2021 European Conference on Computing in Construction
Online eConference
July 26-28, 2021

A USABILITY TEST PROTOCOL FOR EVALUATING MIXED REALITY ENVIRONMENTS

Neziha Yılmaz, Esin Ergen, and Deniz Artan
İstanbul Technical University, İstanbul, Turkey

ABSTRACT

Immersive technologies are being used to enhance the effectiveness and usability of BIM in the construction industry and multiple studies that test the usability of the developed prototypes exist in this domain. However, a structured usability test protocol for the use of immersive technologies is lacking. This paper aims to present a usability test protocol that is developed to measure the performance of a prototype, which is developed for a mixed reality environment in a building. The main steps of this test protocol are 1) cognitive walkthrough, 2) pre-test surveys, 3) usability test video recording 4) post-test surveys.

INTRODUCTION

Immersive technologies such as virtual reality (VR), augmented reality (AR), and mixed reality (MR) are recommended for more effective use of Building Information Modelling (BIM) in the construction industry (Rankohi and Waugh, 2013; Chu, Matthews and Love, 2018).

This area is related to human-computer interaction, which focuses on interfaces between users and computers and explores the design and use of computer technology (Shen et al., 2016). Usability tests are carried out to evaluate the effectiveness of human-computer interaction and to determine how much the developed system meets the user's needs. In these tests, users evaluate the developed system in terms of multiple factors, such as ease of use, learning time, and error rate during task execution. The results indicate the effectiveness of the developed system and also the improvement requirements are identified for future works. In the context of human-computer interaction and product development, standardized usability questionnaires include Usefulness, Satisfaction, and Ease of Use (USE), System Usability Scale (SUS), After Scenario Questionnaire (ASQ), Poststudy System Usability Questionnaire (PSSUQ), and etc.

The use of a mixed reality environment where the user interacts with the virtual model has a great potential in the facility management phase, both for facility management personnel and for the occupants of a building (Irizarry et al., 2013; Chalhoub and Ayer, 2018). It is envisioned that the facility management personnel can access the technical and operational details of the building components at site and enter the related operation and maintenance information to the virtual model in the mixed reality environment (Yılmaz, 2020). On the other hand, the occupants can access some information related to the building components and enter some user feedback in the model, for example, feedback related to the performance of the building and building components (Ergen et al., 2021). Various studies were performed to realize this vision and developed prototypes that were tested for usability by the users (Irizarry et al., 2013; Chalhoub and Ayer, 2018; Chu, Matthew and Love, 2018).

In a recent study, (Chalhoub and Ayer, 2018) examined the use of mixed reality technology for the assembly of electrical elements in the construction phase During the test, video recordings of the users were taken and a questionnaire study was conducted afterward. In another study, the users were asked to find the objects in the room in a virtual/augmented reality-based test environment and answer the questions about the relevant elements (e.g., "What is the last inspection date of the printer?") (Irizarry et al., 2013). During the tests, the duration of the participants to fulfill the tasks was measured, and a questionnaire was applied to the participants after the scenario. (Chu, Matthews, and Love, 2018) questioned the accessibility of the information about door and window elements in an AR environment by using markers on two-dimensional layouts in a test. The duration to complete each defined task was measured and questions were asked about the quality of the interaction.

The review of the existing studies demonstrated that the usability tests utilized for testing the immersive technologies do not include a systematic approach in questionnaire design. Besides, they are focused on the specific tasks of the technical personnel and thus they are not fully applicable to the occupants of the buildings. Moreover, a usability test protocol that is specifically developed for the users of the facility management phase in a mixed reality environment is lacking in the literature (Irizarry et al., 2013; Chalhoub and Ayer, 2018; Chu, Matthew, and Love, 2018).

This paper aims to propose a usability test protocol that is developed to measure the performance of a prototype, which is developed for the users of a mixed reality environment in the facility management phase of a building. To achieve this aim, a literature review was performed to investigate the existing usability test questionnaires used for product evaluation and human-computer interaction and to identify the usability test protocols that are used in the domain. Based on the results of the literature review, a usability test protocol was proposed for the evaluation of mixed reality environments in the facility management phase.

MIXED REALITY ENVIRONMENT

In a mixed reality environment, virtual objects and/or data are attached to the physical environment and the users can dynamically interact with these virtual objects. In this environment, mixed reality technologies that are integrated with depth-sensing cameras, position tracking sensors, and indoor positioning techniques are used.

