Northeastern University

Khoury College of Computer Sciences

Usability in Holographic Augmented Reality applications

Author: Marjan Moshfegh Gohari

Supervision: Prof. Yvonne Coady

December 24, 2021

Contents

1 ABSTRACT **1**

2 INTRODUCTION **1**

3 BACKGROUND **2**

3.1 User Experience and Usability 2

3.2 Usability evaluation 2

3.3 Evaluation method 2

3.3.1 Heuristics evaluation 3

3.3.2 Cognitive walk-through 3

3.3.3 Laboratory observation 4

3.3.4 Questionnaire 4

3.4 Evaluated application 4

4 RELATED WORK **4**

5 METHODOLOGY **5**

5.1 Heuristic evaluation 6

5.1.1 Participants 6

5.1.2 Procedure 6

5.1.3 Result Analysis 8

5.1.4 Finding 9

5.2 Cognitive walk-through 9

5.2.1 Participants 10

5.2.2 Procedure 10

5.2.3 Result analysis 10

5.2.4 Finding 11

5.3 Laboratory observation 12

5.3.1 Participants 12

5.3.2 Procedure 12

5.3.3 Result analysis 12

5.3.4 Finding 14

5.4 Questionnaire 16

5.4.1 Participants 16

5.4.2 Procedure 16

5.4.3 Finding 17

6 SUMMARY **18**

7 CONCLUSION **20**

8 FUTURE WORK **20**

1 ABSTRACT

Augmented Reality is the recent media that is enabling special human-computer interaction experiences. The interactive 3D holographic system is an AR technology that provides users with a hybrid visual experience in which 3D Hologram becomes a part of the real world. As AR applications are relatively novel, most targeted users are novices. Hence, for these users, several actions in the AR apps can be confusing, and they might find it difficult to use them. Moreover, as we move forward with AR technology, understanding how users interact with the holograms is essential. Therefore, usability evaluation plays a crucial role in accepting this kind of application by the targeted group. Since most relevant studies have focused on technological developments and usability of holographic AR applications has not been discussed frequently in the literature, there is a need to explore their usability problems and define user experience guidelines. Usability testing is considered a central phase of User Experience (UX) research and consists of verifying three main variables used to measure usability (efficiency, effectiveness, and satisfaction). This study aims to evaluate the Usability factors and explore usability issues in Holographic Augmented Reality (AR) applications using *HoloX* as a pilot. This study conducted four usability evaluation methods to obtain the more accurate result: cognitive, walk-through, heuristic evaluation, laboratory observation and questionnaire. We have collected qualitative and quantitative data during this process and analyzed and interpreted the collected data, aligning with end-users ' needs and requirements. The outcome obtained from the four usability methods has been represented in this document. Also, a set of design guidelines for AR applications has been proposed. This research also will provide guidance on: What types of data should be collected and how that data can be analyzed.

2 INTRODUCTION

Innovations in display technologies created new experiences of human–computer interaction. Augmented reality (AR), and holographic projection are of these technologies providing users new three-dimensional (3D) visual experiences in real environments [1].

They try to integrate the virtual and real-world together. AR Uses various computer vision and machine learning algorithms to analyze what the user sees to enhance what is relayed to the user more realistically [2].

In recent years, holographic AR applications have gradually developed for mobile devices. As these applications are still a new emerging technology and the target users are novices, they find it difficult to interact and treat with Hologram. Therefore, usability evaluation plays a crucial role in accepting this kind of application by the targeted group. Relevant studies mostly have focused on technological developments [3] and the number of researches that have explored the user's experience or usability problems of holographic AR applications is relatively low. Hence, there is a need to explore their usability problems and propose user experience guidelines. The present academic research was conducted to measure the usability factors and explore the usability problems of HoloX to pilot Holographic AR applications.

Considering the literature, we combined and reinterpreted some usability measures for an augmented reality context by adapting the ISO 9241-11 and Nielsen Heuristics. Then, it conducted follow-up observation and questionnaires. An evaluation is more valid if the evaluation was done in multiple methods in order to confirm the result of the evaluation. Hence, four methods have been chosen to analyze the targeted AR application out of the existing usability evaluation methods. In order to cover all the aspects of usability methods, two methods from usability inspection (namely cognitive walk-through and heuristic evaluation), one from usability testing (laboratory observation) and one from user reports (questionnaire) have been chosen [4].

The research questions answered in this study are as follows:

1. To measure Usability factors based on effectiveness, efficiency and satisfaction?
2. What are the usability Problems experienced by users while using holographic AR applications (HoloX)?
3. What are the design guidelines to improve the holographic AR applications (HoloX)?

The following part of this paper will provide a background of usability and user experience, usability methods adapted in this project and evaluated application. The literature review section reports prior research on usability evaluation for AR applications. The methodology section describes how the research was designed and executed. The result section elaborates the results and analysis of HoloX in terms of usability based on effectiveness, efficiency, and user satisfaction. Finally, in the conclusion section, the conclusion with future research directions is provided.

3 BACKGROUND

3.1 User Experience and Usability

ISO defined user experience as a "person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service." It includes the whole experience of using the system, such as making assessments about user preferences, even identifying moments of delight or frustration [5].

Usability is part of the broader term "user experience" and refers to the extent to which a product can be used by specified users to achieve specified goals. ISO outlines three measurable attributes to define a usable system: System Effectiveness to examine the users' ability to achieve goals the given tasks, System Efficiency to examine the required user resources to achieve goals in the given tasks; and System Satisfaction to record the users' opinions and positive attitudes towards the use of the system [6].

3.2 Usability evaluation

Usability is evaluated by the quality of interaction between a system and a human reaching back to human-computer interaction (HCI). Various usability evaluation methods have been proposed by the research practitioners. These can be classified as inspection, testing and inquired methods. Inspection methods focus on the user interface, testing methods focus on task performance and inquire methods focus on user data [7].

3.3 Evaluation method

The method that is used to measure according to Dunser and Billinghurst [7]:

1. *Objective measurements methods:* result from a repeatable and reliable task of quantitative observations. This method measures: accuracy (average of error), object or user position, time (task completion times), scores from the test, etc.

2. *Subjective measurements methods:* the result based on the personal subjective judgement from the user, and use questionnaires: grading, rating, or depth judgement.

