

SafarNama



Group Number: 14
Group Name: Itehaad
Project Phase: Phase 4 – Evaluation
Group Members: Noverah Khan, Zoraiz Qureshi, Hamza Farooq, Ahmad Farhan, Farukh Rasool

Table of Contents

Hi-fi Prototype (final version)	5
1.1. Changes	5
2. Usability Test Planning	6
2.1. Purpose and objectives	6
2.2. Participants	6
2.3. Scenarios and Tasks	7
2.4. Test materials	9
3. Usability Testing	10
3.1. Test Procedure	10
3.1.1 Environment	10
3.1.2 Methodology	10
3.1.3 Interview Script	11
3.2. Pre-test questionnaire	12
3.3. Post-Test questionnaire	13
3.5. Measurements	15
3.6. References	16
4. Results (Findings)	18
4.1 Quantitative Analysis	18
4.2 Qualitative Analysis	27
4.3 Observations	32
5. Discussion (about evaluation)	33
6. Conclusion	34
7. Appendix	35

1. Hi-fi Prototype (final version)

[Final Hi-Fi Prototype Link \(APK\)](#)

1.1. Changes

Only a few minor edits were made for the Hi-Fi prototype before we began evaluation:

- **Improved Responsiveness:** We noticed some responsiveness issues on various Android devices for our target platform for some specific screens e.g. the Final Quiz, the Tour History and the FAQs. These were immediately fixed and scaled to perform on all devices. This change was necessary as our built application prototype was required to be installed on devices of testers unknown to us for this phase and any UI issues could have led to bias in test results and responses given, and negatively affected user experiences.
- **Tour History updates after tour completion:** Another important and viable task for our application was testing out how feasible it was to inspect the tour history and look at the last tour they played to gather details to thoroughly test it out. For this reason, after the user completes the Lahore Museum tour, they will find a record for their last played tour with the number of coins they gained, time taken and the name statically coded as “You” just to engage them further for this task.
- **Axis Rotation for 3D models:** Axis-based rotation was added for all 3D models for exhibits, which is performable via the twist finger gesture using two fingers. Previously interactions were only limited to translation and scaling of the 3D models. This change was necessary to check how users would test and react to this specific interaction, and also to enable them to thoroughly inspect the 3D models which account to some of our measurements and usability heuristics.

2. Usability Test Planning

2.1. Purpose and objectives

The purpose for usability test planning is to evaluate and assess our application design. Given that the user experience of the application was designed while keeping in mind the target demographic as well as the problem being solved, usability planning will allow us to compare our perceived ideas of user experience against how the user actually experiences the application.

This will allow us to locate any inconsistencies and issues within our design which need to be fixed. Using usability testing will also help us understand what areas of the problem is our solution fixing and what areas are still not solved.

Our testing will primarily consist of testing how the user plays through a particular tour. This is because the tour is the main element of our app which provides the main solution to the problem of children not being interested in History. This will involve testing gameplay elements such as dialogues, quizzes as well as interactions with the models. A secondary target of testing will be the main user interface such as the main menu and the related settings. This will also involve testing of how the user understands and sets up the magic sheet which is the most vital part to our application.

Our objective in Usability testing is to not only locate any usability flaws within the application but also locate interface flaws as well as the acceptability of the solution. Specifically, our focus will be to evaluate the following:

1. How will the user interact with and respond to Augmented Reality?
2. Can the user locate the required functions naturally?
3. Does the user understand how to set up the magical sheet?
4. Does the user understand how to play the game?
5. Is the user satisfied with the implementation of the quizzes and dialog as a solution to the problem?
6. Can the primary user (children) successfully complete a tour without guidance?
7. Will the user be able to follow the story and retain the information presented by the various exhibits?

2.2. Participants

Participants consisted of three primary parties: children, parents / legal guardians, and siblings. Children, from grade 5 to grade 8, were the primary test group. Some tests were conducted with slightly older children as well to verify if that age group would be interested as well.

Demographics of our participants and other associated details (obtained from our consent forms and pre-questionnaire) are as follows. Please note that names have been anonymized to ensure confidentiality and privacy of user information, and mapping of aliases to names is present in supporting documents.

Alias	Gender	Age Range	Medium of Instruction at School	Printer at home	Knows about AR	Lahore Museum Visited
P1	Male	18-24	English	No	Yes	No
P2	Male	14-17	English	No	Yes	Yes
P3	Male	18-24	English	No	Yes	No
P4	Female	14-17	English	Yes	No	Yes
P5	Female	14-17	English	No	Yes	No

P6	Female	14-17	English	No	Yes	No
P7	Female	11-13	English	No	Yes	No
P8	Female	11-13	English	Yes	Yes	Yes
P9	Female	11-13	English	Yes	Yes	Yes
P10	Female	18-24	English	No	Yes	Yes
P11	Male	18-24	English	No	Yes	No
P12	Male	18-24	English	Yes	Yes	Yes
P13	Male	5-10	English	No	No	No
P14	Male	11-13	English	No	No	No
P15	Male	25+	English	No	No	No
P16	Female	25+	English	Yes	No	Yes
P17	Female	25+	English	No	Yes	No

2.3. Scenarios and Tasks

To test almost every feature of the application, we came up with a list of tasks. We tried to come up with tasks that would map the scenarios in the design document. The tasks were designed in such a way to analyze certain aspects of user interaction and experience while using the app. The main aspects that we tried to inspect were:

1. The time taken to understand a certain task and then successfully complete it.
2. Easiness of finding different features and functions provided by the app.
3. The experience of interacting with an exhibit and listening to a story.
4. The capability of students to retain information and then successfully complete a quiz based on that information.

The tasks designed, covered all the features implemented in the application portrayed in Table 2.3.1. Close analysis of the results of the completed tasks and the user responses, while performing the tasks, made us analyze the user experience and what aspects contributed positively and what contributed negatively to the overall user experience. The tasks also focussed on providing an interactive testing experience.

The tasks are designed in such a way that when users complete a task, they will immediately be able to identify a change of state in the app. Every next task is dependent on the completion of the previous task, this dependency allows the user to identify successful task completion and also motivates them to complete upcoming tasks.

Before starting the testing process, the users were given a slight background of the application and the motivation behind making the application. The tasks, given to the users, have been listed below.

The tasks were divided into two categories: one for the children/students and one for the parents. The tasks were divided because of the different use cases of both types of participants. The children were required to start and complete a tour. On the other hand, the parents were required to set up a tour and start it for their children. Thus, there were two different sets of tasks for children and parents, with the *highlighted ones for parents only*.

Task	Description
T1	<p>Read FAQs and About Us</p> <p>You wish to learn more about the app and the makers of the app. You also have some questions regarding the app and want to get quick answers from within the application. Your first task is to find and read the About Us and FAQ's section.</p>
T2	<p>Change Language (After Selecting it the First Time)</p> <p>Imagine that you are not comfortable with using the app in the language you have currently selected. You find it difficult to understand and read in it.</p> <p>Your second task is to attempt to change the language of the app after you have already selected it.</p>
T3	<p>Select and Start Specific Tour</p> <p>Your third task is to find the Lahore Museum tour and Start it. Make sure you set it up first.</p>
T4	<p>Setup Tour using the Magic Sheet</p> <p>(With Printer) Your task is to download the marker magic sheet of the Lahore Museum tour, cut it into separate sections and set up the markers across your room.</p> <p>(Without a Printer) We have provided you with a digital picture which shows the markers printed, cut and placed on a floor. Please open it on your computer and tilt your display horizontally to use it with ease.</p>
T5	<p>Find and Talk to an Exhibit</p> <p>After starting the Lahore Museum, your next task is to find and focus on the first marker, to listen to the story of the first exhibit as instructed.</p>
T6	<p>Interact with the first 3D Model</p> <p>Now that you have found and talked to the exhibit, your task is to interact with its 3D model on the screen. Try experimenting with how you can scale, rotate and move this 3D model. Make sure you place it back on its marker after you are done.</p>
T7	<p>Find and Talk to the next Exhibit (and take its Quiz)</p> <p>You have visited an exhibit and listened to its story. It has stated some facts and information during your interaction with it. You must find and interact with the next exhibit as instructed. Your task is to complete this second exhibit's quiz successfully using the knowledge you have gained.</p>
T8	<p>Complete the Story</p> <p>You have visited the second exhibit and answered its question successfully. Your next task is to scan all of the remaining markers and complete the story.</p>

T9	Take the Final Quiz and Complete the Tour After listening to the last exhibit's story, your task is to take the final quiz and complete the tour by successfully answering all questions in the final quiz.
T 10	Inspect Tour History Now that you have completed the tour. Your task is to locate the Tour History list and find details about the tour you last played.
T 11	Attempt to Find and Unlock the Second Tour Try searching for any other tours in the application apart from the Lahore Museum one and attempt to unlock them. If that tour is not yet available you should be informed that it is coming soon.

Table 2.3.1 Tasks List

2.4. Test materials

For virtual interviews online:

- **User's Android Mobile Phone:** The application is currently only supported on Android, but it is responsive to all screens so the user's phone was recommended for the most personal and realistic experience. It was required to be connected to the video chat application as well for screen sharing.
- **User's Personal Computer:** The user was encouraged to additionally connect to the video chat via their computer to view our markers sheet and also share their own video to capture their expressions during evaluation.
- **SafarNama APK:** For installation from scratch on the user's Android phone.
- **Digital Image of Cut AR Markers:** An image sent to the printed, cut and placed on a wooden floor to simulate the actual experience. It was required to be opened up on their computer and they were asked to assume that this was the end result after they had performed the tour setup, to overcome the issue of printer access virtually.
- **Pre-Questionnaire Form Link**
- **Post-Questionnaire Form Link**
- **Conductor's Computer for Video Conferencing**
- **Consent Form**

For physical interviews:

- **Conductor's Android Mobile Phone:** The application is currently only supported on Android, but it is responsive to all screens so the user's phone was recommended for the most personal and realistic experience. It was required to be connected to the video chat application as well for screen sharing.
- **Camera/Phone for Recording**
- **Printed Magic Marker Sheet:** As required by the application, but it was printed beforehand but cut out by the users themselves.
- **Printer:** For some of the physical tests, a printer was available and the users were asked to collect the pre-printed AR marker sheet from its tray to simulate the actual experience.
- **Scissors**
- **Pre-Questionnaire Form**
- **Post-Questionnaire Form**
- **Consent Form**

3. Usability Testing

3.1. Test Procedure

3 .1 Environment

Our test procedure was conducted in a limited virtual environment rather than a natural environment. The reason for this was the ongoing pandemic due to which we were limited to testing from home and interacting virtually with the participants.

The environment involved using Zoom to observe the participants. The conductor would invite the participant to their Zoom session. The participant was required to have access to two devices, one of which was the test device: an Android phone with the SafarNama app installed and the other which was used to load up the image of the marker sheet. This was usually because most users did not have printers at home and the current situation prevented them from getting the sheet printed out of home. However, it is noted that loading up the image on a phone or any other device is an easy and viable option. After the participant had joined the Zoom session, they would share the screen of their test device so that the conductor could observe and record the participant's actions. A limitation with this approach was that Zoom required Android devices on Android 10 to be able to share audio, thus all of the Zoom recordings did not contain device audio. This setup was still a limited environment for the conductor because the AR nature of the app would have required an open space for the app to be properly tested, however this solution still provided the best alternative.

A few tests were conducted live with participants. These were few in number due to the pandemic. Additionally, schools were operating virtually / on holidays as well which is why we were unable to conduct any tests on teachers or a group of students. The live tests had the benefit of testing if the participant would naturally understand how to operate the marker sheet as well as any difficulties that they would face. This scenario involved the conductor recording the scenario with their phone's camera as well as acting as a facilitator. The participant would be sitting in a natural environment such as their room and would be using the app on an Android device normally. This environment had no barriers or constraints which is why it provided very useful results.

1 .2 Methodology

The method for conducting a usability test consisted of six parts. For the first test, a pilot study was conducted which involved all group members being part of the testing to observe how to conduct the test as well as locate any misunderstandings or errors that may occur in future tests. This was helpful in removing any bias from future tests as well as maintaining a consistent testing procedure.

The first step in the testing involved asking the user for their consent to record the test. If the user was a minor / child, the consent was taken from their parents or legal guardian. The consent was filled by the conductor with the correct information.

The next step involved having the participant fill the pre-test questionnaire. The pre-test questionnaire was a short survey which was used to get an idea of the demographic and basic information of the participant. This would be useful for the finding and analysis part as it would allow us to deduce factors that may have affected how the user experience changed with the demographic of the user.

Following the pre-test questionnaire, the user was given an explanation of the SafarNama app such as its purpose and its goals. This would provide the user a basic understanding of the type of app they were dealing with as well as act similarly to the description section provided to users on the App Store before they can download a game.

The fourth step of usability testing involved the user starting the app and performing the tasks that they were assigned by the conductor. This was the crux of the usability testing as this is where the conductor could observe the problems in the application as well as collect the various testing measurements. The conductor acted mainly as a facilitator by providing the participant with the task that they were asked to perform. The tasks themselves were general tasks which provided no hints to the user such as “locate the first exhibit and interact with it”. Moreover, the facilitator observed the participant and only guided them when they faced any issue with completing the task. Each time this happened, that event was recorded in the findings as a possible error that would need to be resolved. Since a recording was available, the conductor did not manually measure the time taken to complete each task. Instead, this was extracted from the time-frame present in the video recordings.