The use of mixed reality technology has a great potential for facility management. With the help of mixed reality, users will be able to see the BIM model overlaid to the physical building components and interact with the BIM model effectively to access the information related to the building components and to enter information related to those components. The users in the facility management phase can be the technical personnel or the occupants of the building. The technical personnel needs to access the technical and operational data, such as problems observed, the repair dates, repair actions related to a building component (Artan, Yılmaz and Ergen, 2018). Moreover, the expert facility managers in the head office can connect to the mixed reality environment seen by the technical staff in the field and will be able to support him remotely and point out important aspects. On the other hand, the occupants of the building can interact with the BIM model to enter their feedback related to the performance of the building and occupant satisfaction (Yılmaz, 2020).

The studies associated with the use of immersive technologies in facility management and the use of human-computer interaction are limited (Shi et al., 2016). Few studies examined the features that BIM models should have in virtual environments for facility management purposes (Iorio et al., 2011; Zou, Arruda, and Ergan, 2018). Other studies focused on the use of augmented reality (AR) in facility management and periodic maintenance processes (Irizarry et al., 2013; Olbrich et al., 2013; Ammari and Hammad, 2014). However, in these studies, the information and features that models should contain in terms of user interaction were not examined, mixed reality technology was not used and the effectiveness of user interaction was not evaluated.

Since user interaction is prominent in studies using mixed reality, it is important to carry out usability tests evaluating the human-computer interaction of the developed systems and to make evaluations for the end-user. However, there is limited work in this area both in facility management and other areas (Irizarry et al., 2013; Chalhoub and Ayer, 2018; Chu, Matthews and Love, 2018). These studies do not include a systematic approach in the preparation of survey questions and only report general evaluations. Besides, the evaluations are focused on some of the steps performed by the technical staff and a comprehensive study of usability involving various tasks is missing.

USABILITY TEST PROTOCOLS

The usability test protocols that were used in the literature are investigated in this section under two categories.

Cognitive Walkthrough

The cognitive walkthrough is an assessment method that is performed by the researchers to investigate the interaction of the users with an interface through predefined tasks (Rieman, Franzke and Redmiles, 1995; Ball and Bothma, 2017). In this method, the main focus is on exploratory learning, which is defined as how well novice users can and cannot use the interface without any preliminary practice (Desurvire, Kondziela and Atwood, 1992). This evaluation method consists of two stages: preparation and evaluation stage and are carried out by at least two researchers (Martínez and Bandyopadhyay, 2014). These researchers are involved in the prototype design and are familiar with the user interface. During the preparation phase, one of the researchers defines the user profile and determine the tasks, sequence of task selection, and task-defining actions to be integrated into system design. This researcher is also familiar with the workflow and thus he/she can create tasks to test whether the workflow is progressing correctly. In the evaluation phase, the other researcher, who did not participate in the preparation stage, performs the tasks and the related sequential actions which are predetermined by the first researcher. This participant is assumed to be a novice because s/he does not know the purpose determined in the preparation phase and the task steps to realize the purpose. S/he will be expected to catch the errors and bugs in the application/prototype while performing the tasks. The main focus areas are 1) to verify that the interface menus are in the correct, complete and desired order, 2) to test whether there are any system errors, and 3) to determine if the visual design is correct and the visual elements are placed properly on the page and 4) loading time. The check list is generally dependent on the prototype, but there are some standard items to check, for example, "Is the logo clickable?", "Is the forward button active?", "Is the user name asked when logging in/opening the system?". During this process, the researcher who defined the tasks before records the second researcher's behavior (Krug, 2000; Rogers, Sharp and Preece, 2011). The participation of real users in usability tests and freely discovering the prototype takes place in the further steps of the protocol.

The data that is collected in the study is analyzed to evaluate the usability of the developed system and to detect the problems (Molich and Dumas, 2008; Rogers, Sharp, and Preece, 2011). This method can be summarized under four stages in the usability test protocol (Wang, 2008):

- Researcher 1 sets a goal to be achieved (preparation stage).
- Researcher 2 searches the interface for available actions (evaluation stage).
- The user selects an action that seems likely to progress towards the goal (evaluation stage).
- The user takes the action and checks whether the feedback indicates progress towards the goal has been made (evaluation stage).