Four methods among all usability methods have been chosen for this study. The definition of the selected methods are as below:

3.3.1 Heuristics evaluation

Heuristics evaluation is one of the Subjective measurements methods introduced by (Nielsen and Molich 1990), which involves usability experts inspecting the user interface based on a set of heuristics to identify usability problems. In this method, one or more evaluators are recruited, and they should be a novice to the system. These principles are often called heuristics because they are more in the nature of rules of thumb than specific usability guideline [8].

1. *Visibility of system status:* the application should keep the user informed during the interaction.
2. *Match between system and the real world* The application should Follow real-world conventions, making information appear in a natural and logical order
3. *User control and freedom* the application should provide freedom for users to perform actions and undo incorrect actions.
4. *Consistency and standards* the application should have a consistent interface layout and user interaction.
5. *Error prevention* The application should avoid mistakes and prevent undesired action.
6. *Recognition rather than recall* It should be easy to memorize how to work with the application.
7. *Flexibility and efficiency of use* It should be easy to interact with the application for novice users.
8. *Aesthetic and minimalist design* the application should not show irrelevant or rarely used information to the user.
9. *Help users recognize* The application should diagnose and recover from errors.
10. *Help and documentation* The application should have a brief and easy to find tutorial for first-time users.

3.3.2 Cognitive walk-through

Cognitive walk-through is another type of Subjective measurements methods used to examine the interaction between the user and the interface through some pre-defined tasks. They are designed to see whether or not a novice user can easily carry out the defined tasks and subtask without any prior training. Then an evaluator (novice to the user interface)attempts to answer the following four questions for each correct action involved in a task [9, 7] :

1. Will, the user try to achieve the right effect? (Conceptual model)
2. Will the user notice that the correct action is available? (Visibility of actions)
3. Will the user associate the correct action with the desired effect? (Labelling)
4. If the correct action is performed, will the user see that progress is being made towards the solution of the task? (feedback)

3.3.3 Laboratory observation

Laboratory observation is one of the objective measurements methods. In this method, users will be asked to complete the pre-defined tasks in a controlled environment while observers watch and take notes or do screen recording. The goal is to identify any usability problems (qualitative data) and measure the System Effectiveness to examine the users' ability to complete the given tasks, System Efficiency to examine the required user resources to complete the tasks, and System Satisfaction to record the users' opinions and feedback [7].

3.3.4 Questionnaire

Questionnaire is the other subjective measurement method that involves gathering targeted users' preferences and opinions about the user interface related to the operability, effectiveness, understandability, and aesthetics of the user interface [10]. Typically, the evaluator first asks participants to use the product and then fill up or rate pre-defined questions. Users are often eager to please, eager to seem smart, and quickly forget when things have gone wrong if they achieve a goal. These very pleasant tendencies can give way to misleading biases. UX decisions can not be made based on users' reporting alone. Hence, we decided to add observational studies to our UX research methodology.

3.4 Evaluated application

HoloX is the selected application for usability evaluation. HoloX Holographic AR application that has been made by NexTech AR Solutions Company, available for iOS mobile operating systems. HoloX is an application that lets users create, share, and view human holograms on smartphone Devices (currently available for iOS devices). Users see 2D Human holograms as 3D by using Augmented Reality Technology. The application can be used for Announcements, Delivering a message, Virtual training, Product walk-through demos, In-store demonstration, etc.

4 RELATED WORK

Some researchers have done usability evaluation for the AR application with different methodologies and in different contexts. The following table shows some of the previous studies used in this research.

.

	Sang Min ,Won Chang 2013	Satu Elisa Schaeffer 2014	Hsinfu Huang & Chin-wei Chen 2019	Pier Luigi Ingrassia et al. 2020	Neha Tulia,Archana Mantrib 2021	Zamzuri Mohamad, Mohd Ramlie 2021
Research	**Usability Principles for AR Applications in a Smartphone Environment**	**Usability Evaluation for Augmented Reality**	**Assessment of usability factors in an interactive three-dimensional3d holographic projection system for experiential**	**AR Learning Environment for Basic Life Support Training: Usability Study**	**Evaluating Usability of Mobile-Based AR Learning Environments for Early Childhood**	**Examining the user experience of learning with a hologram tutor in the form of a 3D cartoon character**
Problem	As AR applications will require new evaluation methods, improved usability and user convenience should be developed.	Since AR is one of the new emerging technologies, the number of systematic evaluations done in AR interfaces is relatively low.	Innovative interfaces for the display and control of information constitute an essential topic for interactive experiential learning.	Cardiac arrest is a major cause of death worldwide. only a minority of victims receive CPR from bystanders.	As per the existing literature, there are no clear design principles that can be considered while designing mobile AR applications, particularly for kids.	However the choice of actual human hologram has its own advantages for tutor hologram, it's development is difficult and complicated.
Proposed Solution	Developing usability principles for the development and evaluation of smartphone applications	a systematic evaluation of the user interface	Developing an interactive 3D holographic projection system	Presenting an AR system for BLS training that offers realistic haptic feedback through a virtual scenario	Establishing the usability and suitability of the proposed application	Design a 3D cartoon-like hologram tutor and examine aspects of the user experience (UX) with usability test
Evaluation Methodology	**Analysing existing research to do usability principles for smartphone AR application Heuristic evaluation**	**Walk-through Laboratory Observation Questionnaire**	**Assessment of 11 usability factors by Questionnaire**	**Conducting SUS questionnaire for assessing the feasibility of the proposed tools**	Pre test and post test evaluation by analysis of the students' behaviour Qualitative method by teachers' interviews Quantitative method based on **SUS questionnaire**	The UEQ (**questionnaire**) was utilized after 30 minute teaching session to obtain UX feedback
Results	**22 usability principles were developed for user-centred design previous design principles were inadequate for newly developed technologies.**	Several design problems have been found. The results show that there is a need for the developers to revisit the application in order to enhance the design of the interface	The size of the projection object is a user's most crucial usability indicator, followed by cursor sensitivity, ambient brightness, and compatible gestures.	The results showed the proposed application feasible and acceptable as a tool for self-instruction training	The results showed that the developed ARLE is efficient, effective, and highly usable for the kids. Presenting **23 identified usability principles** to develop an interactive mobile-based AR application.	Overall, students feel that learning with the hologram tutor was efficient, practical and organized. And it has a significant positive effect on students' interest and motivation.
Future work	**To conduct a heuristic evaluation and usability testing on various AR applications. Include Heuristic evaluation of various kinds of smartphone device OS such as iOS and Windows Mobile.**	**To include tests with users that have not been instructed before using it. To carry out another session with users that are familiar with the concept of**	Future work can be directed towards accessible user interfaces and natural speech control interface designs.	The positive outcomes of this preliminary study make this prototype worthy of future testing.	Design and assessment of Augmented Reality applications for kids with the proposed principles	To find the best approach to using a 3D cartoon hologram tutor as a representative of the lecturer in the classroom.