The fifth step, after the completion of all tasks, was for the user to be given the post-test questionnaire. This would gauge the main items for analysis in the findings such as the ease of use, the consistency of the app, etc. It consisted of questions that were used to understand the overall experience that the user had with the app, any issues they faced and any suggestions they could provide for the app. The measurement factors were provided on a Likert scale which provided a quantitative measure of the measures. The issues and suggestions were filled in text boxes which could be used for qualitative measurement along with the conductor’s observations of the participant.

Finally, the conductor asked the user for verbal feedback as well so that the user was not limited to the questions on the post-questionnaire. This provided additional context as well as a view of the user’s perspective to the conductor.

An additional consideration while performing the tasks was friendliness and openness with the user. This would be beneficial so that the conductor could observe the user working naturally and without any pressure. This was especially considered in the case of children since they would be easily influenced by the nature of the test. Thus, the facilitator would be more helpful to children similar to how parents and siblings might help them. However, it was also made sure that not too much help was given so that the child’s own decisions were not influenced.

3 .3 Interview Script

This section goes over the general interview script that was followed by each conductor. Each conductor did not follow this script exactly, but made changes as required and only used this script as a general guide of the steps to perform. This would ensure uniformity and consistency amongst all of the conducted tests. The following is the interview script:

(General introductions with the user)

(If the user is not a minor) If it is fine with you, we would like to take your consent to record this testing. The purpose of this recording would only be to review the test for our required testing measurements and analyse the testing for issues in the user experience. Your information would not be disclosed publicly, will only be shared with the team members and will be deleted as soon as the analysis is complete.

(If the user is a minor) We need to speak with your parents or legal guardian to ask for their consent to record this meeting. If it is fine with you, can we speak with them?
(With Parents:) we would like to take your consent to record this testing. The purpose of this recording would only be to review the test for our required testing measurements and analyse the testing for issues in the user experience. Your child’s information would not be disclosed publicly, will only be shared with the team members and will be deleted as soon as the analysis is complete.

(After the consent form has been filled)

Thank you very much for participating in our research study. The goal of this project is to test our application called SafarNama. The results of our evaluation will be kept completely confidential and only shared with our project supervisor. I am

sending you our application [APK](#) (send the link to the apk on Zoom or other similar service), please take your time to download and install it. In the meanwhile, here is a short [questionnaire](#) (send the link on Zoom or other similar service) that we would like you to fill. (Questionnaires are to be filled by children with the facilitator providing help if anything is misunderstood)

(Recording Starts) Before we tell you ourselves, what comes to your mind when you hear our application's name *SafarNama*? Alright. As a general overview, *SafarNama* is an educational virtual touring mobile application to showcase the heritage and culture of Pakistan through gamified Augmented Reality based stories for children using interactive dialogues and 3D models.

You will be asked to use *SafarNama* to perform a series of specific tasks related to the application. Just try to *think out loud* as you try to complete your tasks while using *SafarNama*. Please let me know if you have any questions about it. (Answer any questions by the participant)

Alright, please launch the *SafarNama* app that you have installed. (Keep the [task list](#) open, narrate all tasks and their descriptions, and start timing them as soon as they have understood and start interacting with the screen. Aid the user if they are stuck for more than a minute or unable to perform a certain task.)

(After all tasks are completed)

Please fill out this final questionnaire (for [children/siblings/students](#), for [parents/guardians](#)) to give us a complete overview of your experience using the application.

(After completion of final questionnaire)

How was your experience? Do you have any suggestions for the application? (Probe accordingly)

Thank you for your time!

3.2. Pre-test questionnaire

[Pre-Test Questionnaire Google Form Link](#) (Also available in Supporting Documents)

The questions in the pre-test questionnaire were constructed to get an idea of the audience and the demographics involved. This questionnaire was filled by all participants so that we could relate their performance and experience to their demographics. The purpose of this questionnaire was to obtain quantitative feedback about the user. The reasons for asking each question are explained below:

Age and Gender

The pre-test questionnaire was filled by a mixture of children, teens, siblings, and parents. Thus, these two questions were added so that the team could maintain different analytics for each age group as well as to verify if the sampling set was diverse.

Medium of instruction at school

This question was only asked to the participants who were between the ages of 5 and 20. This was done because only these participants were given tasks to play the tour and complete the story. Therefore, this question could be used to determine if medium of instruction and thus the language for the dialogue in the game became a factor in the understanding and following the dialogue in the story. This would be useful for

evaluating if perceived usefulness and degree of interest changed with the medium given that the app only had English language at the time of testing.

Do you have a printer at home?

This was used for two reasons. The first reason was to check how many people from the demographic actually had printers at home. The second reason for this was so that the facilitator could keep this in mind when conducting the tasks and ask the user to perform the task corresponding to the availability of printers differently.

Previous Experience with AR?

This was used to understand if the user had any previous experience with AR and if they did have some experience, were they able to recall how to use and manipulate the models similarly using markers. This would work with the learnability measurement to see if the user was able to recall how it works based on their memory.

How often do you play mobile games?

This was used to gauge if a user has any experience with mobile apps, particularly games, and if that played a part in them being able to recognize and map the interface with similar apps.

Have you ever visited Lahore Museum?

This question would work along with a similar question in the post-test questionnaire to see if the Lahore Museum stage in the SafarNama app developed an interest in the user to visit it in real life if they had not previously visited it.

3.3. Post-Test questionnaire

The post-test questionnaire was divided into two forms. One form was to be filled by children, their siblings, and students. The other was to be filled by parents and guardians. This was because the tasks for parents and children were separate and focused on different types of evaluative metrics. While the students were asked to start and complete a tour, the parents were only asked to set up the tour and interact with the pre-tour features such as the tour history.

[Post-Test Questionnaire Google Form Link for Children/Siblings/Students](#) (Also present in the Supporting Documents)

The questions in the post testing questionnaire for children/students were constructed while keeping in mind the learning and interactivity aspect of the application. They focused on evaluating the participant's ease-of-use and comfortability with the app. The main focus of the questions was to get an idea of how easy it was to perform different tasks and follow the instructions within the app. Apart from getting a general review, the children were also asked about specific things such as the accuracy and relevance of the 3D models.

The student questions were structured in a way to get quantitative feedback on a linear scale of 1-5. The responses, based on the numerical data, allowed us to get a better visualization of the distribution of reviews of the children. In addition to the quantitative response questions, the students were also asked to answer questions that provide qualitative data. They were asked to provide suggestions about the app and identify interesting things in the app. This allowed the students to openly write out their experience and identify things that were not asked in the questionnaire. The explanation for each question is provided:

Application was easy to learn even after using it for the first time

This was used to gauge how the interface design of the application and if it set up according to Android standards. If this was the case, then the app would provide the user with mappings of where to find what they need without being taught about it.

Whenever I made a mistake, I could recover quickly and easily

This question was used to analyse the aspects of user control and error prevention present within the app in the form of warning dialogs allowing users to cancel or return from unwanted events such as asking for confirmation before starting the final quiz.

I felt comfortable using the application

This was used to understand if the design heuristics of consistency and standards were implemented efficiently to make the application easy-to-use for the user.

I can clearly remember where everything is located after using the application once

This was used to gauge the learnability and recall of the user and evaluate how well they grew accustomed to the interface.

I was clearly informed about what to do and how to progress in the game

This was used to measure if users were able to recognize and understand what they needed to do to progress. This measured the effectiveness of the tour guide in helping the user find the path to the finish.

What animal do you think was the Tour Guide Makhnu?

This was used to check if the users recognized the animal (also the national animal of Pakistan) that the mascot was based-on. Correctly answering this would also mean that the mascot was memorable for the user and would thus be an effective mascot.

How easy was it to start the tour?

This question helped in determining if the user would be able to set up and start a tour by themselves.

How easy was it to read and understand the dialogues of the story?

Given that the application's primary targets are Pakistani children, it is important to measure how effective and easy to understand the dialogue of the story. Without effective dialogue, children will probably be unable to decipher the content of the story which would not be helpful in solving the identified existing issues.

How interesting was the story to you?

This was also an important question as it determined if the app would engage the children. Without interest, the purpose of the app would be lost.

How much do you like the 3D models?

This was used to evaluate if the children formed any interest in the models because the design choice was to go with a mixture of real depictions as well as drawings of the figures. This will help in determining the final design of the models.

How accurate do you think were the 3D models compared to the actual characters?

This was used to check if the children were able to map the 3D models to the depictions of the characters that they had seen within their textbooks or on the internet.

After playing this tour, I am certain that I would like to visit the Lahore Museum and look at these exhibits myself in real life.

This was another important question to see if the Lahore Museum stage within the app helped develop any interest within the user to visit the Lahore Museum in real life. This could be linked to the question from the pre-test questionnaire which asked if the user had previously visited Lahore Museum.

Other than these questions, the user was asked for anything interesting they found within the application to identify what specific elements were useful for developing interest with children. The user's suggestions would also help the team understand the shortcomings in the app as well as features that the target users would love to see within the app.

[Post-Test Questionnaire Google Form Link for Parents/Guardians](#) (Also present in the Supporting Documents)

The questions in the final parents questionnaire focused primarily on the experience and interaction with the app before starting a tour. This is because the parents would mostly be required to set up the markers for their children or check how the children performed on each tour. The questions in this survey also relied mostly on quantitative data to help better visualize and analyze the responses based on statistical inference.

The questions asked in the post-test questionnaire for parents were similar to the questions asked to the children and their explanation is provided above. Parents were additionally asked about suggestions in general for improving the set up of the application so that it was easier to understand and use for them.

3.5. Measurements

Both process and bottom-line data were collected not just from questionnaire responses but also during the evaluation process. For process data, users were firstly asked to share their screens to capture where they were tapping, what specific errors they were making and screens they were landing on incorrectly, their process of recovery. Secondly, they were also asked to think out loud about their actions and permitted to ask for help. Expressions during the process were also captured for most users that gave consent and enabled live video feeds during the evaluation.

The bottom-line data generated includes task times including recovery time upon making any mistakes through those specific tasks and also final tour scores obtained after completion of the main tour task (T9).

Most of our usability measures have been drawn from a systematic review done for such measures specifically for mobile-based augmented reality (MAR) learning applications [1], the slides and other literature comprising usability tests for prototypes created. Moreover, some inspiration was drawn from Ko et al.'s MAR usability principles (five usability principles for AR) [10] consisting of user-information, user-usage, user-cognitive, user-interaction and user-support were used to center our preliminary and post-testing questions around with consideration for each factor partially to the context of the application. The usability measures

are mentioned along with the referenced literature from where it was drawn categorized into performance and preference measures individually for bottom-line and process data respectively. Note that most of the preference measures are exclusive of Nielsen's ten usability heuristics which we also utilized for the design of some of our questions through our evaluation questionnaires.

Performance measures:

- **Time-based Efficiency (Time taken to complete tasks including recovery time):** Using times from recorded videos (in seconds because of their relative significance) from the initiation of the tasks to the achievement of the task goal as soon as the users started interacting with the screen. This time includes recovery time as well in case the users make any error during the process such as tapping on the wrong button, going back or starting the tour without setting it up etc.
- **Final tour score:** A self-developed metric to test the user's knowledge and understanding of the story content involved in the application and also its retainment and recall from memory. After the completion of task T9 which requires the user to give the final quiz and complete the application tour based on the information they have learned about the exhibits and tour's historical context. The minimum achievable score is 150 (due to some previous score gained while they played through the tour), with three questions being asked in the final quiz each contributing to 50 points.
- **Number of uses of help:** Any intermediate questions or calls for support from users were counted during tasks. This was to gauge how well they understood the task, whether they became confused in the process and judge how effective the tutorials and integrated in the application were in addressing these issues. Any bias from the narration of the task description by the conductor is also considerable here as it is also a factor determining this variable.

Preference measures:

- **Learnability [2]:** To measure the learnability of the interface generally, whether the users were able to recall the gestures to interact with the 3D exhibits, how to tap on the dialogues to proceed and access the menu again from the home screen.
- **Ease of Use [6]:** Measured comfort with the interface, navigation to perform certain tasks through the main user interface, hold the device with correct orientation, navigate through the AR tour dialogues and complete the story.
- **Perceived usefulness [8]:** Degree of perceived usefulness and potential of the application for showcasing the heritage and culture of the tour played and as an educational source for gaining historical knowledge.
- **Precision of 3-Dimensional images [4]:** Perceived accuracy of the 3D model to actual historical depictions of the characters or their exhibit counterparts present at the location of the virtual tour
- **Object manipulation [4]:** Ease of manipulation of the 3D objects augmented from exhibits using specific finger gestures for interaction and adaptability of these controls.
- **Degree of interest for the content [9]:** To measure the degree of interest they developed for the storyline, its historical context, the characters and the AR experience.
- **Facial expressions and body movements (Frowning, Smiling, Surprise, Concentration/Focus, Leaning close to screen) [5]:** All facial expressions were captured for the users that gave consent and permitted sharing of their video, and their expressions were generally measured under these specified categories to evaluate the

emotional experiences they were having. These emotional responses were vital to understanding what functions were difficult to understand for the user or made them annoyed.

- **Interest (would use again) [7]:** Gauge the reaction and interest of users as they were to inquire whether the degree of preference to which they would actually use the application again and also visit the tour for that specified location in real life themselves.