Another method that is generally used at this stage is to think-aloud, and the researcher is asked to explain his/her thoughts during the actions he takes (Shneiderman and Plaisant, 2010). In this way, objective data collected from behaviors and thoughts can be combined when analyzing the recordings. Finally, the researcher can be questioned at the end of the study to obtain subjective data (Shneiderman and Plaisant, 2010; Fernandez, Insfran and Abrahão, 2011; Rogers, Sharp and Preece, 2011).

Subjective data from video records and objective data from questionnaires are combined to list the problems of the prototype. These problems can be grouped under certain headings such as functionality, effectiveness, understandability, and aesthetics of the system or the user interface (Kostaras and Xenos, 2009). Besides, these problems can be assigned priority levels such as 1-cosmetic), 2-minor, 3-major, and 4-catastrophic (Walden et al., 2020). At the end of the cognitive walkthrough the detected errors are fixed in the prototype.

Laboratory Evaluations

Laboratory evaluations take place in a controlled environment where subjects are assigned, variables are provided and manipulated under simulated real-life scenarios (Razak et al., 2010; Martínez and Bandyopadhyay, 2014). An alternative to the laboratory evaluation method is the field evaluation method which is performed under dynamic conditions that cannot be simulated in a laboratory environment. However, this method is found to be compelling since it is not a controlled environment. The laboratory evaluation method is advantageous over the field evaluation method when assessing user interfaces since the researchers can obtain video/audio recordings of user interactions with the user interface (Zhang et al., 2003; Kumar and Mohite, 2018). In the laboratory evaluation, the researcher has full control over the assignment of variables in the test environment (e.g., sound, light, ethernet infrastructure, human circulation, objects in the environment) and the determination of the participants depending on the target audience. Therefore, the field evaluation method is not considered in this study.

Data collection and evaluation steps are similar to the cognitive walkthrough, but the main difference is that the participants of the laboratory evaluations are the real or potential users whereas the participants of the cognitive walkthrough are the researchers. Participants can be chosen from experienced people who have used similar technologies and applications, or they can be selected from people with no previous experience. The participants are expected to perform a predetermined series of tasks in a usability laboratory environment.

Researchers can define two types of tasks for the users: 1) structured tasks and 2) unstructured tasks (Alshamari and Mayhew, 2008). The structured tasks are designed by the researchers who create a step-by-step to-do list that users will follow to complete a task. These tasks can be described in detail by presenting a realistic scenario that explains what the user should do in each step. For example: "(1) Open the application (2) Enter the user ID (3) Enter password (4) Open home page (5) Click on the wall where the window is placed (6) View the model of the mechanical equipment on the wall (7) Determine the system of the yellow pipes (e.g., wastewater, drinking water) (8) Close the application". The unstructured tasks are briefly described to allow the participants to find their way in completing the defined task. For example: "Please state the system of mechnaical pipesbehind the wall."

The usability of the system can be determined via video recordings of the user's interactions, and/or surveys (Martínez and Bandyopadhyay, 2014). The videos of the users are recorded to determine the behaviors and tendencies of the users when they interact with the user interface while performing the assigned tasks. At this stage, various factors will be taken into consideration in evaluating the usability of the system, such as the task completion duration, the number of errors, the number of correctly completed tasks, and the problems encountered in the natural flow of the tests. This stage is intended to provide quantitative data on the tests. The second method to evaluate the usability of the system is to conduct surveys to obtain user's opinions about the tested system (Razak et al., 2010). In this method, the task steps defined in the laboratory environment are followed and video/audio recordings are taken. Afterward, the participants' subjective evaluations are collected by a survey in the lab environment. This part involves questioning the user's opinions and preferences about the functionality, effectiveness, understandability, and aesthetics of the system or the user interface (Kostaras and Xenos, 2009).

There are two main types of surveys that are performed:

(1) Pre-test Surveys: It will be conducted before the user tests the system to determine the participant profiles and to collect reference data for the usability measurements of the system. For example, age, gender, profession, computer literacy.

(2) Post-test Surveys: It will be applied to obtain qualitative evaluations of the tests. Usability testing processes generally end with with the analysis of users' ideas and preferences about the developed system/interface/prototype/application. The main types of post-test surveys in the literature and sample questions are provided in the next section.

The advantage of the survey method is its ease of application and low cost-benefit ratio (Walden et al., 2020). However, the survey method alone is not sufficient for obtaining raw data and user interface analysis since it only contains the opinions of the users. Collecting raw data of user behavior from other methods such as video record (Nielsen, 1994) or recording verbalize positive and negative reactions as they worked through tasks (Russ et. al., 2018) is found to be more reliable than the survey (Razak et al., 2010).