Figure 1: Previous research on usability evaluation in AR applications

5 METHODOLOGY

Before conducting evaluation, the researcher applied for ethical research approval to ensure that research involving humans meets high scientific and ethical standards that respect and protect the participants. Human Research Ethics Board reviewed the disciplines, fields of research, and research methods in the University of Victoria. The researcher asked participants to sign the consent form based on the obtained ethical approval before starting the evaluation. Human Research Ethics Board reviewed the disciplines, fields of research, and research methods in the University of Victoria. The researcher asked participants to sign the consent form based on the obtained ethical approval before starting the evaluation. In our methodologies, the process of usability evaluation in each method is: Plan, choosing the participant, performing the test, analyzing the result, documenting the finding. The following form is the usability test plan which has been used for each method:

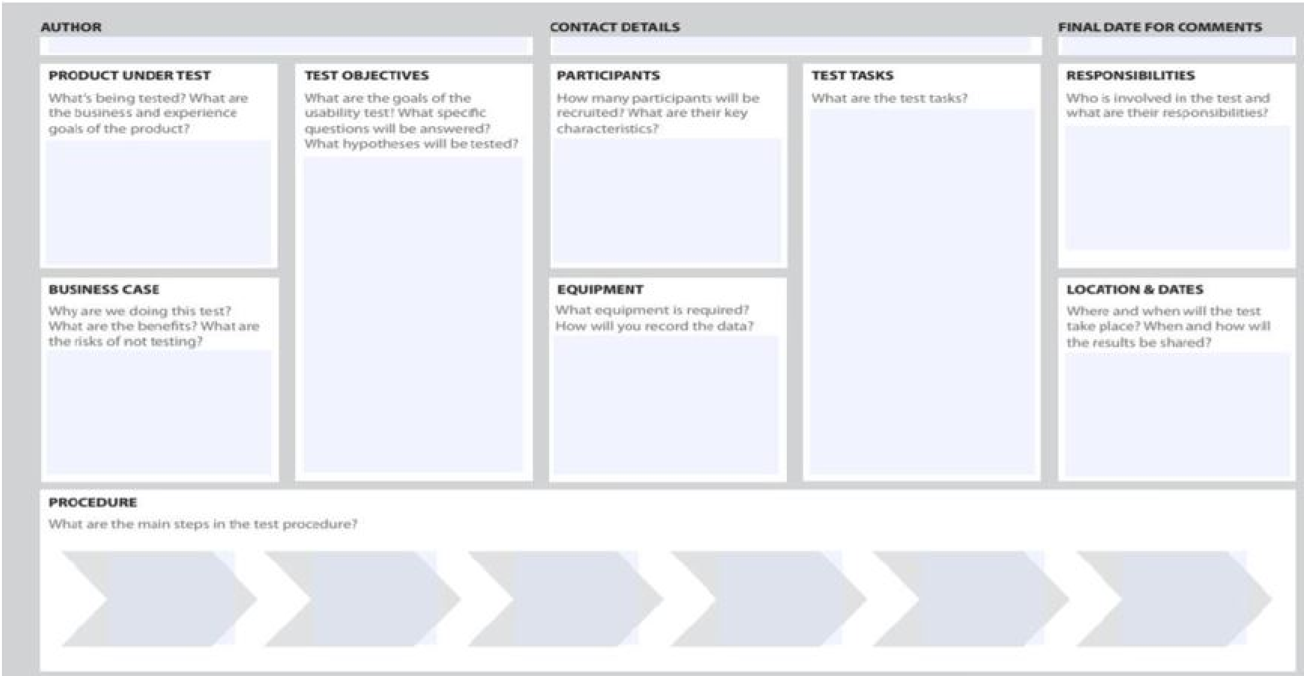


Figure 2: Usability evaluation plan

5.1 Heuristic evaluation

Heuristic evaluation is one of the Subjective measurement methods has been used in this research. Since, it makes it easier to uncover usability issues in the user interface.

5.1.1 Participants

Three experts have participated in this evaluation method. One of them was the Human-computer interaction professor at Northeastern University. The other two were the Northeastern university's students who have successfully passed the Mixed Reality and Human-computer interaction course and had some experience in these areas. None of the evaluators had previous experience with the HoloX application, which was considered a novice.

5.1.2 Procedure

To create the usability measures, we reinterpreted the ten heuristics set proposed by Nielsen [8] considering the context of AR applications(heuristics 1-10) and extended with four new heuristics, according to the literature and the application goals described by the ISO 9241-11 [11](heuristics 11-14). This allowed us to overcome a drawback of this heuristic method, which initial focus was on desktop applications.

Then, we asked three experts to freely work with the applications (No pre-defined task) to evaluate them based on the checklist and offer possible solutions. Lastly, the calculation of the heuristic evaluation results is performed. The following are our interpreted and extended Nielsen Heuristics.:

1. *Visibility of system status:* Does the application keep the user informed during the interaction? (Information such as progress stages and system state should be provided continuously) For example, when creating

Hologram, the application should notify the user what the next step is. When scanning surfaces, the user should be informed when the physical surfaces are difficult to detect, or the object can be placed. If it takes time to load an object, the system should inform the user.

2. *Match between system and the real world* Does the application design follow real-world conventions? For example, Hologram should appear as real as possible and must be coherent with the scene. They should be fixed to surfaces, as it is in the real world.

3. *User control and freedom* Does the application provide freedom for users to perform actions and undo incorrect actions? For example, when creating Hologram, if the user decides to retake or delete the Hologram and when scanning surfaces if the user places the Hologram in an unwanted location or deletes the object by mistake, the system should support undo and redo, and it should confirm with the user when the deletion is selected.

4. *Consistency and standards* Does the application have a consistent interface layout and user interaction? For example, Gestures used to resize, move the Hologram, capture or record video should be consistent and similar to other applications to avoid mistakes. User-centred languages should be used.

5. *Error prevention* Does the application avoid mistakes and prevent undesired action? For example, when a user submits a video, the system should present the user with a confirmation option before they submit. If the user does not confirm, it will return the previous step.