3.6. References

1. Lim, Kok & Selamat, Ali & Alias, Rose & Krejcar, Ondrej & Fujita, Hamido. (2019). Usability Measures in Mobile-Based Augmented Reality Learning Applications: A Systematic Review. *Applied Sciences*. 9. 2718. 10.3390/app9132718.
2. Albrecht UV, Folta-Schoofs K, Behrends M, von Jan U. Effects of Mobile Augmented Reality Learning Compared to Textbook Learning on Medical Students: Randomized Controlled Pilot Study. *J Med Internet Res* 2013;15(8):e182.
3. Peleg-Adler, R.; Lanir, J.; Korman, M. The effects of aging on the use of handheld augmented reality in a route planning task. *Comput. Hum. Behav.* 2018,81, 52–62.
4. Jamali, S.S.; Shiratuddin, M.F.; Wong, K.W.; Oskam, C.L. Utilising Mobile-Augmented Reality for Learning Human Anatomy. *Procedia-Soc. Behav. Sci.* 2015,197, 659–668.
5. Brito, Pedro & Stoyanova, Jasmina. (2017). Marker versus Markerless Augmented Reality. Which Has More Impact on Users?. *International Journal of Human-Computer Interaction*. 34. 1-15. 10.1080/10447318.2017.1393974.
6. Luís Fernando Maia, Carleandro Nolêto, Messias Lima, Cristiane Ferreira, Cláudia Marinho, Windson Viana, Fernando Trinta. LAGARTO: A LocAtion based Games Authoring Tool enhanced with augmented reality features. *Entertainment Computing*, Volume 22, 2017, pp. 3-13, ISSN 1875-9521.
7. Selviany, A.; Kaburuan, E.R.; Junaedi, D. User interface model for Indonesian Animal apps to kids using Augmented Reality. In *Proceedings of the 2017 International Conference on Orange Technologies (ICOT)*, Singapore, 8–10 December 2017; pp. 134-138.
8. Adrian Shatte, Jason Holdsworth, Ickjai Lee. Mobile augmented reality based context-aware library management system. *Expert Systems with Applications*, Volume 41, Issue 5, 2014, pp. 2174-2185. ISSN 0957-4174.
9. Fenu, Cristina & Pittarello, Fabio. (2018). Svevo Tour: The Design and the Experimentation of an Augmented Reality Application for Engaging Visitors of a Literary Museum. *International Journal of Human-Computer Studies*. 114. 10.1016/j.ijhcs.2018.01.009.
10. Ko, Sang Min & Chang, Won & Ji, Yong Gu. (2013). Usability Principles for Augmented Reality Applications in a Smartphone Environment. *International Journal of Human-Computer Interaction*. 29. 10.1080/10447318.2012.722466.

4. Results (Findings)

4.1 Quantitative Analysis

Questionnaires

Two online surveys were conducted, one before and one after the testing, to get the quantitative data from primary users. Each user, testing the app, was asked to fill the survey. 17 responses were recorded from both the surveys.

Of the 17 respondents, 9 were female and 8 were male as shown in the Appendix (Fig 7.1.1). Thus, the sample responses were almost equally divided amongst both genders. The respondents belonged to five age groups: 5-10 years, 11-13 years, 14-17 years, 18-24 years and 25+ years old shown in Fig 4.1.1. There is a differentiable split in the age groups, allowing a variable level of responses from those of children to those of parents and guardians. All of the respondents, aged between 5-20, had studied in or were studying in an English medium school, seen in Fig 4.1.1.

This allowed them to understand the English version of the app with comparatively low difficulty. Users were also asked about the availability of the printer at their homes. 12 respondents did not have a printer, while the remaining 5 did have direct access to a printer as seen in Appendix (Fig 7.1.2). Thus, it would seem that our app should put more emphasis on the ability to open up the marker sheet on a separate device and scan the marker from there. This was possible, but almost all of the users were unaware that they could scan the markers using another device. The respondents were then asked whether they have played or interacted with applications, like SnapChat and PokemonGO, based on Augmented Reality. 12 people said that they have used an application based on AR while the remaining 5 had never used such an application as seen in Appendix (Fig 7.1.3). Thus, it was the first experience with AR technology for them and it was seen in the tests that all of the participants were quickly able to pick up how to use the markers with their phones. Thus, it was a simple process to understand and follow for most users. Moreover, since most users had experience with AR technology, it would seem that the general public would have similar experiences as well and would thus be able to quickly understand our application. In addition to this, 7 users said that they played mobile games on a daily basis while 8 respondents stated that they played games weekly and occasionally, as seen in Appendix (Fig 7.1.4). There was not any single user that had never played a mobile game, which is why more users could use their knowledge from other similar games within this game as well. The users were also asked whether or not they have visited the Lahore Museum prior to using the app for the first time. 10 people said that they had never visited the Lahore Museum before. On the other hand, 7 of the respondents stated that they had visited the Museum at least once, as shown in the Appendix (Fig 7.1.5). Thus, for the 10 people who had not visited Lahore Museum previously, we could check if playing the app induced any more interest in visiting the Lahore Museum. The majority of the users who had not previously visited the Museum before displayed positive responses to visiting the Lahore Museum after playing our app.

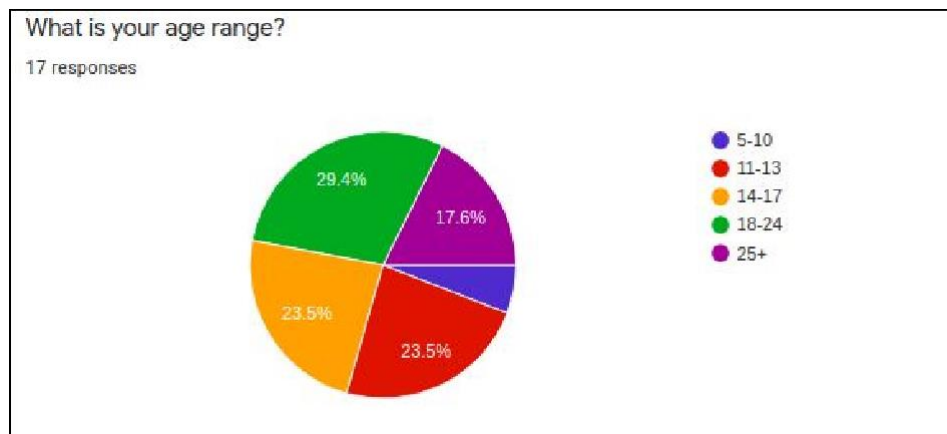


Fig 4.1.1. Variability in Respondent Ages

The second survey consisted of two variants, one for the children/students and one for parents/guardians. The first variant, for students and children, focused on getting statistical feedback from the users. There were 13 responses for the children / student variant. After testing the app, the users were first asked, on a scale of 1-5, how easy was it to learn the application even after using it for the first time. 9 out of 13 respondents strongly agreed that the app was easy to use. 1 out of 13 children strongly disagreed and implied that it was very difficult to learn the app usage when playing for the first time. Thus, it would seem that we need to revisit and fix some specific aspects of the tutorial and guides within the application to make it easier to understand for the users. These stats show that 12 of the users found the app to be easy to use. With the children belonging to different age groups, these results show a positive trend and adds to the ease of app usage. These statistics can be seen in Appendix (Fig 7.1.6). These results also greatly support the preference measures of **ease of use** and **learnability**, discussed above.

Next, the students were asked whether or not they could recover from a mistake made quickly and easily. 10 out of 12 respondents, agreed that they could quickly recover from a mistake made while using the application. Only one respondent said that they could not easily recover from a mistake. Again the majority of respondents were easily able to recover from mistakes, and shows that the app is **easy to use** for children as seen in the Appendix (Fig 7.1.7). It also adds to the heuristic, **error recognition and recovery**.

The respondents were asked whether they felt comfortable using the application. 12 out of 13 children agreed with the statement while only a single child was neutral on the question as shown in the Appendix (Fig 7.1.8). This shows that they found the content of the tour quite interesting and did not find it too difficult to understand and interpret and hence could comfortably use the app. This maps one of the usability measures - **degree of interest for the content**. The users were very much interested in the content of the story and the application as a whole.

To get an idea of whether the children were able identify the mascot, and relate with its context, they were asked to write down what animal they think was the Tour Guide Makhnu. 11 out of 13 children identified the mascot correctly, as Markhor. Only 2 kids did not recognize what animal the mascot resembled. The results are shown in Appendix (Fig 7.1.9). This showed that the mascot was an accurate as well as an interesting depiction of the national animal of Pakistan that could encourage interest in the children.

The next question, in the questionnaire, focussed on finding out the interaction and experience of the children when playing a tour. They were asked to rate the ease of starting the tour. The response statistics are shown in Appendix (Fig 7.1.10). The results showed variable responses. 9 out of 13 kids found the tour starting task to be easy and doable. 3 out of 13 kids found this task to be either difficult or very difficult. These results highlight one of the potential areas of improvement. The user experience was negatively affected by the difficulty present in starting a tour. The preference measure, **ease of use** is satisfied here.

To get an idea of how the users felt when reading the dialogue, they were asked to rate the dialogues based on the level of ease of reading and understanding. Appendix (Fig 7.1.11) shows the results. 5 children found the dialogues to be easy to read and understand while 5 students found them to be difficult. 3 children were unsure to take either side.

To find out whether or not the students found the story of the tour interesting, they were asked to rate the story on a scale of 1-5, with 1 being very boring and 5 being very interesting. 12 out of 13 respondents found the story interesting. Only one child found the story to be boring. The majority of children, finding the story to be interesting showed that they were eager to find out about the story and what happens next in it. This also motivated them to complete the tour and understand and retain the information to successfully complete every quiz. This maps one of the usability measures - **degree of interest for the content**. The users were very much interested in the content of the story and the application as a whole. The results are shown in Appendix (Fig 7.1.12).

One main part of the user experience were the 3D models. A positive review and interaction with these models would have ensured a successful story. The responses for this question were also crucial to find out the effectiveness of the preference measure: **Precision of 3-Dimensional images**. Appendix (Fig 7.1.13) shows the statistics for the question where the users were asked to rate how much they liked the 3D models. 11 out of 13 users said that they liked the 3D models a lot. 1 user was neutral and neither liked nor disliked the models. Only 1 child that did not like the 3D models. The majority here shows that 3D models played a pivotal role in adding to positive user experience. The models increase the interest to complete the story and also encourages them to interact with it, hence positively adding to the **object manipulation** preference measure, as more interaction would make the users learn about how to use finger gestures to interact with a certain model. Users were also asked to rate the similarity of the models with their actual exhibits. The responses support the **Visual Appeal** heuristic, as the models caught the eye of the users and engaged them in the tour. Appendix (Fig 7.1.14) shows that all the users thought that the models were accurately made and resembles the real exhibits.

The table below shows the average score of questions asked from the children/students in the survey. The score is computed by averaging out the user responses given on likert scale of 1 - 5, with 5 being the highest value and 1 being the lowest. A score of 5 shows that a user strongly agreed with the statement of the question, on the other hand a score of 1 suggests that the respondent strongly disagreed with the statement of the question. A high average score means that the user agreed with the question statement strongly and vice versa.

Preference Metric / Heuristic	Question	Average Score (out of 5)
Learnability	The application was easy to learn even after using it for the first time	4.31
Error recognition and recovery	Whenever I made a mistake using the application, I could recover quickly and easily	3.69
Ease of Use	I felt comfortable using the application	4.62
Recognition rather than recall	I can easily remember where everything is located after using the application once	4.23
Help and documentation	I was clearly informed about what to do and how to progress in the	4.38

	game	
Degree of interest for the content	How interesting was the story to you?	4.46
Visual appeal	How much did you like the 3D models?	4.08
Precision of 3-Dimensional Images	How accurate (similar) do you think were the 3D models to their actual exhibits or characters? (Those models that you had seen prior to using this app.)	4.15
Interest (would use again)	After playing this tour, I am certain that I would like to visit the Lahore Museum and look at these exhibits myself in real life.	4.00

Table 4.1.1 Preference Metrics, Questions and Average Likert Scale Scores

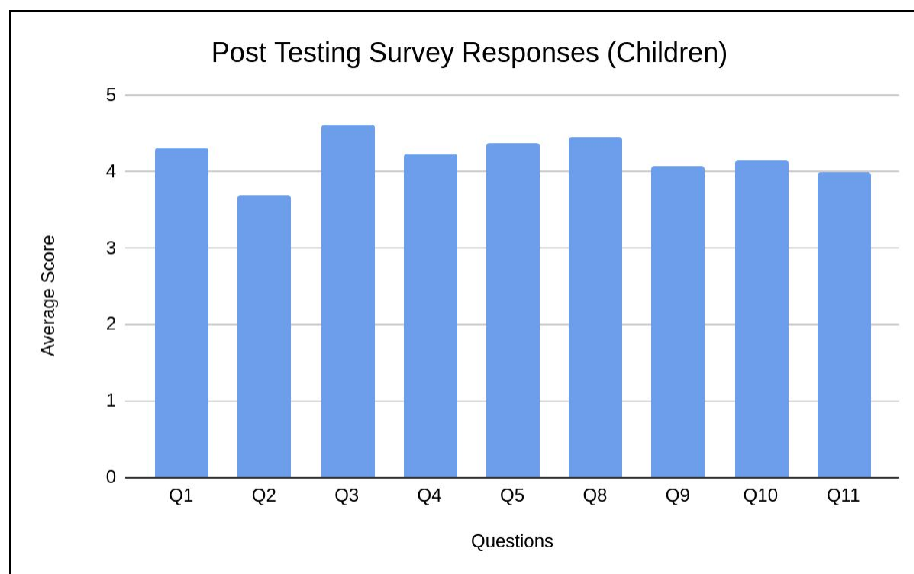


Fig 4.1.2 Average Post-Testing Questionnaire Likert Scores for Respondents

Variant 2 of the second survey focussed on getting statistical feedback from the parents/guardians regarding the application. Their feedback is based on the interaction with the app until starting the tour for their children. A total of 4 parents filled this survey.