The data obtained in the laboratory evaluations are analyzed under various headings. For example (1) effectiveness (i.e., accuracy/user ability to complete assigned tasks/level of interaction in terms of speed and errors) (Nielsen, 1994; Zhou and Chan, 2017; Constantinescu et al., 2019; Walden et al., 2020); (2) efficiency (i.e., time to complete/resources required to complete assigned tasks such as time) (Constantinescu et al., 2019; Walden et al., 2020); (3) error (i.e., number of errors/ability to recover from errors, existence of serious errors) (Nielsen, 1994; Zhou and Chan, 2017); (4) flexibility (i.e., level of adaptation to various tasks) (Zhou & Chan, 2017); (5) learnability (i.e., ease/level of learning needed to accomplish a task) (Nielsen, 1994; Zhou and Chan, 2017; Dowding et al., 2019; Walden et al., 2020); (6) assessment of how pleasurable it is to use /attitude (Shackel, 2009; Zhou and Chan, 2017; Constantinescu et al., 2019); (7) usefullnes (i.e., ease/difficulty/failed) (Dowding et al., 2019).

USABILITY QUESTIONNAIRES

In this section, major usability questionnaires that were developed to test the usability of a product, system, or websites were reviewed. One of the most widely used usability questionnaires is the USE (Usefulness, Satisfaction, and Ease of Use) Questionnaire (Lund, 2001). It is developed to test the usability of software, hardware, services, user support materials, and products (i.e., portable vacuum cleaner (Albertazzi, Okimoto and Ferreira, 2012) under four categories: Usefulness, Satisfaction, Ease of Use, and Ease of Learning. It includes generic statements that evaluate the product, such as, "it helps me be more productive, "it is easy to learn to use it", and "I would recommend it to a friend".

Another commonly used usability questionnaire is the SUS (System Usability Scale), developed by Brooke (1996), which is used for the evaluation of websites (Grier et al., 2013; Lewis, Utesch and Maher, 2013). It includes a survey of ten questions developed at Digital Equipment Corp and each question is the rating of statements on a five-point scale ranging from "Strongly Disagree" to "Strongly Agree". SUS can be used as a measure of usability in standard task-based testing or as a backward analysis in surveys (Lewis, 2018). It is similar to the USE questionnaire, however, it includes more detailed statements, such as "I think that I would need the support of a technical person to be able to use this system". Some researchers have changed the wording in SUS such as using "website" or a product name in place of "system" (Finstad, 2006; Bangor, Kortum and Miller, 2009; Lewis and Sauro, 2009).

Usability Metric for User Experience (UMUX) was developed by Finstad (2010) to test the usability of a software prototype. It is preferred when a shorter questionnaire is needed (Lewis, 2018). UMUX consists of the following four statements that are measured using a 7-point Likert Scale: (1) the system's capabilities meet my requirements, (2) using this system is a frustrating experience, (3) this system is easy to use, (4) i have to spend too much time correcting things with this system (Finstad, 2010). The used expressions are in both positive and negative tones similar to SUS (Lewis, 2018).

After Scenario Questionnaire (ASQ) (Lewis, 1991), was created to test the usability in a scenario after the scenario-based tasks were completed by the user. The test consists of three statements: (1) overall, I am satisfied with the ease of completing the tasks in this scenario, (2) overall, I am satisfied with the amount of time it took to complete the tasks in this scenario, (3) overall, I am satisfied with the support information (on-line help, messages, documentation) when completing the tasks). The user satisfaction is measured with a 7-point Likert scale (1 = strongly agree and 7 = strongly disagree).

Poststudy System Usability Questionnaire (PSSUQ) was originated from SUMS (System Usability MetricS) (Lewis, 1991, 1992), and the main purpose is to record and corroborate the functions, systems, and procedures for measuring the usability of any interface—hardware, software, mobile applications, or websites. The main categories of the questionnaire are measuring system usability, including performance, usability problems, and user satisfaction. The PSSUQ is developed to be used at the end of a standard task-based usability study, and the statements are expressed in the past tense to refer to the completed tasks. Examples of some of the PSSUQ statements are: "Overall, I am satisfied with how easy it was to use this system", "The system gave error messages that clearly told me how to fix problems", "It was easy to find the information I needed", "I liked using the interface of this system".

