Journal on Multimodal User Interfaces

# Developing a Scenario-Based Video Game Generation Framework for Computer and Virtual Reality Environments: A Comparative Usability Study

--Manuscript Draft--

| | |
|---|---|
| **Manuscript Number:** | |
| **Full Title:** | Developing a Scenario-Based Video Game Generation Framework for Computer and Virtual Reality Environments: A Comparative Usability Study |
| **Article Type:** | Original Paper |
| **Corresponding Author:** | Elif Surer<br>Orta Dogu Teknik Universitesi<br>TURKEY |
| **Corresponding Author Secondary Information:** | |
| **Corresponding Author's Institution:** | Orta Dogu Teknik Universitesi |
| **Corresponding Author's Secondary Institution:** | |
| **First Author:** | Elif Surer |
| **First Author Secondary Information:** | |
| **Order of Authors:** | Elif Surer |
| | Mustafa Erkayaoğlu |
| | Zeynep Nur Öztürk |
| | Furkan Yücel |
| | Emin Alp Bıyık |
| | Burak Altan |
| | Büşra Şenderin |
| | Zeliha Oğuz |
| | Servet Gürer |
| | H. Şebnem Düzgün |
| **Order of Authors Secondary Information:** | |
| **Funding Information:** | Horizon 2020 Framework Programme (740521) — Dr Elif Surer |
| **Abstract:** | Serious games ---games that have additional purposes rather than only entertainment--- aim to educate, solve, and plan several real-life tasks and circumstances in an interactive, efficient, and user-friendly way. Emergency training and planning provide structured curricula, rule-based action items, and interdisciplinary collaborative entities to imitate and teach real-life tasks. This rule-based structure enables the curricula to be transferred into other systematic learning platforms. Although emergency training includes these highly structured and repetitive action responses, a general framework to map the training scenarios' actions, roles, and collaborative structures to serious games' game mechanics and game dialogues, is still not available. To address this issue, in this study, a scenario-based game generator, which maps domain-oriented tasks to game rules and game mechanics, was developed. Also, two serious games (i.e., Hospital game and BioGarden game) addressing the training mechanisms of Chemical, Biological, Radiological, Nuclear, and Explosives (CBRNe) domain, were developed by both the game developers and the scenario-based game generator for comparative analysis. Finally, the outcomes of these games were mapped to the virtual reality (VR) environment to provide a thorough training programme. To test the usability, immersion, presence and technology acceptance aspects of the proposed |

| | |
|---|---|
| | game generator's outcomes, 15 participants tested a complete set of games and answered the questionnaires of the corresponding phenomenon. The results show that although the game generator uses higher CPU time, memory usage, and rendering time, it highly outperforms the game development pipeline performance of the game developers and provides usable and immersive games. Thus, this study provides a promising game generator which bridges the CBRNe practitioners and game developers to transform real-life training scenarios into video games efficiently and quickly. |
| **Suggested Reviewers:** | |

Title Page

**Noname manuscript No.**
(will be inserted by the editor)

# Developing a Scenario-Based Video Game Generation Framework for Computer and Virtual Reality Environments: A Comparative Usability Study

**Elif Surer* · Mustafa Erkayaoğlu · Zeynep Nur Öztürk · Furkan Yücel · Emin Alp Bıyık · Burak Altan · Büşra Şenderin · Zeliha Oğuz · Servet Gürer · H. Şebnem Düzgün**

Received: 24 May 2020 / Accepted: date

**Corresponding Author: Elif Surer**
**Department of Modeling and Simulation, Graduate School of Informatics,**
**Middle East Technical University, 06800 Ankara, Turkey**
**Tel.: +90-312-2107888**
**Fax: +90-312-2103745**
**E-mail: elifs@metu.edu.tr**

M. Erkayaoğlu
Department of Mining Engineering, Middle East Technical University, 06800 Ankara, Turkey
Tel.: +90-312-2105812
Fax: +90-312-2105822
E-mail: emustafa@metu.edu.tr

Z. N. Öztürk
Department of Computer Engineering, Bilkent University, 06800 Ankara, Turkey
Tel.: +90-312-2901218
Fax: +90-312-2664047
E-mail: nur.ozturk@ug.bilkent.edu.tr

F. Yücel
Department of Modeling and Simulation, Graduate School of Informatics,
Middle East Technical University, 06800 Ankara, Turkey
Tel.: +90-312-2103740
Fax: +90-312-2103745
E-mail: furkan.yucel@metu.edu.tr