After testing the app, the parents were first asked, on a scale of 1-5, how easy was it to learn the application even after using it for the first time. 3 out of 4 respondents agreed that the app was easy to use. 1 out of 4 parents was neutral when answering this question. These stats show that 3 of the users found the app to be easy to use. These statistics can be seen in Appendix (Fig 7.1.15). These results also support the performance measure, *ease of use*, discussed above.

Next, the parents were asked whether or not they could recover from a mistake made quickly and easily. All the respondents agreed that they could quickly recover from a mistake made while using the application. This shows that the app is easy to use for parents also as seen in the Appendix (Fig 7.1.16).

The parent respondents were asked whether they felt comfortable using the application. All of the parents agreed with the statement as shown in the Appendix (Fig 7.1.17). This shows that they found the, pre tour, content and features quite interesting and could comfortably use the app. This maps one of the usability measures: Degree of interest for the content. The users were very much interested in the content, leading up to the story and the application as a whole.

The parents were also asked whether they could easily remember where certain features were located, within the app. All 4 respondents agreed with the statement. This shows that the features and content was correctly located and was easy to access and found by the parents.

Appendix (Fig 7.1.18) shows the user response statistics.

Next, as seen in Appendix (Fig 7.1.19), the respondents were asked whether the instructions, given to them for performing certain tasks within the app, were clear enough. Every parent agreed that they were clearly informed about what to do and how to progress in the game.

Finally, the parents were asked to rate the ease of starting the tour. The response statistics are shown in Appendix (Fig 7.1.20). The results showed variable responses. 2 out of 4 parents found the tour starting task to be very easy and doable. The other 2 parents found this task to be difficult. These results highlight one of the potential areas of improvement. The user experience was negatively affected by the difficulty present in starting a tour.

Task Time Efficiency

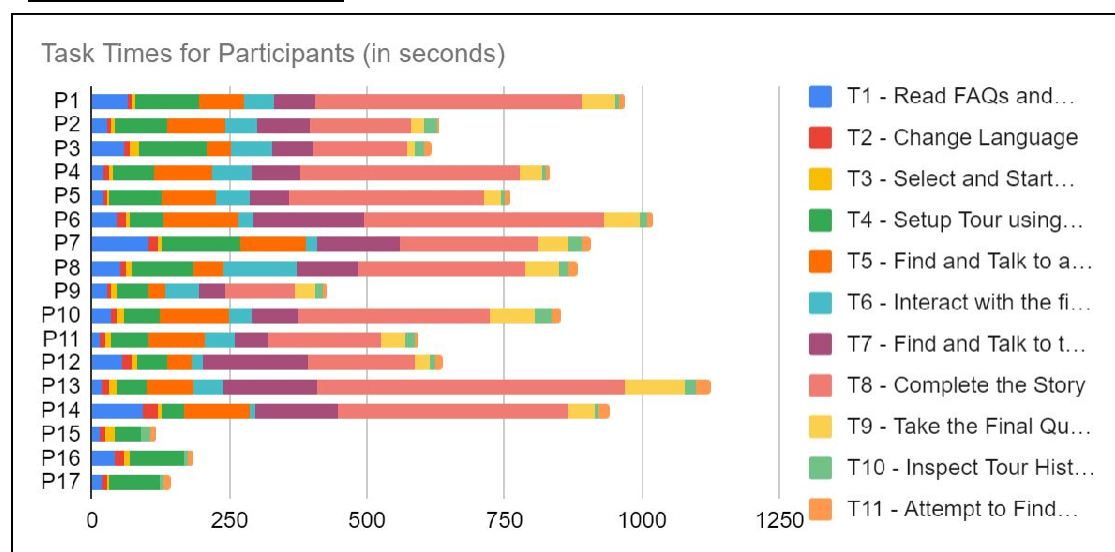


Fig 4.1.3 Task Times for Participants

To measure our first usability metric, task time efficiency, for performance measurement analysis, times for all core application tasks were recorded in seconds from initiation of the task to the achievement of the task goal for all participants. Note that participants P15-P17 are parents and guardians and thus for them, only specific tasks associated with them were tested. Figure 4.1.3 displays the task times for each task for every participant that we obtained from our usability testing.

The results show that Task T3 took the shortest average time to execute (9s) because it involved accessing the first visible tour and starting it with the tap of a single button. This verifies that location and starting a tour from the application is the most convenient task for the user which is accomplished as soon as it launches. Similarly, the average task times for all tasks involving navigating to specific screens (T2, T10, T11) also took relatively short times (ranging from 10-15s). Task T8 took the longest time on average (317s) to execute which is reasonable as it was the major task of completing the entire story of the Augmented Reality tour, after reaching the second exhibit. The setup task (T4) also took a relatively higher average time of 83 seconds because of additional complexity created through the limitation of virtual interviews as users were required to download and display a digital image of the marker sheet instead of printing it

out directly as instructed by the application. Task T6 which involved interacting with the 3D exhibits to scale, rotate and translate them by experimentation also took higher time (average of 53 seconds) even though they involved three different but simple finger gestures. This can be attributed to the lack of a proper tutorial for such 3D interactions and lack of intuitiveness of the gestures to be discovered with ease. This was followed by task T7 (112s) which involved finding that second exhibit itself and passing its quiz successfully on the first try which was not the case for all participants. The average time taken for the tasks specific to parents and guardians (T1, T2, T3, T4, T10, T11) was approximately the same as it involved general navigation through the application interface.

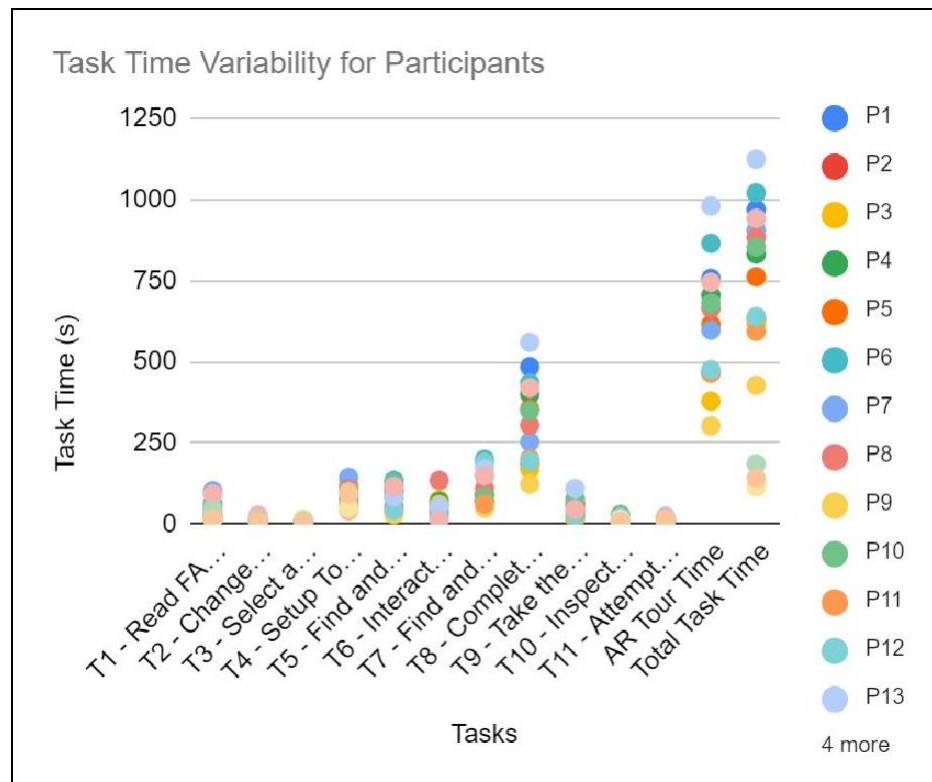


Fig 4.1.4. Task Time Variability for Participants

Fig 4.1.4 demonstrates the variability of task times for every participant. It verifies our previous observation by showing that the highest variance in task times occurs for the tasks T7 and T8. As successfully answering the quiz of the second exhibit in the tour for task T7 depended on user demographics and knowledge of the historical facts along with other factors, the time taken to complete it varies per participant significantly. For this same reason, the time taken to complete the story successfully by answering all proceeding quiz questions correctly for task T8 shows the most significant variance as it was dependent on the participants. As a result, the total task time and the Augmented Reality tour times show significant variability for our sample of participants.

Application Tour Scores

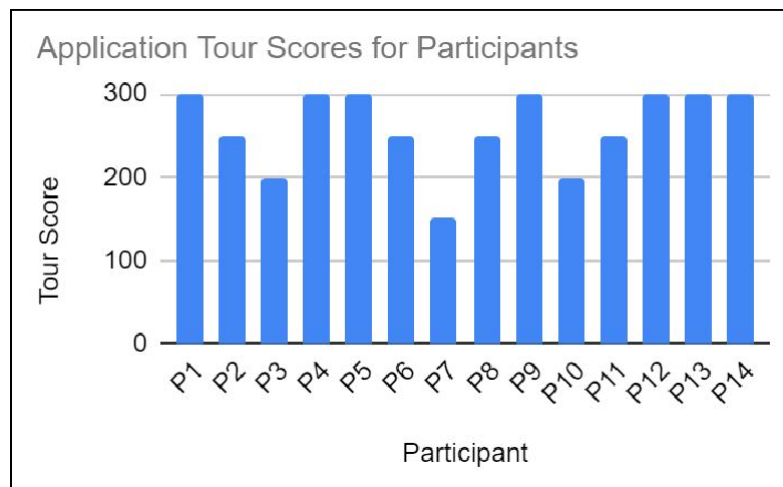


Fig 4.1.5. Application Tour Scores for Participants

Fig 4.1.5 showcases the results obtained for the final tour scores or coins from the main Lahore Museum tour of the application which the users were required to play through the evaluation. The minimum possible achievable score was 150, due to three internal quizzes inquired by exhibits through the story each contributing to 50 points. This score was as a result completely dependent on the final quiz which tests their knowledge and memory of the historical facts covered throughout the virtual tour. The average tour score for our participants was estimated to be 260.71 which is relatively high and shows that they answered at least 2 out of 3 questions correctly in the final quiz. This verifies that the participants of our evaluation clearly understood the story content to some extent and retained most of the facts covered to succeed in this final quiz.

Calls for Help/Support:

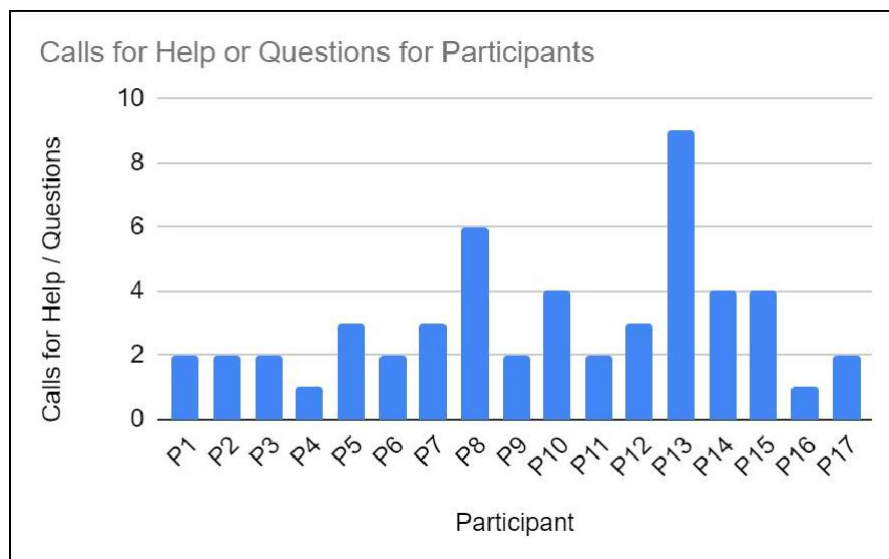


Fig 4.1.6. Calls for Help / Questions for Participants

Fig 4.1.6 displays the results of our third performance measurement which is the number of calls for help or questions asked by each participant individually, the average of which is 3. This is still a relatively high number which shows that some of the instructions for the tasks and their execution were unclear to participants and hints towards lack of understanding for the instructions provided by the application. Participant P13 has the highest number of calls (9) which may be attributed to the fact that they belonged to the smallest age group in our sample of

5-10. However, exact correlations for the number of calls to age cannot be drawn because we are limited by our small sample size used for the evaluation.

4.2 Qualitative Analysis

Pre-questionnaire

"I really liked to see the mughal stories and their way of living" - P9

"It was pretty good. I saw a lot of historical artifacts (e.g. weapons, armours, paintings, etc.)" - P2

"It was very nice I loved all the different statues, paintings, scrolls and coins" - P4

"It was really fun to see the statue of Queen Victoria." - P8

"It was nostalgic." - P12

"My trip to the Lahore Museum was quite rushed, and would have been made better with the help of a guide since most of the time, we didn't really know what we were looking at." - P10

According to the majority of the participants, their physical trip to the Lahore museum was a very enjoyable and memorable one. They enjoyed looking at the various historical artifacts (paintings, coins, weapons, statues, ancient writings, etc.) in the galleries of various ancient civilizations (Mughal, Hindu and Jain, Harappan and Buddhist, etc.). They also took a keen interest in the stories of the ancient rulers of the Subcontinent and their lifestyles, which were depicted through the various paintings and royal heirlooms. This shows that generally, the participants were interested in visiting a historical place. According to one student, however, the trip was rather rushed and he couldn't get much information about the artifacts he saw. He stated that the trip could have been improved by the presence of a tour guide who would have provided insights on the description and historical backgrounds of the museum artifacts. This shows that knowledge of an artifact's history is significant in enhancing the overall tour experience of the visitor.