Computer System Usability Questionnaire (CSUQ) was developed by IBM to test the usability of a computer system (Lewis, 1995). It consists of 19 questions categorized under system usefulness, information quality, and interface quality dimensions and it has a 7-point Likert scale ranging from "Strongly Disagree" to "Strongly Agree". Some statements from the CSUQ are: "It is simple to use this system", "Whenever I make a mistake using the system, I recover easily and quickly", "This system has all the functions and capabilities I expected it to have".

Questionnaire for User Interface Satisfaction (QUIS) is used to measure the usability of a human-computer interface (Chin, Diehl and Norman, 1988). QUIS includes 27 questions that evaluate the effectiveness of the website design (Tullis and Stetson, 2004). Some of its statements are as follows: "system speed (i.e., slow to fast)", "designed for all levels users (i.e., never to always)", "sequence of screens (i.e., confusing to very clear)" and statements evaluation range is from 1 to 9.

PROPOSED USABILITY TEST PROTOCOL

Based on the literature review, a usability test-protocol was developed for testing prototypes that are developed for a mixed reality environment to be used in the facility management phase. The main steps of the protocol are 1) cognitive walkthrough, 2) pre-test surveys, 3) usability test and video recording 4) post-test surveys.

In the first step, cognitive walkthrough takes place with the participation of at least two researchers. One of the researchers prepares a set of tasks that will be performed in the laboratory environment. For example, 'click homepage', 'select every object in the environment', 'go to previous and next pages'. The other researcher who did not participate in the preparation of the tasks performs the task during the test. Meanwhile, the movements of the researcher are recorded. The obtained data are analyzed according to the determined evaluation criteria. The criteria used can be customized according to the needs of the prototype or tasks. After determining the problems related to the prototype according to the evaluation criteria, the detected errors are fixed in the prototype. Since this step is performed before testing the usability, it is not part of the developed protocol.

In the second step, the users are selected among the target group and pre-test questionnaires are performed to determine the user profiles and to collect reference data for the usability measurements of the system.

In the third step, the introductory information about the prototype is provided to each user and they are asked to perform structured tasks by using the prototype. For example; "Click on the digital label of the window in the room you are in. In the popup that appears, click on the manufacturer company information from the information listed for the window. Update the existing information and enter ABC Company. Close the pop-up window." Structured tasks in this step is used in order to prevent users from drifting away and wasting time by trying different alternatives in performing tasks since their task durations need to be collected precisely to evaluate the performance of the prototype. Structured tasks can be defined according to the prototype requirements. Users will be allowed to discover the prototype and to become familiar with the interface by allowing a certain amount of time before the tests. The user's interaction with the technology including their questions –if any- is recorded during the test. The aim is to determine the behaviors and tendencies of the users when they perform the assigned tasks and interact with the technology.

At this stage, quantitative data are collected. According to lab test notes and video records the obtained data gathered as following heading which are compiled from the literature and considering the needs of the MR environment: task completion duration, number of errors, number of correct task completion, number of clicks while performing the tasks, problems encountered in the natural flow of the tests, number of requests for assistance, number of failed attempts to establish a relationship between the MR environment and the physical environment, number of attempts that resulted in getting lost in the MR environment. Also, the questions that were asked by the users are listed.

Finally, post-test questionnaires are conducted after the test to obtain qualitative evaluations of the tests. The sample questions that are customized for a mixed reality environment are provided in Table 1. These questions were developed based on usability questionaries such as SUS, USE, ASQ, and based on the technological needs defined in Yilmaz (2020). The expressions in the post questionnaire are composed of positive & negative tones together to prevent participants from random voting. The 7-point Likert scale is recommended for the created post questionnaire similar to existing usability tests.

The results of the surveys and analysis of the recordings are combined to evaluate the usability of the prototype and suggest improvements.

