0 (0%)
- 5: 0 (0%)

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

5 responses