Post-questionnaire

"I like how the models interact with the user." - P10

"The statues were amazing and it was fun playing that game." - P8 "The dialogues were very interesting and the 3d models were very nice and they could be rotated in all directions" - P4

"The UI was just brilliant and excellent and I found it really interesting." - P3

After using the app, our participants liked the interaction with the 3d models of the various historical characters from the Lahore Museum and their unique dialogues were also very fun. They could also rotate the models to get a 360 degree view and compare it to the images that were displayed before each 3d model appeared and conversed with the users. One user even commented on the user interface of our app, stating that it was great and enhanced his overall experience.

"I like the events/stories explained to us during the game." - P9

"I found interesting pieces of history" - P12

"I liked the historical details, which makes this application very informative and intuitive." - P2

The users also liked the historical details that were provided in the form of a story presented by the characters themselves. This way the users could learn more about the Sub continent's 1857 War of Independence in a more interactive and enjoyable way. This story-telling feature makes the app not just a game but also a history book and the quiz feature ensures that they retain the knowledge that they receive. That knowledge is then further tested through short quizzes on each character's story.

"The makhnu was cute." - P6

"I really liked the tour guide Makhno." - P11

"I loved Markhor." - P7

"I liked the letters that guided me to find my way." - P5

Another interesting feature of the app that many users loved was Makhnu the Markhor, who was the mascot of the app and also the tour guide who helps the user understand various steps and walks them through the entire journey. The guide is a markhor, which is the national animal of Pakistan (which a lot of users didn't know previously, but they do now, thanks to the app). They really liked Makhnu's adorable appearance and the expressions that he gives throughout the tour, depending on what the user has done (e.g a sad expression means the user lost marks on a quiz, a happy one means that the user aced the quiz and so on).

Interviews

"History is my favourite subject and I am especially interested in the Mughal Era. I even like to solve fun quizzes related to it online. I have never seen anything this fun and interesting to attract your interest to this subject. The 3D models were exactly that of paintings I have seen like that of Tipu Sultan. I would really like to play a similar tour for the Ottoman Empire so I definitely suggest making one." - P2

One of our participants (P2) was absolutely delighted to see our app and loved the whole tour of the Lahore Museum including the 3D models, their story related to the 1857 War of Independence and even the quizzes. His response especially validated our preferential measure on **perceived usefulness** as he foresaw and indicated great potential in the application. He very much liked the subject of History and also solved quizzes online, so this app provided him with a good combination of historical stories, artifacts and quizzes in a single platform. Moreover, he suggested that our app should introduce more tours related to the Mughal Empire as that era was close to his heart, and also asked us to introduce tours related to other dynasties like the Turkish Ottoman Empire. These are some of the features that we will soon add to our app to increase the overall experience through several models and exciting stories. He also liked the 3D models of the app and thought they were very accurate in terms of their physical appearance, in comparison to the artifacts displayed in the Lahore Museum.

"It really needs a better tutorial and I was confused why I had to drag exhibit B to C to talk to it which was not the case. But still it's an amazing application and has a lot of potential. The UI is simple, minimalistic, straightforward and not cluttered." - P12

"It is an amazing application. I would really just recommend a good tutorial for the AR feature where I got confused as there are some people who are not informed about those things. I had forgotten what AR is in itself. Everything else was great." - P3

We noticed through our performance measure related to time-based efficiency, that generally the users took more time while setting up the exhibits and scanning them to find the models. One member also mentioned that the app should include a tutorial in order to explain more about the sheet printing, cutting, scanning and usage in order to interact with the models. He was confused after the first exhibit and didn't know that he had to drag the scanner to the second marker in order to proceed to the next model's story. One reason for this ambiguity was the online scanning process as a lot of users didn't have printers at home, so we had to send images of the sheet to them and they had to scan the pictures of the markers (in their laptops) through the app scanner in their phones. Thus, the experience would be a lot easier if the users have printed copies of the markers and scan them through the app. Overall, the user thought that the user interface was

simple, easy to use and the options were clearly visible. Furthermore, he stated that the app's idea was really unique and interesting, and with a few improvements, the app could do really well as it has a lot of potential.

"The mascot was very cute. But the models did not look real which I did not like. The game was fun to play but some of the words were difficult to understand for me." - P6

"I liked the cartoon models. The quiz was very hard. I could not understand what the models were saying." - P7

According to some users the language used by the models while narrating the story included high vocabulary which made it difficult for them to understand what the models were saying and similarly, the quizzes were also very tough for them. A possible reason for this could be the age of the users or the grade in which they study. To improve the storytelling, we could introduce animated videos or images that would be played/shown while the characters narrate their stories. This would help the users understand and visualise the content in the stories in an even better way and they would be able to retain that knowledge, and would thus, be able to perform better in the quizzes as well. The app's tour guide Makhnu was liked by almost all of the users and they thought he was very cute and adorable. As for the 3D models, the users had conflicting views - some thought the animations were very well done and accurate in resemblance to the actual museum artifacts, while some users thought that the models needed some improvement in terms of their colours, texture etc. This could be improved in the future by spending more time on all the minor details of the statues to enhance the overall appearance of the models or visiting the Lahore Museum itself to get a closer look at the artifacts in order to imitate them (as 3D models).

"I really like the concept of application, although I have learned this all during my O'Levels but learning and experiencing through this way is quite interesting and engaging. I liked how the models interact with the user and the contextual words are really appealing like Afreen and Namaste." - P10

Users really liked the overall concept of the app (history, story, tour, artifacts and quizzes), especially the story as it was taken from their O level's Pakistan Studies syllabus. So, they could use the app in order to grasp the information related to the Indian War of Independence, 1857, as the story provided by the app covered the chapter related to the war in their history books. This way, they could learn their course chapter through interaction with fun characters involved in the story, which as a result, would help the users retain knowledge more easily through visualisation and quizzes. Moreover, the users found the dialogues spoken by each character to be unique and distinguishing. For example, the Mughal emperor Bahadur Shah greeted the user by saying 'Salam' while Rani Lakshmibai of Jhansi said 'Namaste'. These minor details added extra value to the overall user experience.

Suggestions

We obtained various suggestions from our post-questionnaire when the users were asked to evaluate the application and suggest what could have been done better for us by providing suggestions for the application as well as us in general.

"AR features and the 3D models can be improved further. Except that the overall app is superb." - P1

"There should be options(like arrows) to easily rotate the 3D models." - P11

"Please make it more detailed and add more civilisations (Ottoman, British, French, etc.)" - P2

"It should include other tours too and also the models should be coloured." - P4

According to the users, the 3D models need some improvements, with more realistic detail in structure and textures. Moreover, arrows should also be added with the models to control their movement as it was hard for some users to rotate the models using their fingers (especially if they used an image of the sheet on their laptop for scanning rather than printing and cutting the sheet, and then scanning it.). Some users also suggested that the app should introduce tours to several other historical places and should also have tours related to other civilizations and empires (French, Ottoman, etc) apart from the ones related to the Subcontinent only.

"Make the models more real. English was difficult. Quiz was difficult. Models talk too much." - P7

"The English was a bit too advanced for me." - P6

According to some users, the vocabulary used by the models while telling the story was harder and so they faced difficulty in understanding the meaning of some parts of the story. Reasons for this could be the age of the user or the grade in which they study (i.e, a 7th grade student would have more advanced vocabulary as compared to a 4th grader). A user also faced difficulty while attempting the quizzes and thought that the questions were tough. A reason for this could be the user's lack of attention while listening to/ reading the models' stories. Moreover, one user stated the models talked too much. A reason for this could perhaps be the user's lack of interest in history or stories (novels etc) in general. We found out more related to this problem through the following performance measure: the time taken by each user to complete the tasks, the final tour score and the number of times a user called the interviewer for help. Some users who were in grade 5 or below, took longer time to read the stories as compared to users who were older in age. Their tour scores were also low as compared to older students. This showed us that there was a difference in the understanding of the story by the users in terms of the level of language and vocabulary. Also, some users called the interviewers for help, to learn the meaning of certain words and phrases.

"Maybe a user guide that opens as a pop-up because most people do not refer to FAQs unless prompted." - P10

"I think there can be a short tutorial for how to interact with the exhibits and the 3d models." - P12

"A better guide could be in place for people not that informative about Augmented Reality." - P3

Some users felt that the app should provide more guidance in order to make the overall experience better and easier. According to them, there should be a short tutorial in order to explain how the user needs to interact with the 3d models and exhibits. Moreover, there should be a section which provides information related to Augmented Reality and explain how it is used in the app as a lot of users aren't familiar with the concept of AR or VR. Although the app does contain an 'About Us' section which provides its description and an 'FAQs' section which answers the general queries related to its usage, a lot of people do not refer to these sections. We found out more about this problem through the following performance measure: number of times the user called the interviewer for help. Users mostly got stuck in the 'Get Sheet' segment and asked for more information on how to go about the steps related to sheet scanning and finding the correct exhibit. Even though all their queries were answered in the FAQ section, they still asked the interviewers for help.

"The application shouldn't take so much time to continue." - P8

"App should contain a notifications tab so the user gets notified about the tours. App should close immediately when wanted. This app should contain settings options so users can mute the sound of the keyboard if they want to." -

P5

Some users mentioned that the app takes more time to load when one character's story ends and the next character appears in order to tell his/her story. One user stated that the app should immediately close when the user wants to exit rather than asking again if the user wants to quit. However, this option will be useful for those users who accidentally click the 'back' option but don't really want to exit the tour, so in that case this option would save them from the trouble of replaying the whole tour (in case they are in the middle or near the end). Another user pointed out that the app settings should contain the option of increasing/ decreasing the volume. This is a very useful feature which we plan on including in the future. One user also mentioned that the app should provide notifications to users regarding the addition of any new tours or any updates in the app's features. This notifications option is something that we plan on adding once we add other tours and models to our app.

4.3 Observations

This section includes the observations that the conductors made when testing the application. This could include emotional responses of the users while using the application as well as issues that the conductor found while conducting the testing session.

Issues faced during the testing consisted of two types of issues. One type consisted of glitches and bugs present within the game. These consisted of the following:

1. The user was unable to find a direct way to return to the main menu after clicking the start button of a stage. This would not work even after pressing the return button on the phone. Thus, the user could not undo an action that they made by accident.
2. When conversing with a model, the dialogue of the model does not disappear if the user stops pointing their phone at the model's marker. This breaks the immersion of the app as the dialogue is now not associated with any model.
3. Interactions with any model are not possible while the dialogue box is present on the screen.
4. Tipu Sultan informs the user to meet with Bahadur Shah II by presenting an image of Bahadur Shah II. However, in specific conditions, when the user repeats Tipu Sultan's quiz, the image of Bahadur Shah II does not disappear from the screen after the user is finished talking with Tipu Sultan.
5. When an exhibit asks the user a question, the answers appear before the question is completely written. In case the user selects an answer before the question is completely printed, the remaining text of the question is mixed with the proceeding dialogue which causes the dialog to appear as a mixture of random alphabets.
6. Some dialog was repeated randomly by Makhnu the Markhor under specific conditions after the user incorrectly answered an exhibit's question.
7. Users had the ability to translate the models. Sometimes, they accidentally moved the model off-screen which made them believe that the model was not being captured by the application in the camera whereas it was just off-screen.
8. Users were often confused about how the game only works in vertical orientation. They attempted to play the game while holding the device horizontally, however the game did not change its orientation.
9. In case multiple markers were present on the screen randomly, only one would display its exhibit which would cause issues for the user when they would want to interact with a specific model and a different model would be shown.
10. The transparent locks which depicted that an exhibit was locked were not clearly visible within the camera which is why the user would not see it or interact with it.

The other type of issues that the conductor witnessed were set up and understanding issues that the users faced when trying to understand the functionality of the application. These consist of the following issues:

1. Some users did not tap on the last dialog of a model or Makhnu which meant that the dialog would stay present on the screen. In this scenario, the dialog box prevented the user from interacting with the environment. Thus, they could not figure out why they could speak to the other exhibits after passing a specific one
2. One of the tasks involved asking the user to interact with the model. In this scenario, they were not aware of what interactions they could attempt on the model other than tapping it. They were thus told by the facilitator that they could pinch to change the size of the model or rotate two fingers to rotate the model.
3. Each exhibit, such as Tipu Sultan or Rani of Jhansi, would ask a single question from the user to verify if the user had properly listened to the previous exhibit's story. However, this question and its options would suddenly appear in the dialog without any warning dialog and would cause the user to accidentally tap on an answer thinking it is a dialog.
4. Most of the users from the smaller age-groups found it difficult to interpret the dialog due to the use of some advanced words and phrases.
5. Most parents did not understand that one of the interesting elements of this game was to place the markers afar or in hidden locations which would make it more interesting for the children to go and look for them similar to a scavenger hunt.