6. *Recognition rather than recall* Is it easy to memorize how to work with the application (e.g. create and edit Hologram)?

   For example, useful gestures for resizing the Hologram, the marker functionalities and positioning, all possible actions should be easy to memorize (e.g. undo/redo, delete, change colour).

7. *Flexibility and efficiency of use* Is it easy to interact with the application for novice users? For example, the minimum of action should be required to create a profile, record, and view a Hologram. There should not be any inconvenience in the operation while the user is holding the device in one hand. The system should allow the user to scan the surface, change the location or other properties of the Hologram easily without watching the tutorial. There should be a menu for speeding up the interactions.

8. *Aesthetic and minimalist design* Does the application show irrelevant or rarely used information to the user? For example, the process of creating, editing, publish and viewing Hologram should be easy to understand for all users. The system should have an easy-to-understand visual design for an interactive AR environment. *As the question in this Heuristic is the only negative question, the scale should be converted when analyzing.

9. *Help users recognize* Does the application diagnose and recover from errors? For example, the moment the system has trouble detecting surfaces, the user should be informed of the reason and help the user correct the error.

10. *Help and documentation* Does the application have a brief and easy to find tutorial for first-time users? Such a system usually should not have explicit documentation because it should be intuitive to use. However, the

application should have a brief tutorial for first-time users. For example, recording Hologram and surface detection is an action that most users are not familiar with. So, there should be a brief tutorial and explicit indication telling the user how to create a Hologram and how to detect surfaces. In addition, the possibility of translating and rotating the objects should also be clearly communicated to the user.

.

Below Is the added Heuristic according to the ISO 9241-11:

11. *Accessibility of off-screen objects* Is it easy for novice users to find items when those items are outside the field of view? In addition to visual elements, multimedia elements like audio should be provided.

12. *Environment setup* Does the application require a simple environment setup? For example, recording video, setting up the environment and background, scanning the surface should be as simple as possible.

13. *Accuracy* Does the application achieve an accurate outcome? For example, the created Hologram or the detected surface should be right and accurate.

14. *Satisfaction* Is the navigation, interaction and screen design enjoyable and pleasant? Interaction is an important aspect of AR applications, and the user must have positive attitudes toward the system.

5.1.3 Result Analysis

Reviewing literature [?]the data analysis in this method is performed as below:
The number of participants in the heuristic evaluation method amounted to 3 people. To calculate the results of each scale selected by participants, the formula T x Pn is used:
. T = Likert scale for each column (Strongly Disagree = 1, Disagree = 2, Agree = 3, Strongly Agree= 4).
. Pn = The number of participants who select each scale
Then the total is calculated for each Heuristic. To measure the percentage of each Heuristic, the following calculation has been used:
. Y: Number of the highest Likert Value multiplied by the number of respondents
. Y: T x Pn = 4 x 3= 12
. Percentage of each Heuristic= Total / y * 100
After getting the percentage then the interval is found for the interpretation criteria as follows:
. Interval = 100 / The number of point scales
. Interval = 100/4 (Four-Point scales), Interval = 25

No	Value Interval	Information
1	0% - 24,99%	Bad
2	25% - 49,99%	Not Good
3	50% - 74,99%	Good
4	75% - 100%	Very Good

Figure 3: Interval table

The below table is a percentage of each Heuristic from the result of the questionnaire(checklist):

No	Heuristic Principle	Total users	Strongly agree 4	Agree 3	Disagree 2	Strongly disagree 1	Each Heuristic Total score	Each Heuristic Percentage
1-	**Visibility of system status**	3	0	2	0	1	7	60 %
2-	**User control and freedom**	3	0	2	1	0	8	66 %
3-	**Match between system and the real world**	3	0	2	1	0	8	66 %
4-	**Consistency and standards**	3	0	2	1	0	8	66 %
5-	**Error prevention**	3	0	1	2	0	7	60 %
6-	**Recognition rather than recall**	3	0	2	1	0	8	66 %
7-	**Flexibility and efficiency of use**	3	2	0	1	0	10	83 %
8-	**Aesthetic and minimalist design**	3	2	1	0	0	11	91%
9-	**Help users recognize**	3	1	1	0	1	8	66 %
10-	**Help users recognize**	3	0	2	0	1	7	60 %
11-	**Accessibility of off-screen objects.**	3	0	0	2	1	5	42 %
12-	**Environment setup**	3	0	2	1	0	8	66 %
13-	**Accuracy**	3	2	0	1	0	10	83 %
14-	**Satisfaction**	3	0	3	0	0	9	75 %
								Avg= 68 %

Figure 4: Heuristic result table

5.1.4 Finding

From the above table, three notable results have been obtained:

1. First, the average value of the Heuristic percentages is 68
2. Heuristic 11 (Accessibility of off-screen objects) has the lowest score, considered in the "Not good” category.
3. “Aesthetic and minimalist design” with the highest score has been considered the best Heuristic applied in the evaluated application.

5.2 Cognitive walk-through

The details of conducting this evaluation method for the "Create Hologram" task are explained in this section. The other two tasks (View and Share) have not been analyzed in the walk-through evaluation due to the limitations of the available time and experts.

5.2.1 Participants

Users who have participated in this evaluation are the same as Heuristic evaluation. As mentioned before, the users in this method should be experts and novices to the system. The users in this evaluation are named as below:

- P1, P2: The two experts from Northeastern University
- P3: Professor in Northeastern University

5.2.2 Procedure

In the first step, the researcher has defined “Create Hologram” to examine the interaction between user and interface. Then users have been asked to perform the "Create Hologram" Task, which contains some pre-defined subtask.

TASK NAME: Create Hologram
Sub-Task:
ST 1: Press the button to start creating a Hologram
ST 2: Take your background capture
ST 3: Record your video
ST 4: Stop the video
ST 5: Submit your Hologram
ST 6: Publish the generated Hologram

Figure 5: Predefined task/sub-task

Finally, users were asked to answer the four (Yea/No) questions for each subtask for the last step. Following is the pre-defined task and subtasks provided by the researcher :

. Q1. Will the user try to achieve the right effect? (Conceptual model)

. Q2. Will the user notice that the correct action is available? (Visibility of actions)

. Q3. Will the user associate the correct action with the desired effect? (Labelling)

. Q4. If the correct action is performed, will the user see that progress is being made towards the solution of the task?

At the end of each asked users were asked to provide their feedback.