E. A. Bıyık
Department of Modeling and Simulation, Graduate School of Informatics,
Middle East Technical University, 06800 Ankara, Turkey
Tel.: +90-312-2103740
Fax: +90-312-2103745
E-mail: emin.biyik@metu.edu.tr

B. Altan
Department of Modeling and Simulation, Graduate School of Informatics,
Middle East Technical University, 06800 Ankara, Turkey
Tel.: +90-312-2103740
Fax: +90-312-2103745
E-mail: burak.altan@metu.edu.tr

B. Şenderin
Department of Modeling and Simulation, Graduate School of Informatics,
Middle East Technical University, 06800 Ankara, Turkey
Tel.: +90-312-2103740
Fax: +90-312-2103745
E-mail: busra.senderin@metu.edu.tr

Z. Oğuz
Department of Psychology, Bilkent University, 06800 Ankara, Turkey
Tel.: +90-312-2903415
Fax: +90-312-2664127
E-mail: zeliha.oguz@ug.bilkent.edu.tr

S. Gürer
Department of Mining Engineering, Middle East Technical University, 06800 Ankara, Turkey
Tel.: +90-312-2105812
Fax: +90-312-2105822
E-mail: e119605@metu.edu.tr

H. Ş. Düzgün
Colorado School of Mines, Brown Hall 268, CO 80401, USA
Tel.: +1-303-384-2357
Fax: +1-303-273-3719
E-mail: duzgun@mines.edu

**Noname manuscript No.**
(will be inserted by the editor)

# Developing a Scenario-Based Video Game Generation Framework for Computer and Virtual Reality Environments: A Comparative Usability Study

**Author List**

Received: 24 May 2020 / Accepted: date

**Abstract** Serious games —games that have additional purposes rather than only entertainment— aim to educate, solve, and plan several real-life tasks and circumstances in an interactive, efficient, and user-friendly way. Emergency training and planning provide structured curricula, rule-based action items, and interdisciplinary collaborative entities to imitate and teach real-life tasks. This rule-based structure enables the curricula to be transferred into other systematic learning platforms. Although emergency training includes these highly structured and repetitive action responses, a general framework to map the training scenarios' actions, roles, and collaborative structures to serious games' game mechanics and game dialogues, is still not available. To address this issue, in this study, a scenario-based game generator, which maps domain-oriented tasks to game rules and game mechanics, was developed. Also, two serious games (i.e., Hospital game and BioGarden game) addressing the training mechanisms of Chemical, Biological, Radiological, Nuclear, and Explosives (CBRNe) domain, were developed by both the game developers and the scenario-based game generator for comparative analysis. Finally, the outcomes of these games were mapped to the virtual reality (VR) environment to provide a thorough training programme. To test the usability, immersion, presence and technology acceptance aspects of the proposed game generator's outcomes, 15 participants tested a complete set of games and answered the questionnaires of the corresponding phenomenon. The results show that although the game generator uses higher CPU time, memory usage, and rendering time, it highly outperforms the game development pipeline performance of the game developers and provides usable and immersive games. Thus, this study provides a promising game generator which bridges the CBRNe practitioners and game developers to transform real-life training scenarios into video games efficiently and quickly.

**Keywords** Serious Games · Video Game Generator · CBRNe · System Usability Scale · Technology Acceptance Model

## 1 Introduction

Serious gaming [1][2], the umbrella term describing the video and board games having additional goals rather than only entertainment, is widely used in several domains such as health [3], defense [4][5], and education [6][7]. The most common focus of the serious games has been the games' physiological and spatial effects on the players [4]. Several studies demonstrate that playing video games can improve spatial, cognitive, and motor skills [8][9]. Some studies show that these improvements could affect the players in the long term [10][11]. Game technologies and serious games have been used in different disciplines for the purposes of simulation and training, and these technologies have been adapted to not only computer environments, but also virtual reality (VR) environments. Milgram et al. [12] provided a detailed scheme to better identify the differences between different realities and VR has been defined as *"Computer generated artificial simulations, generally recreation of the real environment"*. Using the game technologies in VR enhanced the research on spatial tasks while also enabling the research on immersion, presence, and usability.