6. Some users could not properly understand the gestures required to rotate a model.
7. Very few users did not know that they needed to tap on the dialog box to advance the dialog.
8. Users who scanned the sheet from laptop or monitor screens which were vertically aligned could only see the model from a top-down view. They proceeded to try to rotate the model so that it was seen from the side rather than the top, however the application did not support this form of 3D rotation.
9. Some users did not like the dull colors of the models and wanted more saturated and realistic depictions of the models.
10. Users felt that the face of the 3D models were of low quality and not clearly visible when the model was zoomed in which broke their immersion.

The conductor also observed the emotional responses of the users to various parts of the application. This was vital to understand what functions were interesting to the users and which were difficult to understand for the user or made them annoyed. Most users had a smile on their faces and were excited whenever they saw Makhnu the Markhor, specifically its animation at the end of a stage. They adored the design of Makhnu and felt a more personal relation with it due to its various reactions. More interested expressions were seen on the user's faces when they saw the models for the first time in their own rooms. Some were so excited that they decided to explore the model in detail on their phones. However, there were some aspects of the application that made the users annoyed. Some users were bored and even sighed because the dialogs of the models were too long without any reaction or animation which made the dialog similar to a textbook. Some users were also confused at the vocabulary since they were young and could not understand the meaning of some advanced words. Overall, the emotional responses provided us with clues on what changes to make which could lead to better and more positive emotional responses from the users.

5. Discussion (about evaluation)

We primarily noticed that the application was given more preference by an audience slightly above the age group we had imagined because of the interest required for the story content and quizzes as well as the difficulty of language. Thus, in order to improve that we could include more imagery in the story telling part. While the models narrate the story, the user would be able to see images of the historical event in the form of an animated video or images. This way, users of all age groups would be able to understand the stories and would be able to solve the quizzes in a better way. Moreover, we could also include an option of 'level of difficulty/vocabulary' in the settings, which would show 'easy', 'medium', 'hard'. Then students in grade 5 or below could select the 'easy' or 'medium' option which would then show the same models and stories but the english would be easier, according to their level of understanding.

We also noticed a general lack of exposure and understanding to Augmented Reality. More suggestions asked for improvements to be made towards the interaction features as they were taken for granted due to this assumption. more tutorials should be based around the interaction part to improve usability. Similarly, the performance metrics validated the usability of the application in general but there is still room for improvement for some of the primary tasks by inclusion of more intermediate tutorials as even suggested by our participants themselves. Regarding our metric on realism and accuracy of the 3D models judging their aesthetics, there is also more room for improvement and they can be reconstructed in both structure and texture to increase the heuristic of visual appeal for children of our targeted age groups.

Even though all their queries were answered in the FAQ section, they still asked the interviewers for help. This showed that they quickly forgot that the settings included the FAQs option. So, for that, we could provide an 'info' option to the users for quick reference throughout the tour. This way if the users get stuck during any part of the tour, they would easily click the always available, 'For more info' option on the top right side of their screens and would be able to see the FAQs directly. We also found that some users had trouble with the sound/volume and were distracted by the typing sound when the models were narrating their story through the dialogue boxes.

They had to close the app and go to their phones' settings option in order to eliminate the keyboard sounds. In order to overcome this issue we plan on adding a 'volume' option in the settings which would then allow the users to easily increase or decrease the app volume.

Another thing we found was that some of the users were very much interested in historical civilisations and empires and history in general. So, in order to enhance their overall experience with our app, we plan on introducing more tours related to other historical places like Badshahi Masjid, Sheesh Mahal, Lahore Army museum etc, where the users will be able to interact with more models and listen to exciting new stories. Besides this, we also plan on introducing other empires and civilisations apart from those related to the Sub Continent (e.g. Ottoman Empire, Chinese Dynasties, ancient Egyptian civilisations, etc.). Each empire/ civilisation will then have its own tours, models and stories. Thus, this evaluation has indicated that there is potential for expansion of the application both vertically and horizontally. Consequently, as a future implication, we believe that further consideration should be given to this domain of designing mobile augmented reality applications utilizing gamified stories for delivering historical and educational virtual experiences to children through similar interactive learning approaches.

6. Conclusion

With the restriction on physical trips and outdoor activities and the lockdown of schools in the COVID-19 pandemic, children have been confined to their homes for months without any school or family trips. The lack of active learning in schools and over dependence on the classical form of teaching has left a void in the learning experience of children and led to a declining interest in history related subjects.

It was imperative to provide a solution that can provide the children with an opportunity to go on trips and at the same time, learn from their experience. After an extensive user research, design and development process throughout the semester, we are proud to provide a solution that provides an opportunity to the parents to be a part of the learning experience of their children and keep a track of their children's performance. Apart from providing a fun and interactive platform to go on trips virtually, the mobile application also provides a learning aspect for the children. Thus, it combines tours, stories and quizzes,

The evaluation phase helped us identify both the positives and negatives of the application, even given the constraints of the partially virtual interviews. Most users appreciated the features and functionality provided by the app. They admired the use of Augmented Reality and 3D models to provide a realistic experience to the users. There were also certain complications identified by the users that proved to be vital contributors to the overall user experience. Marker identification was somewhat complicated. In case a quiz was not successfully completed, some users found it difficult to identify which marker to return to. Multiple usability heuristics and measurements were tested including efficiency, effectiveness, learnability and others. The positives and negatives of the app were identified and mentioned in the document. The suggestions from the surveys were also identified as vital for improving the application and making it better for the overall user experience. Thanks to the feedback by the users and our own observations, we will implement several changes in the app and include various exciting new features to make the app more fun and interesting.