5.2.3 Result analysis

The data were collected in this methodology is as below:

		Q1	Q2	Q3	Q4	Feedback
ST 1	P1	✓	✓	✓	✓	It is intuitive
	P2	✓	✓	✓	✓	Good enough
	P3	✓	✓	✓	✓	OK
ST 2	P1	✓	✓	✓	✓	It was ok for me.
	P2	✓	✓	✗	✓	Before taking the capture, I did not know that the background pic was needed. Therefore, the picture I took was not correct.
	P3	✓	✓	✗	✗	Instruction was not offered before starting the task.
ST 3	P1	✓	✓	✓	✗	it is not clear that it is recording video; The light was not blinking
	P2	✓	✓	✓	✗	The transition to get out of frame with only audio feedback was quick and rushed. It could be good to have a visual indicator that the count down begins and is controlled by the user to begin the three-second count down.
	P3	✓	✓	✓	✗	Across the board - visual feedback for completing or making progress on a task is the most immediate required actionable.
ST 4	P1	✓	✓	✓	✓	It was intuitive and simple.
	P2	✓	✓	✓	✗	Stopping, the video went smoothly. I was not sure how long the video could be or if I would be automatically stopped. From far away, I assumed the stop button indicated that the video was being recorded, not that it was a stop button.
	P3	✓	✗	✗	✗	Across the board - visual feedback for completing or making progress on a task is the most immediate required actionable.
ST 5	P1	✓	✓	✓	✓	It was a little unclear what submitting a hologram implies at first.
	P2	✓	✓	✓	✗	The circle that shows the Hologram being uploaded does not specify how the process is going and how long we have to wait.
	P3	✓	✗	✗	✗	Across the board - visual feedback for completing or making progress on a task is the most immediate required actionable.
ST 6	P1	✓	✓	✓	✓	Similarly, it was a little unclear what publishing a hologram implies.
	P2	✓	✗	✗	✓	At first, I did not know how to publish the Hologram; however, I soon found the button.
	P3	✗	✗	✗	✗	Across the board - visual feedback for completing or making progress on a task is the most immediate required actionable.

Figure 6: Cognitive-walk-through result table

5.2.4 Finding

The result shows that the easiest and understandable subtask with the more positive response is sub-task 1, in contrast with subtask six, which has the most negative response.

In general, based on the obtained feedback, the most special issue in the view of users was the lack of proper instruction and feedback. Moreover, two software bugs have identified during the evaluation phase:

- The application crashed Two times during the evaluation phase.
- The generated Holograms sometimes disappear and appear again after creating another hologram.

5.3 Laboratory observation

In order to evaluate the AR application from the perspective of the target users, a laboratory evaluation has been conducted. In this study, the laboratory evaluation method aims to measure the Usability factors based on effectiveness, efficiency and satisfaction and explore that to what extent the application is usable.

5.3.1 Participants

The users who participated in this usability method were master's degree students of Computer Science in Northeastern University. All the participated users were familiar with the concept of AR but novice to the system. Users have not been instructed before using it.

5.3.2 Procedure

Following are the details of the laboratory evaluation method :

The test condition was the same for all users, including the same location, same device (an iPhone 12) and same network. First, the unstructured tasks have been given to the users. During a usability test, the observer has done screen recording and note-taking to assess the extent of system effectiveness, efficiency, and satisfaction. After performing each task, users answered the Single Ease Question (SEQ) with the one pre-defined question. After performing all tasks, they filled out the System Usability Scale (SUS) with the ten pre-defined questions by (Brooke and Weerdmeester 2011). Finally, the researcher examined the overall usability of the application using the Single Usability Metrics (SUM), which is defined by efficiency, effectiveness, and satisfaction.

5.3.3 Result analysis

Following formulas have been used for the calculation of each usability factor and SUM: [12]

1. *Effectiveness*

   - Task completion rate: The number of complete tasks successfully by each user

$$\textit{Task completion rate} = \frac{\textit{Number of completed tasks}}{\textit{Total number of tasks}}$$

   - Number of errors: The number of errors made by the user in each task

$$\textit{Error Rate} = \frac{\textit{Total number of errors}}{\textit{Total Opportunities}}$$

2. *Efficiency*

- Time on task: How long it takes to complete each task

.

$$\textit{Task time} = \textit{End time} - \textit{Start time}$$

Efficiency is measured in terms of task time which is calculated in this way:

$$Overall\ Relative\ Efficiency = \frac{\sum_{j=1}^{R}\sum_{i=1}^{N} n_{ij}t_{ij}}{\sum_{j=1}^{R}\sum_{i=1}^{N} t_{ij}} \times 100\%$$

Where:

N = The total number of tasks (goals) , R = The number of users

nij = The result of task i by user j; if the user successfully completes the task, then Nij = 1; if not, then Nij = 0

tij = The time spent by user j to complete task i.

3. *Satisfaction*

- Task Level Satisfaction(SEQ): How satisfied participants are with each

SEQ Avg= Sum of the scores/Number of users

- Test Level Satisfaction(SUS): How satisfied participants are with the product.

((q1-1)+(5-q2)+ (q3-1)+ (5-q4)=. . . + (q9-1)+(5-q10))*2.5

SUS Calculation [13]: users have ranked each of the ten templates questions from 1 to 5, based on their level of agreement. For each of the odd-numbered questions, 1 is subtracted from the score and for each of the even-numbered questions, their value is subtracted from 5. These new values have been taken and added up to the total score, then multiplied by 2.5

4. *Single Usability Metrics (SUM)*

SUM allows a researcher to reliably sum up metrics from every sub-metric of every task to a single overall value for a product or experience. The three -tasks in this study (create, view, share) have nine sub-metrics(effectiveness, efficiency, satisfaction for each task) which can be shown in a single number using SUM score. SUM represents the application's usability at a glance. Any SUM score below 50 percentage is relatively poor, and anything above 50 percentage is relatively good [13].