Table 1: Post-questionnaire for a proposed usability test protocol

GENERAL STRUCTURE
The prototype makes the things I want to accomplish easier to get done*a
The prototype meets my needs*a
I would recommend this prototype to a friend*a
I needed to learn a lot of things before I could get going with this prototype*b
I found the prototype very cumbersome to use*b
I think that I would need the support of a technical person to be able to use this prototype*b
I would imagine that most people would learn to use this prototype very quickly*b
I found the prototype unnecessarily complex*b
I thought the prototype was easy to use*b
I thought that I would like to use this prototype frequently*b
This prototype has all the functions and capabilities I expect it to have*c
I have to spend too much time correcting things with this prototype*d
I found the prototype difficult*e
The prototype corrected my mistakes easily*e
The prototype was designed for all levels of users*e
The prototype was too slow to react to my input*e
Learning to operate the prototype was difficult*e
USER INTERFACE
The interface of this prototype was pleasant*c
The organization of information on the prototype screens was clear*c
The sequence of screens of the prototype was confusing*e
The use of terms throughout the prototype was inconsistent*e
The position of messages on the screen of the prototype was consistent*e

Prompts for input were confusing[*e]
Error messages were helpful[*e]
Help messages on the screen of the prototype were helpful[*e]
MR ENVIRONMENT
I had trouble finding virtual objects overlaid on physical elements[*g]
I had a hard time selecting virtual objects/elements[*g]
I was distracted by pop-up windows that block physical elements[*g]
I was able to navigate wherever I wanted[*g]
The mixed reality prototype included all assets in the test environment[*g]
The mixed reality prototype included all supporting information about assets in the test environment[*g]
The duration to overlay and render virtual elements on physical elements was very long[*g]
The reaction time of the pop-up windows was too long[*g]
I could easily turn the virtual elements on and off[*g]
The speed of detecting virtual elements in place was satisfactory[*g]
The speed of accessing the data of virtual elements was satisfactory[*g]
The speed of entering data related to virtual elements was satisfactory[*g]
I find the messaging functionality of the mixed reality prototype useful[*g]
I find the video chat function of the mixed reality prototype useful[*g]
I find the multi-user functionality of the mixed reality prototype useful[*g]
The mixed reality prototype enabled instant rather than periodic communication[*g]
I find the used screen size sufficient[*g]
I prefer the hands-free/one-handed/two-handed use of the mixed reality prototype[*g]
I prefer to use the prototype via a MR head-set[*g]
I prefer to use the prototype via a MR smart glass[*g]
I prefer to use the prototype via a tablet or smartphone[*g]
OPEN ENDED
What functionality would you add to the MR prototype? *g
What functionality would you extract from the MR prototype? *g
Would you prefer the traditional method or this prototype for for long term usage? *g

[*a:] *Adopted from USE;* [*b:] *Adopted from SUS;* [*c:] *Adopted from PSSUQ;* [*d:] *Adopted from UMUX;* [*e:] *Adopted from QUIS;* [*f:] *ASQ;* [*g:] *Added by authors to customize to MR;*

CONCLUSIONS

In this paper, a usability test protocol was developed to measure the performance of prototypes that are integrated with a mixed reality environment and used in the facility management phase.

A literature review was performed to identify the usability test protocols that are used in the domain and investigate the existing usability test questionnaires used for product evaluation and human-computer interaction. Based on the results of the literature review, a comprehensive usability test protocol was proposed for the evaluation of mixed reality environments in the facility management phase. While post questionaries are used in the evaluation of mixed reality or other immersive environments in the literature, there is no protocol specialized for a mixed reality environment to be used in facility management domain. This study presents the developed usability test steps and fills this gap in the literature.

The main steps of this test protocol are 1) cognitive walkthrough, 2) pre-test surveys, 3) usability test video recording 4) post-test surveys. In the first step, one of the researchers defines the user profile and determine the task actions and another researcher performs the tasks. Video recording can be taken during the cognitive walkthrough. In the second step, a pre-test survey is conducted to determine the profiles of the participants. The questions can include age, gender, previous experience with MR environment, etc. In the third stage, a usability test is carried out in the laboratory environment. At this stage, the tasks defined by the researchers are performed by the participants and a video recording is taken to collect quantitive data such as task completion duration, number of errors, number of correct task completion. In the last stage, a post-survey is performed for qualitative evaluations. A sample post questionnaire survey has been prepared by integrating existing usability questionnaires in the literature and by combining customized statements for testing the MR environment.

The developed usability test protocol can be used by the researchers that will test the usability of an MR-based prototype which is specially developed for the facility management phase.

REFERENCES

Albertazzi, D., Okimoto, M. L., & Ferreira, M. G. G. (2012) ‘Developing an usability test to evaluate the use of augmented reality to improve the first interaction with a product’, *Work*, 41(Supplement 1), pp.1160-1163.

Alshamari, M. and Mayhew, P. (2008) ‘Task design: Its impact on usability testing’, in *2008 Third International Conference on Internet and Web Applications and Services*. IEEE, pp. 583–589.