7. Appendix

7.1. Figures for Surveys

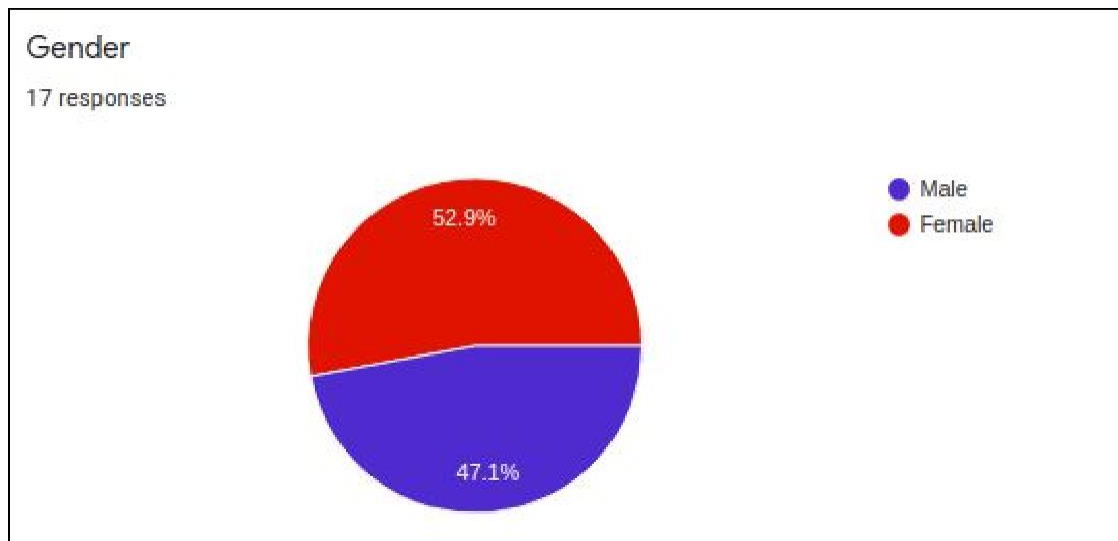


Fig 7.1.1. Respondents Gender

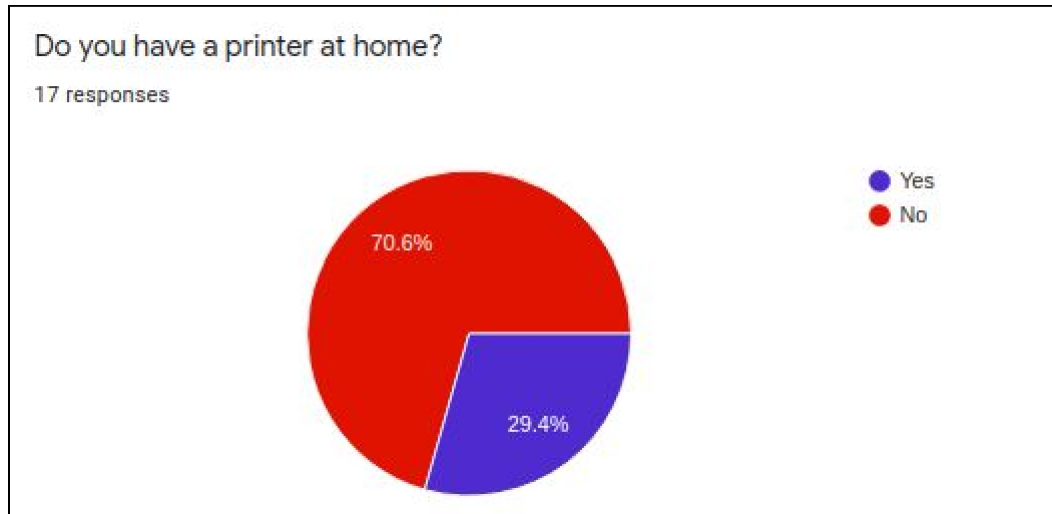


Fig 7.1.2. Printer Access

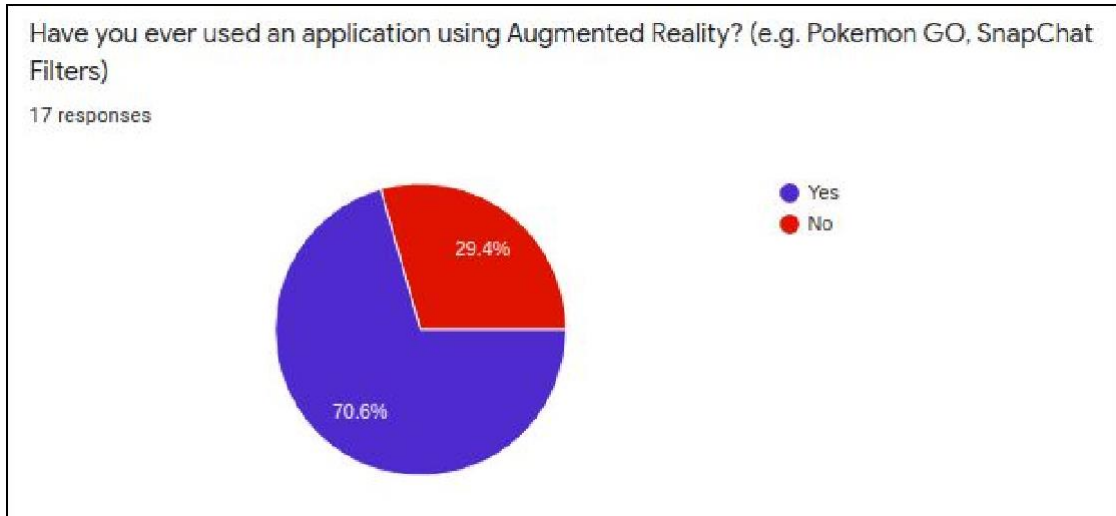


Fig 7.1.3. Interaction with AR based applications

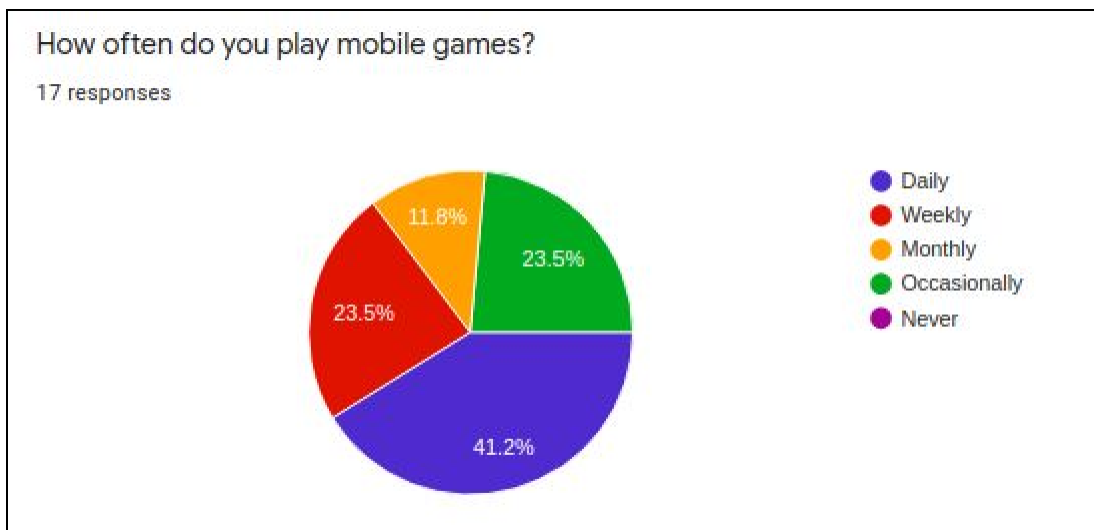


Fig 7.1.4. Respondents playing mobile games

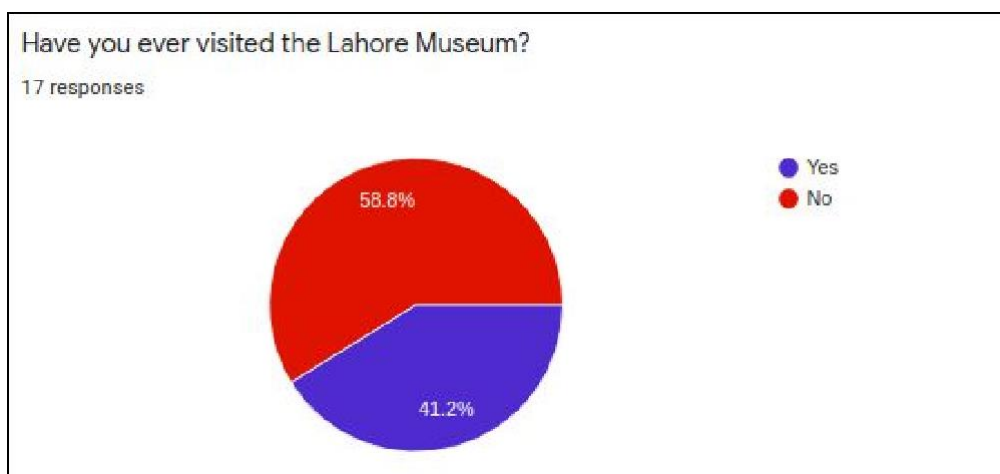


Fig 7.1.5 Visit to Lahore Museum

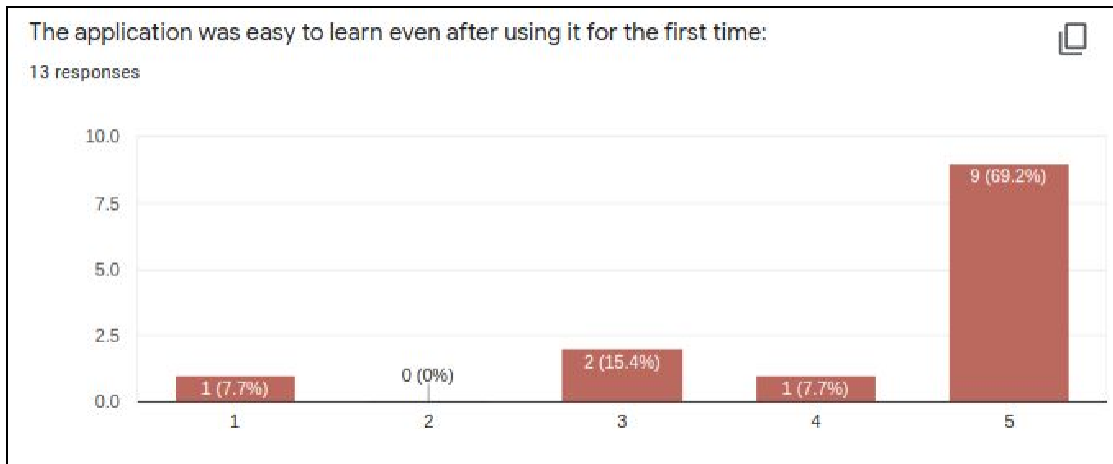


Fig 7.1.6 Ease of use of application

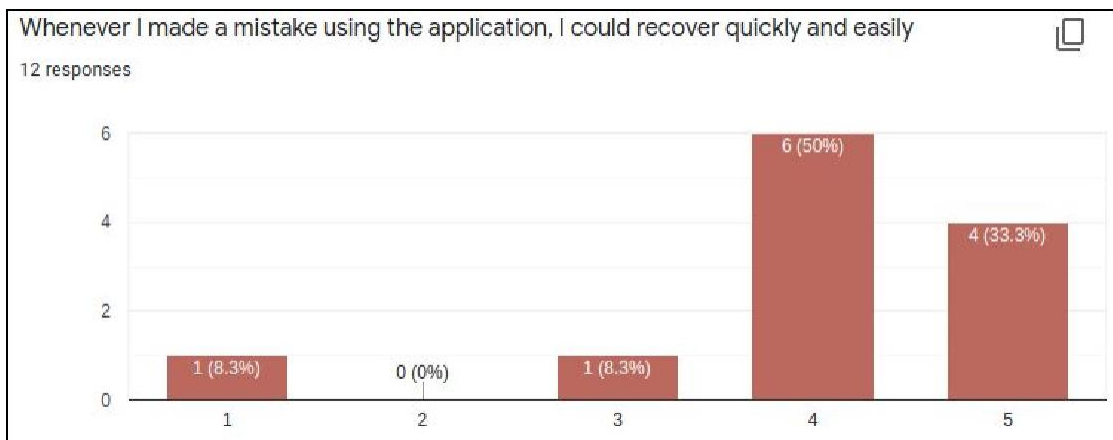


Fig 7.1.7 Recovery from mistake

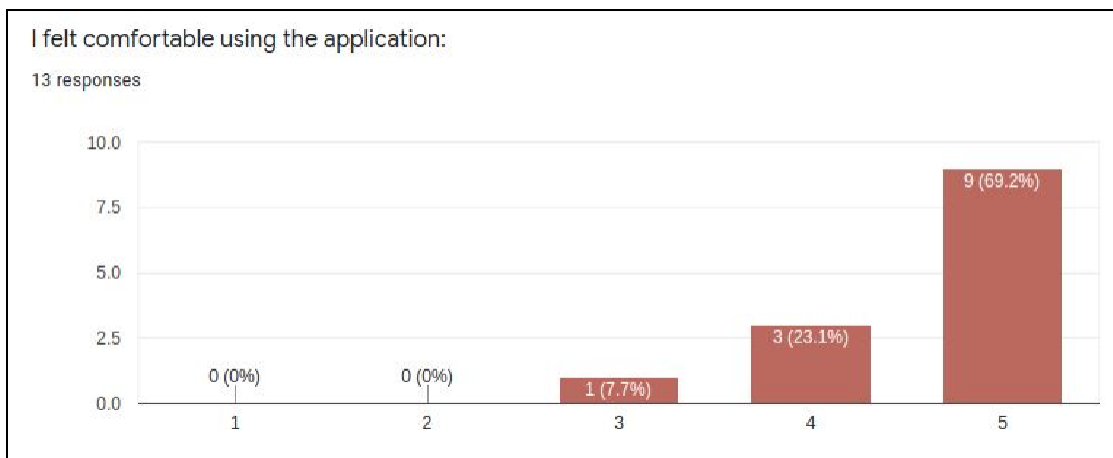


Fig 7.1.8 Comfort level when using the application

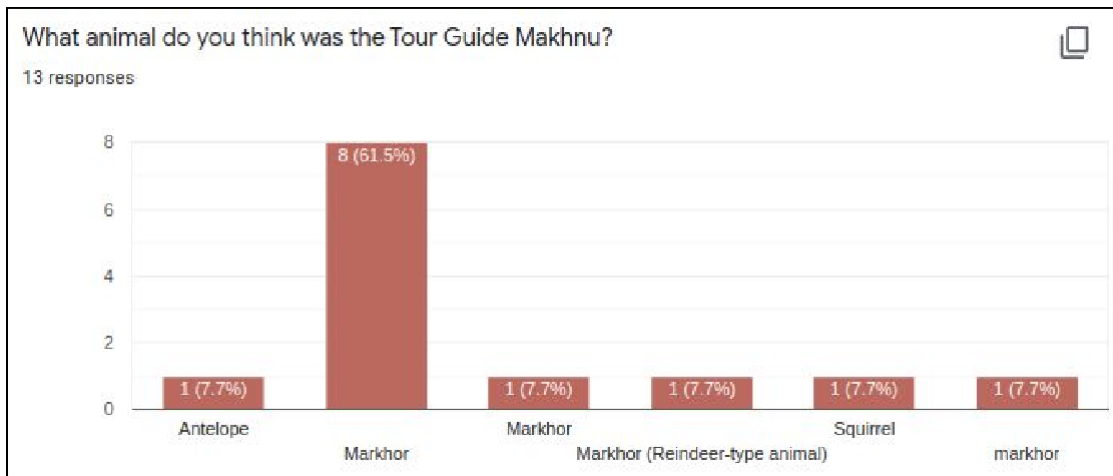


Fig 7.1.9 Identification of Mascot

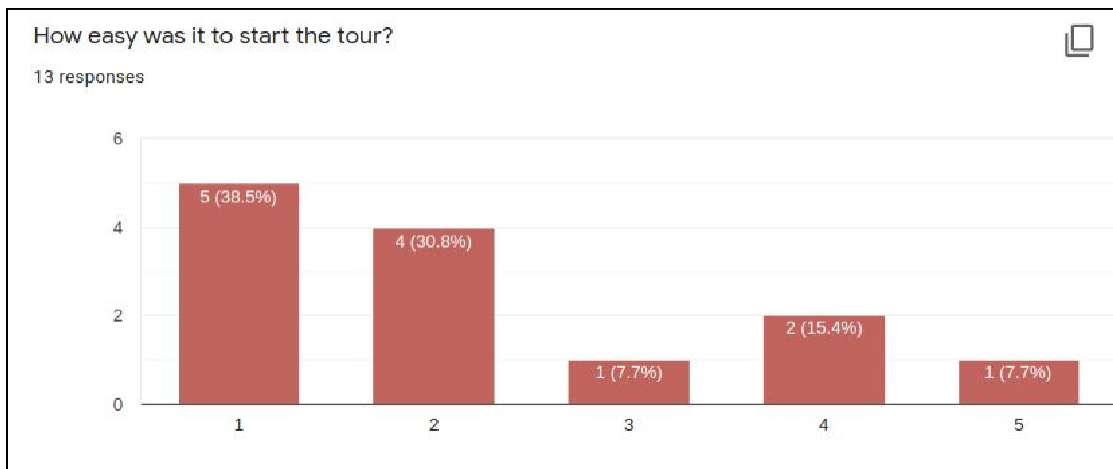


Fig 7.1.10 Ease to start a tour