Single usability score is obtained from the below calculation:

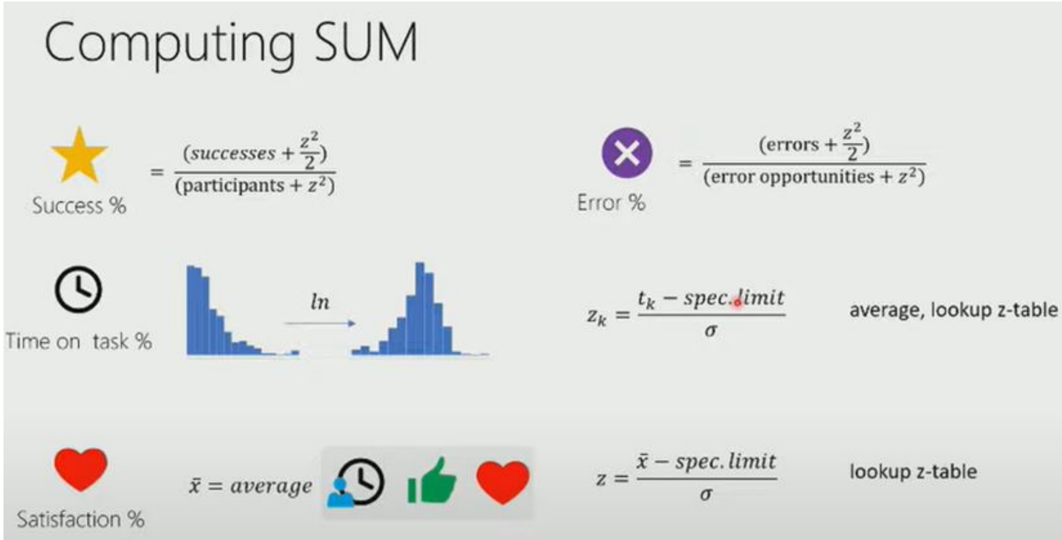


Figure 7: Single Usability measurement formulas

5.3.4 Finding

The results from the all the calculation for each measurement are summarised in table below:

	Time On Task			Satisfaction (Seq)			Completion			Error		
Task names	Create	View	Share	Create	View	Share	Create	View	Share	Create	View	Share
Mean	35.8	23.4	42.1	5.6	5.1	4	0.8	0.8	0.7	1.1	3.6	3.5
SD	8.77	14.35	12.98	1.06	1.58	1.53						
Spec (specification limit)	37.5	21	31.5	5.5	5.5	5.5	0.78	0.78	0.78	< 0.7	< 0.7	< 0.7
Result from calculations	***75.8***	***76.3***	***64.4***	***5.6***	***5.2***	***4.0***	***82%***	***82%***	***73%***	***0.4***	***0.7***	***0.9***

Figure 8: Usability measurement for each task

The specification limit for satisfaction, Completion and error, which determines the good score, is defined in the literature. The specification limit for time on task is calculated by multiplying 2.5 to the required time for performing the task by expert [12].

The laboratory tests and the methodology used for each usability factor calculation helped in representing efficiency (figure 9),effectiveness (figure 10), and satisfaction (figure 11) for each task.

1. *Efficiency*

Figure 9: Based on time on task

2. *Effectiveness*

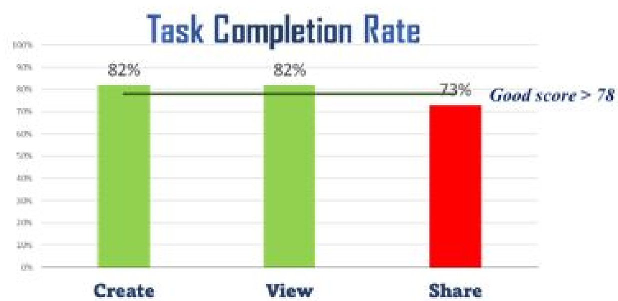


Figure 10: based on the number of completed tasks

Figure 11: based on the number of error

3. *Satisfaction*

Figure 12: level of user satisfaction for each task based on SEQ questionnaire

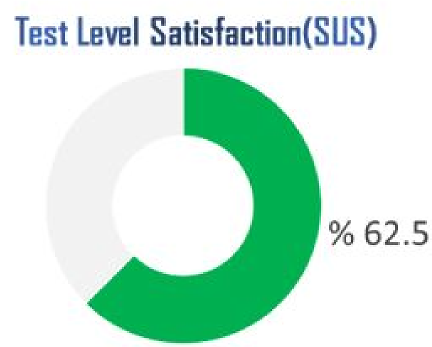


Figure 13: level of user satisfaction for the application based on SUS questionnaire

Here's an overview of how the app scores should measure: • 80.3 or higher is an A. People love the application and will recommend it to their friends • 68 or thereabouts gets you a C. Y It is doing ok but could improve • 51 or under gets you a big fat F. it needs to be fixed.

5.4 Questionnaire

Using this method, we collected users' opinions about some important aspects of the user interface in the context of the AR applications with some pre-defined questions.

5.4.1 Participants

The users who participated in this usability method were the same users who participated in laboratory observation. (11 northeastern university students)

5.4.2 Procedure

After the laboratory evaluation, participants were asked to rate the pre-defined questions on a scale from 1 (strongly disagree) to 5 (strongly agree) and give their overall feedback about the application. Reviewing literature and considering some factors in the heuristic evaluation, 30 questions have been designed in this method [4, 14, 15, 16, 17?].

Then, we analyzed the subjective data collected from the questionnaires. Based on the result, we documented the explored usability issue by the users and offered some design guidelines for improving the design interface.

5.4.3 Finding

In the table below all 30 pre-defined question provided in the questionnaire and the average score of each question have been provided:

No	Questions	Avg
1	When beginning to use the device and application, it was straightforward.	3.1
2	The system provided me with feedback on what I was working on.	2.9
3	The messages that appeared in the application were self-explanatory.	3.5
4	I was in control of the application all the time	3.3
5	I could easily undo/redo any action if I felt to do it.	3.4
6	If I mistakenly perform an action, I could easily stop or remove it.	3.4
7	The navigation through the application was easy.	3.2
8	The application had many errors and/or crashed.	2.5
9	The option to select the desired function was always clear and available all the time, and the icon images helped	2.9
10	It was easy to find the desired options at any time.	2.9
11	The application was visually well-designed.	3.4
12	If error messages appeared, they were clear in their description, and probable steps to recover from them were provided.	3.1
13	When I needed help, there was helping document to complete my task successfully.	3.5
14	The application provides hints throughout to guide you with the next steps.	3.6
15	The system did not behave as I expected.	2.7
16	I could not always achieve what I wanted the system to do.	2.1
17	I kept making mistakes interacting with the system.	2.9
18	Accurate pointing to detect surfaces was easy to achieve.	4.5
19	Hand interactions were difficult to perform.	2.2
20	When creating a hologram, voice interactions were properly recognized by the system.	3.6
21	The audio instructions provided were easy to understand.	3.9
22	The quality of generated Hologram was very realistic	4.6
23	Interaction with the system was fast enough.	3.4
24	If moving while using the application, the augmentations stay still regarding the place they should appear in relation to my movement.	4.2
25	The Hologram appears inappropriate size.	4.5
26	Generated Hologram and the background are easily differentiated from each other (i.e. is the brightness and contrasts appropriate)	4.4
27	The information offered by the application organized and grouped clearly?	2.8
28	The information which requires action is identified easily. (It is highlighted or differentiated in any other ways).	3.3
29	It is easy to create a Hologram.	4.0
30	The time of the usage of the application was appropriate	4.3

The following figure illustrates that the avg score of questions 18, 21,22,24,25,26,29 and 30 has obtained a good score. Reviewing the mentioned questions, It is easy to conclude that the application's main function, such as creating a Hologram, detecting surface, the size, voice, and quality of the Hologram, is highly qualified.