There are some research domains where serious gaming and VR have recently been introduced to the scene and CBRNe is one of them. CBRNe is an acronym for

Address(es) of author(s) should be given

Chemical, Biological, Radiological, Nuclear, and Explosives, and recent research on this domain focuses on personnel training, emergency planning and organizing of field, tabletop, simulation, and serious gaming exercises for preparedness [13]. In CBRNe exercises and training, repetiteveness, structure, scenarios, roles of the practitioners are very well designed. Although these trainings have such well-defined structures, serious gaming and VR have not taken these pre-defined rules into consideration while creating well-defined games.

In this study, we propose a scenario-based game generator where the actors, scenarios and the locations of the CBRNe training exercises are mapped to state diagrams and compose simple games. At the beginning, these "plain" games are generated with simple shapes, in order to test the correctness of the scenario only. Then, these simple shapes are easily replaced with real game assets using the tag information of the game objects. To compare the results of this plain game generator and asset game generator with real games, two of the games are also separately developed by game developers and game generator's outcome and the game developer's outcome were compared in terms of performance, usability, and technology acceptance. Finally, game generator's and game developer's outcomes were mapped to VR and additional presence, immersion and usability test were also performed on users.

## 2 Literature Survey

The idea behind the serious gaming term can be traced back to the Renaissance era, where the philosophers used the term "serio ludere" which can be translated as "serious play (theatre)" —i.e., using humor in plays to emphasize serious matters [14]. The first time "Serious Game" was used —as closest to its current meaning— in Clark Abt's Serious Games [15] book. In his book, Abt described serious games as follows:

> *"Games may be played seriously or casually. We are concerned with serious games in the sense that these games have an explicit and carefully thought-out educational purpose and are not intended to be played primarily for amusement. This does not mean that serious games are not, or should not be, entertaining."* [15]

Abt's serious games approach was used in order to solve real life educational, governmental and industrial problems [16]. Early serious games were mostly named simulations and their main focus was to simulate military scenarios —i.e., a training and testbed for dangerous-to-live and emergency scenarios [17].

One of the use cases of serious gaming in emergency planning is on firefighter training. In a study by Heldal [18], firefighter training was examined using serious games and tools. To do so, qualitative questionnaires and observations on two use cases (i.e., ship evacuation in Baltic Sea and railway accident with cyanide leakage) were used to analyze the impacts of serious gaming on non-users. Results showed that serious games would be useful in emergency training situations, and in-depth training scenarios and evaluation methods were necessary. In another study, Lukosch et al. [19] performed the steps of the traditional design process with the contributions of the end-users. The primary purpose of the study was to check if the simulations could be used to train situational awareness skills, and the end-user participation simulation demonstrated the positive impact of using simulations. However, the main limitation of the study was not having a game-scenario-based approach, and future research would focus on this aspect while creating virtual agents.

The use of VR simulation was also a common topic in the literature. Ingrassia et al. [20] focused on testing and comparing performances of 56 medical students during mass casualty triage in real-world and VR. The results showed that VR and live simulation were both useful in improving the accomplishments of the medical students. Ragazzoni et al. [21] also focused on VR training's medical aspect, where the objective was to increase the staff safety in life-or-death risks. Hybrid simulation for infection control and Ebola treatment were also successfully performed virtually, and the results demonstrated that the awareness of the health personnel increased.

Serious gaming in CBRNe has been a recent topic, and there are some misinterpretations on the definitions and core concepts, such as the misuse of the words 'game' or 'simulation'. To overcome these misinterpretations, a pre-development survey was developed [22] to be used before implementing the serious games of the European Network Of CBRN TraIning CEnters (eNOTICE) project [23]. In the pre-development survey, 24 questions were asked to the practitioners and experts of CBRNe under the following subgroups: 1) Participant's video gaming background, 2) Participant's knowledge on serious games, and 3) Participant's expectations on eNOTICE serious games. Results from 14 CBRNe professionals showed that the majority of the participants were highly positive on using serious games in CBRNe and provided open concepts, suggestions, and guidelines to develop serious games for CBRNe domain [24].