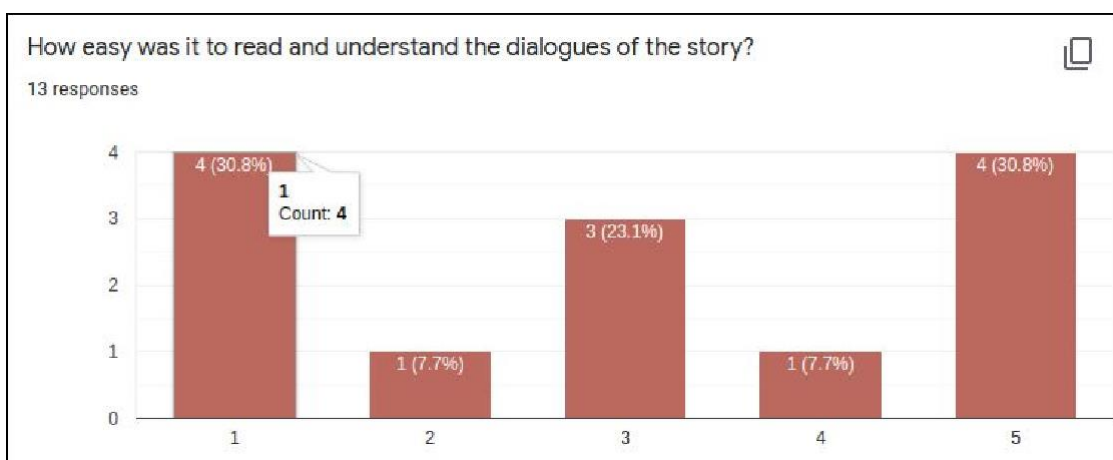


Fig 7.1.11 Ease of reading and understanding the story

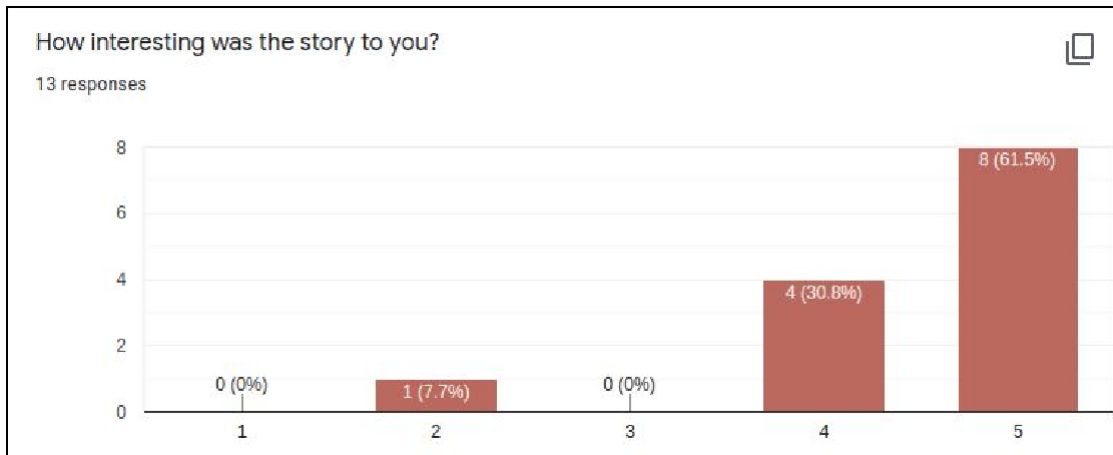


Fig 7.1.12 How interesting was the story

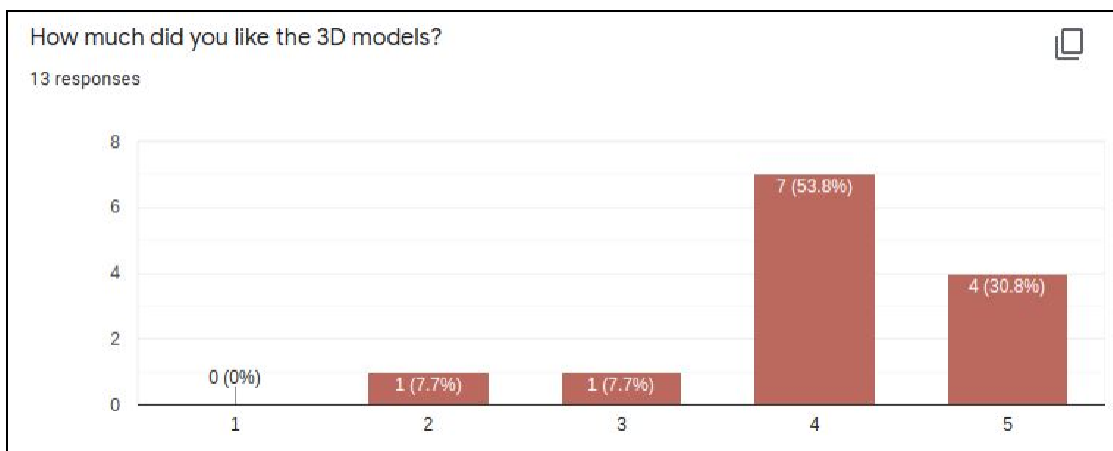


Fig 7.1.13 3D Models

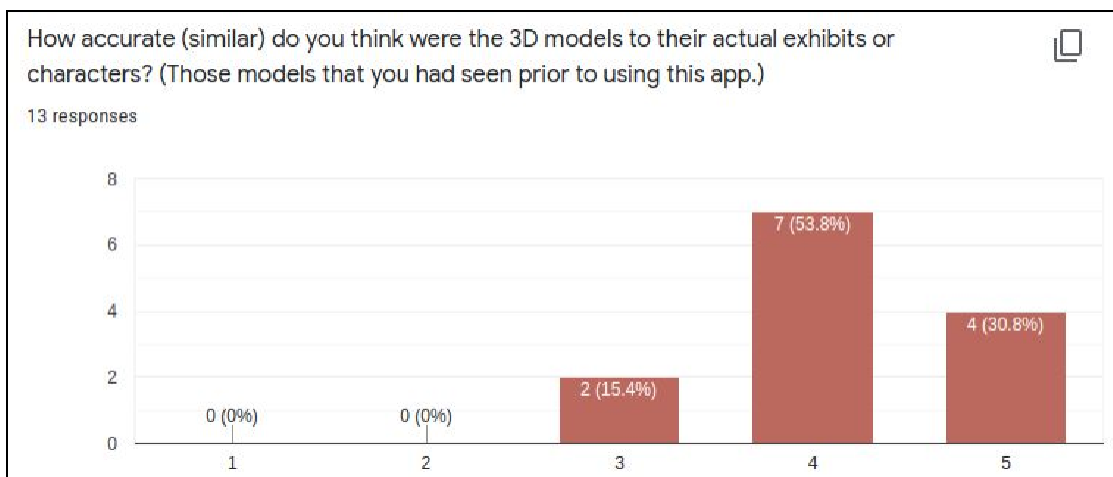


Fig 7.1.14 Accuracy of 3D Models

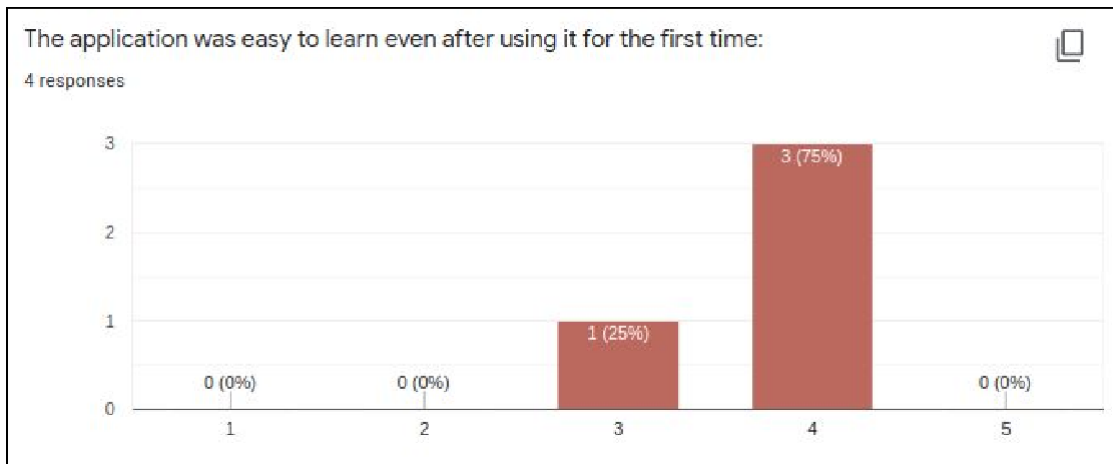


Fig 7.1.15 Ease of use of application (for parents)

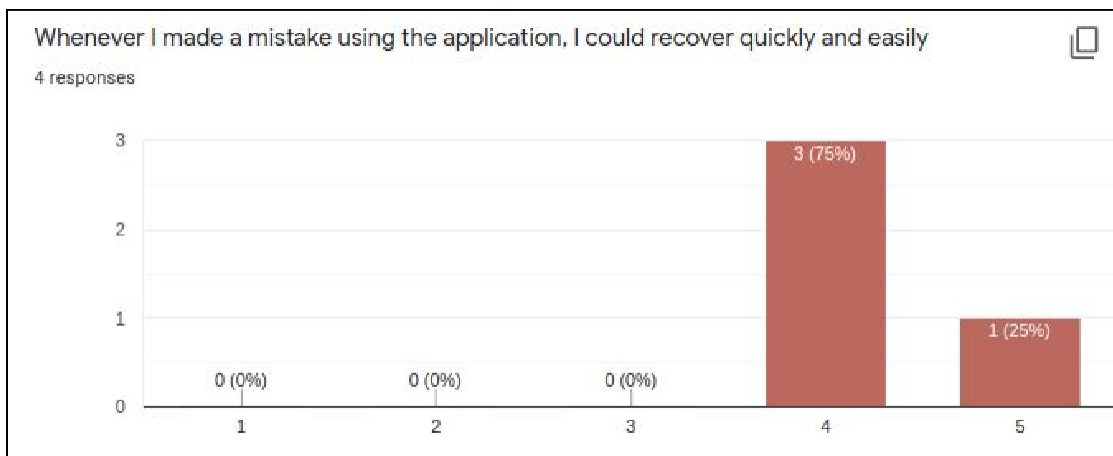


Fig 7.1.16 Recovery from mistake (for parents)

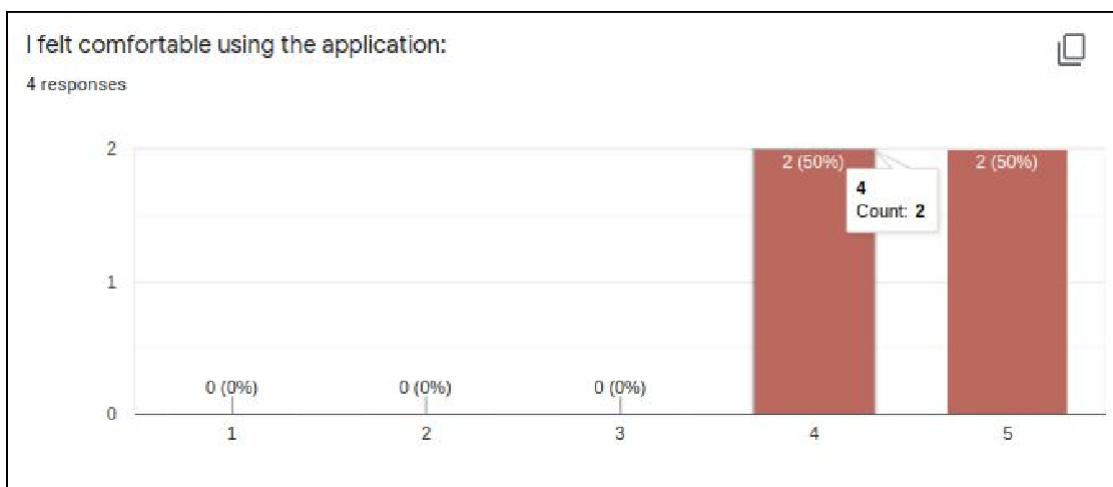


Fig 7.1.17 Comfort level when using the application (for parents)

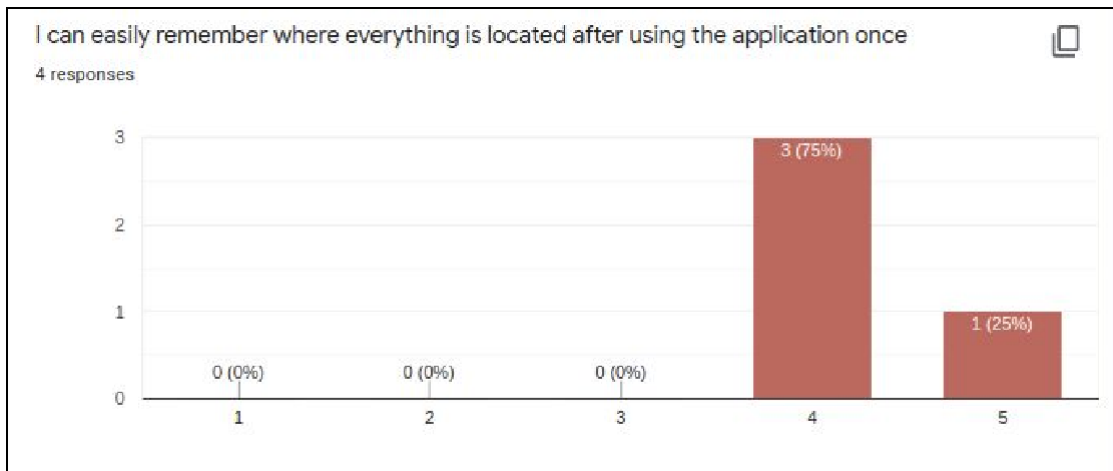


Fig 7.1.18 Remembering location of things

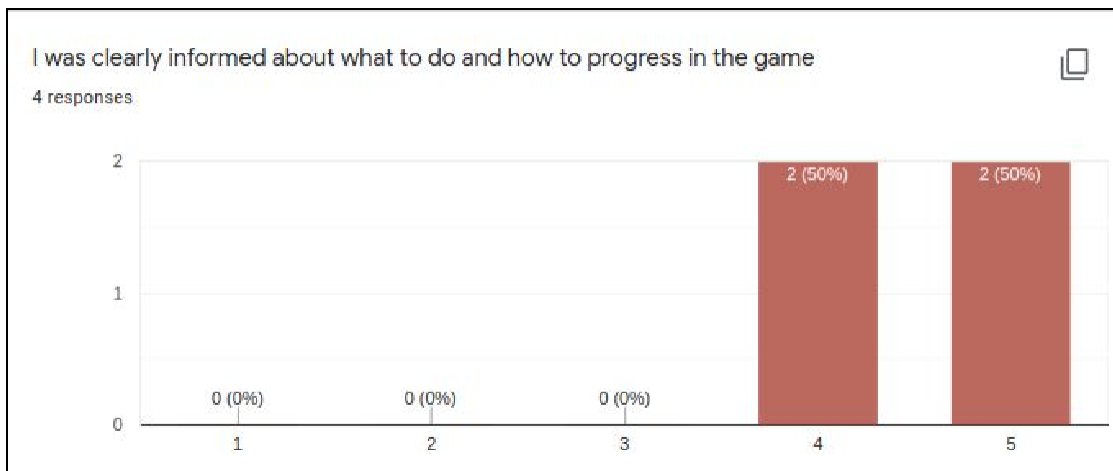


Fig 7.1.19 Progression in the game

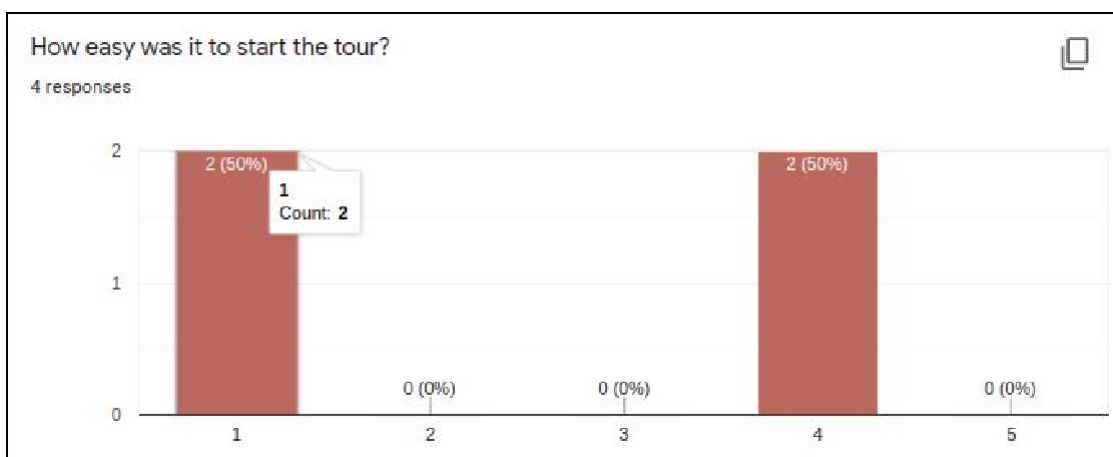


Fig 7.1.20 Ease to start a tour (for parents)