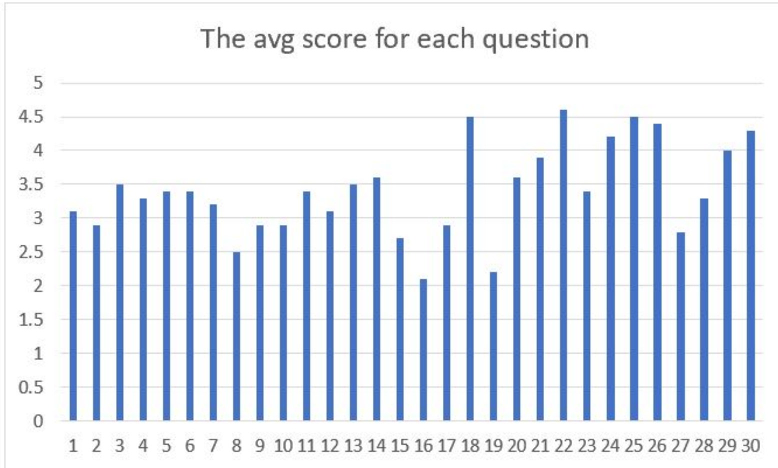


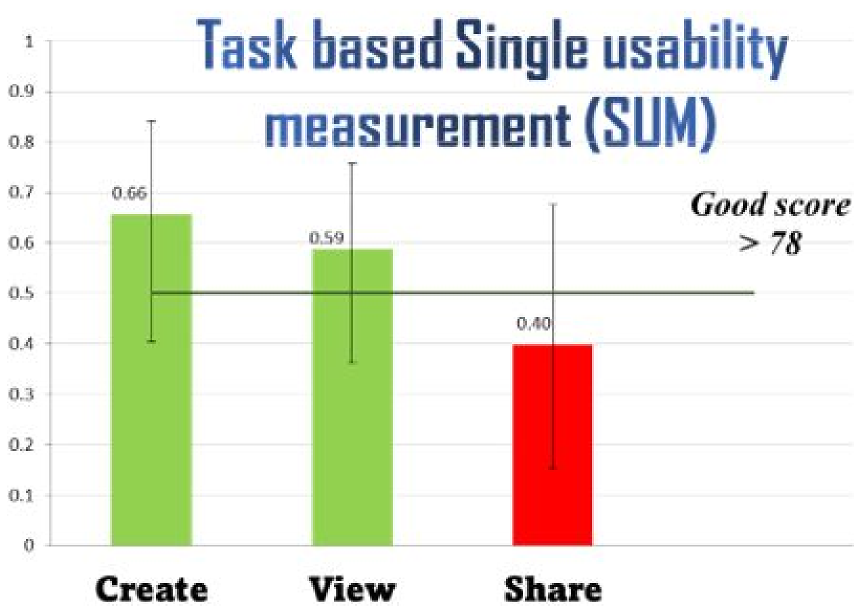
Figure 14: Result obtained from questionnaire

6 SUMMARY

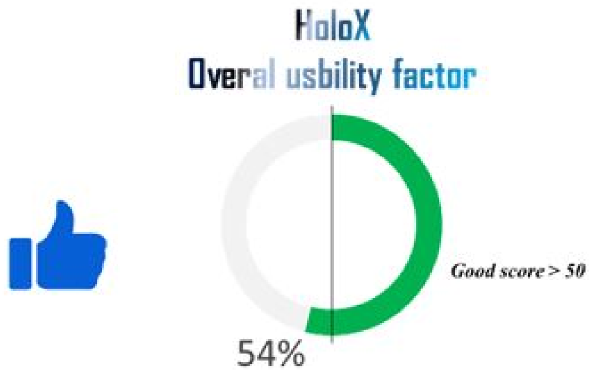
The summary of our research have been provided in this section based on our Research Questions: *Research questions:*

1- To measure Usability factors based on effectiveness, efficiency and satisfaction?

The below bar chart presents the usability factor of HoloX with a single number for each task. It can be seen Create and View task in this application is overally usable, and share task needs design interface enhancement.



The pie chart below is representing the overal HoloX usability factor with the single number which is defined as a usable application based on the ISO definition(effectiveness, efficiency and satisfaction) .



2- What are the usability Problems experienced by users while using holographic AR applications (HoloX)?

In the table below the most common usability issue that have been declared by the users can be found:

Navigation between the pages is a little confusing.
Users are not aware of the available option on the main page.
The UI alone is not showing the user how to use the app.
Users are not aware the background picture is needed before starting to create a Hologram.
Users make a mistake in taking background pictures in a way that they should go out of the frame.
It is not clear how long the video could be or if it will be automatically stopped.
Recording the video does not indicate it is recording.
The transition to get out of frame with only audio feedback is quick and rushed.
The circle that shows the Hologram being uploaded does not specify how the process is going and how long we have to wait.
It is unclear what publishing a hologram implies.
When publishing the Hologram, It does not show that the hologram title input is compulsory.
There is no option to delete the failed hologram.
Share profile picture's icon is confusing with scanning the QR's icon.
Share a hologram is out of the system visibility.
There is not any feedback on when and where the barcode is downloaded.
It is not indicated to the user how the object can be manipulated.
Inappropriate labels and icons, for example, Profile Icon led users to scan the QR code, share Hologram leads users to download Hologram, QR code button leads the user to share their profile, help option is located in the setting menu.
When detecting the surface, the label above the scene is confusing with the button. Users try several times to locate the Hologram by pressing that label.
None of the users utilized the Help menu option provided in the user interface. Instead, all of them asked for manual help from the evaluator.
When scanning the surface, there are some buttons that are not working.
The application crashed five times during the evaluation phase.
The generated Holograms sometimes disappear and appear again after creating another hologram.
After setting the profile picture, the app does not show it in the profile.
The combination of the button and the label associated with changing the profile picture is confusing for the users.
When a user logout and the other user login into the system, still the profile picture of the previous person is shown.