Ammari, K. El and Hammad, A. (2014) ‘Collaborative BIM-based markerless mixed reality framework for facilities maintenance’, in *Computing in Civil and Building Engineering (2014)*, pp. 657–664.

Artan, D., Yılmaz, N., and Ergen, P. E. (2018) '*Identification of Queries for Decision Making in Building Operation and Maintenance*' in *5th International Project and Construction Management Conference (IPCMC 2018)*.

Ball, L. H. and Bothma, T. J. D. (2017) ‘A Usability Evaluation of the Prototype Afrikaanse idiome-woordeboek’, *Lexikos*. Bureau of the WAT, 27(1), pp. 78–106.

Bangor, A., Kortum, P. and Miller, J. (2009) ‘Determining what individual SUS scores mean: Adding an adjective rating scale’, *Journal of usability studies*. Citeseer, 4(3), pp. 114–123.

Brooke, J. (1996) ‘SUS: a “quick and dirty’usability’, *Usability evaluation in industry*. CRC press, p. 189.

Chalhoub, J. and Ayer, S. K. (2018) ‘Using Mixed Reality for electrical construction design communication’, *Automation in Construction*. Elsevier, 86, pp. 1–10.

Chin, J. P., Diehl, V. A. and Norman, K. L. (1988) ‘Development of an instrument measuring user satisfaction of the human-computer interface’, in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 213–218.

Chu, M., Matthews, J. and Love, P. E. D. (2018) ‘Integrating mobile building information modelling and augmented reality systems: an experimental study’, *Automation in Construction*. Elsevier, 85, pp. 305–316.

Circuit Stream. 2020. *What Is Mixed Reality — All You Need To Know | Circuit Stream*. [online] Available at: <https://circuitstream.com/blog/guide-to-mixed-reality/> [Accessed 20 January 2021].

Constantinescu, G. *et al.* (2019) ‘Usability testing of an mHealth device for swallowing therapy in head and neck cancer survivors’, *Health informatics journal*. SAGE Publications Sage UK: London, England, 25(4), pp. 1373–1382.

Desurvire, H., Kondziela, J. and Atwood, M. E. (1992) ‘What is gained and lost when using methods other than empirical testing’, in *Posters and short talks of the 1992 SIGCHI conference on Human factors in computing systems*, pp. 125–126.

Ergen, E., Kula, B., Guven, G., and Artan, D. (2021). 'Formalization of Occupant Feedback and Integration with BIM in Office Buildings', *Journal of Computing in Civil Engineering*, *35*(1), pp. 04020055.

Fernandez, A., Insfran, E. and Abrahão, S. (2011) ‘Usability evaluation methods for the web: A systematic mapping study’, *Information and software Technology*. Elsevier, 53(8), pp. 789–817.

Finstad, K. (2006) ‘The system usability scale and non-native English speakers’, *Journal of usability studies*. Usability Professionals’ Association Bloomingdale, IL, 1(4), pp. 185–188.

Finstad, K. (2010) ‘The usability metric for user experience’, *Interacting with Computers*. Oxford University Press Oxford, UK, 22(5), pp. 323–327.

Grier, R. A. *et al.* (2013) ‘The system usability scale: Beyond standard usability testing’, in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications Sage CA: Los Angeles, CA, pp. 187–191.

Iorio, J. *et al.* (2011) ‘Factors impacting usage patterns of collaborative tools designed to support global virtual design project networks’, *Journal of Information Technology in Construction (ITcon)*, 16(14), pp. 209–230.

Irizarry, J. *et al.* (2013) ‘InfoSPOT: A mobile Augmented Reality method for accessing building information through a situation awareness approach’, *Automation in construction*. Elsevier, 33, pp. 11–23.

Kostaras, N. N. and Xenos, M. N. (2009) ‘Assessing the usability of augmented reality systems’, in *Proceedings of the 13th Panhellenic Conference on Informatics*, pp. 197–201.

Krug, S. (2000) *Don’t make me think!: a common sense approach to Web usability*. Pearson Education India.

Kumar, B. A. and Mohite, P. (2018) ‘Usability of mobile learning applications: a systematic literature review’, *Journal of Computers in Education*. Springer, 5(1), pp. 1–17.

Lewis, J. R. (1991) ‘User satisfaction questionnaires for usability studies: 1991 manual of directions for the ASQ and PSSUQ’, in *Tech. Rep. No. 54.609*. International Business Machines Corporation Boca Raton.

Lewis, J. R. (1992) ‘Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ’, in *Proceedings of the Human Factors Society Annual Meeting*. SAGE Publications Sage CA: Los Angeles, CA, pp. 1259–1260.

Lewis, J. R. (1995) ‘IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use’, *International Journal of Human-Computer Interaction*. Taylor & Francis, 7(1), pp. 57–78.

Lewis, J. R. (2018) ‘Measuring perceived usability: The CSUQ, SUS, and UMUX’, *International Journal of Human–Computer Interaction*. Taylor & Francis, 34(12), pp. 1148–1156.

Lewis, J. R. and Sauro, J. (2009) ‘The factor structure of the system usability scale’, in *International conference on human centered design*. Springer, pp. 94–103.

Lewis, J. R., Utesch, B. S. and Maher, D. E. (2013) ‘UMUX-LITE: when there’s no time for the SUS’, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2099–2102.

Lund, A. M. (2001) ‘Measuring usability with the use questionnaire12’, *Usability interface*, 8(2), pp. 3–6.

Martínez, H. and Bandyopadhyay, P. (2014) ‘Analysis of Four Usability Evaluation Methods Applied to Augmented Reality Applications’, *Tampere University of Technology, Tampere*.

Molich, R. and Dumas, J. S. (2008) ‘Comparative usability evaluation (CUE-4)’, *Behaviour & Information Technology*. Taylor & Francis, 27(3), pp. 263–281.

Nielsen, J. (1994) *Usability engineering*. Morgan Kaufmann, p. 358.

Olbrich, M. *et al.* (2013) ‘Augmented reality supporting user-centric building information management’, *The visual computer*. Springer, 29(10), pp. 1093–1105.

Rankohi, S. and Waugh, L. (2013) ‘Review and analysis of augmented reality literature for construction industry’, *Visualization in Engineering*. Springer, 1(1), p. 9.

Razak, F. H. A. *et al.* (2010) ‘Usability testing with children: Laboratory vs field studies’, in *2010 International Conference on User Science and Engineering (i-USEr)*. IEEE, pp. 104–109.

Rieman, J., Franzke, M. and Redmiles, D. (1995) ‘Usability evaluation with the cognitive walkthrough’, in *Conference companion on Human factors in computing systems*, pp. 387–388.

Rogers, Y., Sharp, H. and Preece, J. (2011) ‘Interaction design: beyond human-computer interaction’. John Wiley & Sons Hoboken.

Shen, L. *et al.* (2016) ‘Human-Building Interaction (HBI): A User-Centered Approach to Energy Efficiency Innovations’, in *ACEEE Summer Study on Energy Efficiency in Buildings*.

Shi, Y. *et al.* (2016) ‘A multiuser shared virtual environment for facility management’, *Procedia Engineering*. Elsevier, 145, pp. 120–127.

Shneiderman, B. and Plaisant, C. (2010) *Designing the user interface: strategies for effective human-computer interaction*. Pearson Education India.

Tullis, T. S. and Stetson, J. N. (2004) ‘A comparison of questionnaires for assessing website usability’, in *Usability professional association conference*. Minneapolis, USA.

Walden, A. *et al.* (2020) ‘User-centered design principles in the development of clinical research tools’, *Clinical Trials*. SAGE Publications Sage UK: London, England, 17(6), pp. 703–711.

Wang, X. (2008) ‘Using cognitive walkthrough procedure to prototype and evaluate dynamic menu interfaces: A design improvement’, in *2008 12th International Conference on Computer Supported Cooperative Work in Design*. IEEE, pp. 76–80.

Yilmaz, N. (2020) *Determining The Requirements Of Using Mixed Reality Technology In Facility Management.* M.Sc. Thesis. İstanbul Technical University in İstanbul.

Zhang, J. *et al.* (2003) ‘Using usability heuristics to evaluate patient safety of medical devices’, *Journal of biomedical informatics*. Elsevier, 36(1–2), pp. 23–30.

Zhou, R. and Chan, A. H. S. (2017) ‘Using a fuzzy comprehensive evaluation method to determine product usability: A proposed theoretical framework’, *Work*. IOS Press, 56(1), pp. 9–19.

Zou, Z., Arruda, L. and Ergan, S. (2018) ‘Characteristics of models that impact transformation of BIMs to virtual environments to support facility management operations’, *Journal of Civil Engineering and Management*, 24(6), pp. 481–498.