3- What are the design guidelines to improve the holographic AR applications (HoloX)?

The design guidelines based on the collected feedback in all of the four usablity method has been provided in below table:

Improving navigational between the pages
Providing feedback would be the biggest solution
Providing visual prompts like meta-UI elements for preventing error
Providing a visual aid before starting the task
Showing animation gestures in the first time of use
Providing proper labels and icons for sharing profile, sharing Hologram and View Hologram
Process the Hologram from the video locally on the phone without the need for the internet
With an extra UI component, the user can be presented that the Hologram manipulation option is available.
Designing the user interface in such a manner that it will have the least learnability curve.
It could be good if it will be controlled by the user to begin the three-second count down.
When recording the video, a timer or bilking light could help to indicate the video is recording.
Providing a hint "the image will appear in the gallery" after downloading the Hologram
Providing an option to play the Hologram again without having to exit and scan the QR code again

7 CONCLUSION

In this research, we presented the analysis of four usability evaluation methods, including heuristic evaluation, cognitive walk-through laboratory observation and questionnaire. A developed application called HoloX has been chosen as the pilot for the Holographic AR application. The goal of this study measures the usability factor and explores the design problem in this application. The usability factor calculated by SUM(single usability factor) is equal to 54 percent, which indicates that the system is usable. During the evaluation carried out in this study, several design problems have been explored in each methodology. The results obtained show that there is a need for the developers to revisit the application to enhance the interface's design. Then Some design guidelines have been provided based on the users' feedback.

8 FUTURE WORK

Due to the limitation of time, we could not wait to evaluate the new version of HoloX. Researchers can use the applied methodology to evaluate the newly released version of the application. One of the other ideas of future work for the proposed study is to create a larger laboratory experiment including more users to obtain more statistically consistent results in laboratory observations and questionnaires. Including more experts for the heuristic evaluation and cognitive walk trough would provide a large advantage. Apart from more users, the cognitive walk-through was performed just to create a Hologram because of a lack of time and experts. Conducting another cognitive walk-through evaluation for view and sharing the Hologram will help to explore more details about the usability issue in those parts. Finally, conducting further methods apart from the four studied methods would provide a better overview of which usability evaluation method can be more suitable for the evaluation of AR interfaces.

References

[1] Thomas M Alisi, Alberto Del Bimbo, and Alessandro Valli. Natural interfaces to enhance visitors' experiences. *IEEE MultiMedia*, 12(3):80–85, 2005.

[2] Yee-Yin Choong, Kevin Mangold, Kurtis Goad, and Samia Benjida. Pscr 2021: Augmented-reality (ar) usability evaluation framework for pscr. PSCR 2021: The Digital Experience, Boulder, CO, US, 2021-09-28 04:09:00 2021.

[3] Hsinfu Huang and Chin-wei Chen. Creating different learning experiences: assessment of usability factors in an interactive three-dimensional holographic projection system for experiential learning. *Universal Access in the Information Society*, 18(3):443–453, 2019.

[4] Satu Elisa Schaeffer. Usability evaluation for augmented reality. 2014.

[5] Valéria Farinazzo Martins, Leandro Valdeviesso Soares, Vitor Barbosa Mattos, Eduardo Luiz Makihara, Renato Nakahishi Pinto, and Marcelo de Paiva Guimarães. Usability metrics for augmented reality applications. In *2012 XXXVIII Conferencia Latinoamericana En Informatica (CLEI)*, pages 1–10. IEEE, 2012.

[6] Nigel Bevan. Iso 9241: Ergonomic requirements for office work with visual display terminals (vdts)-part 11: Guidance on usability. *Tc*, 159:61, 1998.

[7] Sugandha Gupta. A comparative study of usability evaluation methods. *International Journal of Computer Trends and Technology*, 22(3):103–106, 2015.

[8] Jakob Nielsen and Rolf Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 249–256, 1990.

[9] Cathleen Wharton, John Rieman, Clayton Lewis, and Peter Polson. The cognitive walkthrough method: A practitioner's guide. In *Usability inspection methods*, pages 105–140. 1994.

[10] Andreas Holzinger. Usability engineering methods for software developers. *Communications of the ACM*, 48(1):71–74, 2005.

[11] Tristan C Endsley, Kelly A Sprehn, Ryan M Brill, Kimberly J Ryan, Emily C Vincent, and James M Martin. Augmented reality design heuristics: Designing for dynamic interactions. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 61, pages 2100–2104. SAGE Publications Sage CA: Los Angeles, CA, 2017.

[12] Jeff Sauro and Erika Kindlund. A method to standardize usability metrics into a single score. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 401–409, 2005.

[13] Michael Van Waardhuizen, Jennifer McLean-Oliver, Nancy Perry, and Joe Munko. Explorations on single usability metrics. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–8, 2019.

[14] Woon-Hyung Lee and Hyun-Kyung Lee. The usability attributes and evaluation measurements of mobile media ar (augmented reality). *Cogent Arts & Humanities*, 3(1):1241171, 2016.

[15] Ramy Hammady, Minhua Ma, and Anna Powell. User experience of markerless augmented reality applications in cultural heritage museums:'museumeye'as a case study. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pages 349–369. Springer, 2018.

[16] Pier Luigi Ingrassia, Giulia Mormando, Eleonora Giudici, Francesco Strada, Fabio Carfagna, Fabrizio Lamberti, and Andrea Bottino. Augmented reality learning environment for basic life support and defibrillation training: usability study. *Journal of medical Internet research*, 22(5):e14910, 2020.

[17] Renan Luigi Martins Guarese and Anderson Maciel. Development and usability analysis of a mixed reality gps navigation application for the microsoft hololens. In *Computer Graphics International Conference*, pages 431–437. Springer, 2019.

[18] J Nielsen. How to conduct a heuristic evaluation. nielsen norman group. *Available online (last accessed November 2016) at: http://www. nngroup. com/articles/how-to-conduct-a-heuristic-evaluation*, 1995.