Evaluating developed tools, games and the related environments is necessary to check if the games realized their primary objectives. Among several evaluation methods, the main aspects that are evaluated in the literature are Presence, Immersion, Technology Acceptance and Usability. Presence can be defined as a person's experience of being in a place while he or she is in another place. This definition was redefined for virtual environments as a person's experience of being in a virtual, computer generated world while he or she is in the real world [25]. To evaluate presence in VR-based games, a detailed questionnaire on presence for virtual environments is widely used [26]. Involvement and immersion are two important aspects measured in the questionnaire and they explain the relationship of involvement and immersion states with presence as the user's degree of focus and the user's involvement on the virtual environment. Another measure, Technology Acceptance Model (TAM) was introduced by Davis [27] and TAM became a widely accepted model for studying user's acceptance of technology. TAM focuses on perceived usefulness and perceived ease-of-use and was finalized by Vanketash and Davis [28]. Evaulating usability is another important aspect of virtual environment research and Brooke's [29] proposed questionnaire method, System Usability Scale (SUS), has been used widely to evaluate usability on websites, applications, and virtual environments.

In this study, a game generator that maps linear real-life scenarios to serious games with training objectives is generated as the scenario-based game generator. The crucial part of this study is the CBRNe domain, where the roles, tasks, and goals of the actors are clearly defined. The fundamental objectives of the serious games are as follows: 1) Providing a new environment for further training, 2) Building synergy, and 3) Adopting distinctive territories and different concepts to the CBRNe community. Two games are implemented with both this scenario-based game generator and also by game developers, which are real scenarios of eNOTICE project's joint activities in Nimes (France) and Brussels (Belgium). The analysis of the study was performed based on their CPU usage, rendering, memory usage, and game development pipeline on the generator-based and developer-based games. Also, two phases of the software development were tested by 15 participants in terms of usability, technology acceptance, immersion and presence. The quantitative comparative analyses on the game generator, game developer-based games and their virtual environment versions were analyzed thoroughly followed by qualitative comments and suggestions from the users.

## 3 Materials and Methods

In this section, the details of the scenario-based game generator (Figures 1-4), two serious games that were developed by both the game designers and the game generator, and their evaluation are explained in detail.

### 3.1 Scenario and Task Definitions

The theme of the scenario, active players, location, and interaction mechanisms were collected in advance from the practitioners. Workflow and state diagrams were used to create a detailed scenario where different roles and entities communicate with each other. At the beginning, two different CBRNe scenarios that were based on real practices were mapped to the workflow and state diagram structures. One of the scenarios is based on the subset of the BioGarden exercise (i.e., linear version of the scenario in which the players do not change the flow of the events), which was held in June 2018 in Belgium as part of eNOTICE joint activities. The other scenario is based on the Nimes exercise, which was held in France in January 2018 again as part of eNOTICE joint activities.

Defined scenarios and interaction mechanisms were mapped to game ideas, linear game stories, and interaction mechanisms. Interaction mechanisms were converted into concrete tasks and user roles. A generic system, built on top of the initial scenario definitions, was conceptualized and implemented. Then, the generated system was fine-tuned with goals, feedback measurements, and score adaptations.

The scenarios of the exercises were designed by different institutions such as fire departments, research centers, and hospitals. Thus, breaking the scenarios down into actions and events was a crucial step so that the game mechanics, reward mechanisms, and scoring could be systematized. Also, different roles in the scenarios were assigned to different player types so that the active role of the player and the role of non-player characters (NPCs) were defined.

Before starting to implement the Hospital game, a detailed survey, which was briefly mentioned above [24], was conducted on 14 professionals and researchers from CBRNe field working in several European Union projects —7 of them being game players. The scope and the purposes of the study were as follows: 1) Learning the user's gamer profile, 2) Understanding the user's perspective on serious games, 3) Retrieving the expectations of the user, 4) Clarifying the differences between the video/serious games and simulations, and 5) Asking for suggestions. The initial results of the survey showed

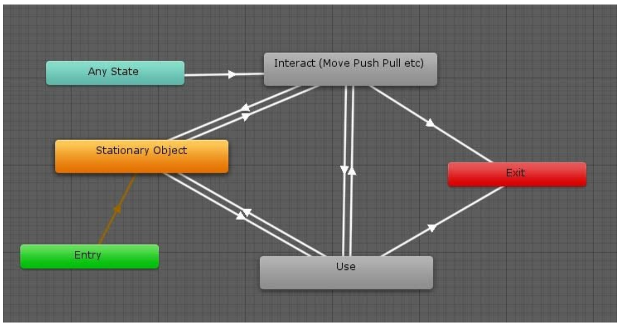



Fig. 1: Initial interaction mechanism including Entry and Exit States, a Stationary Object interacting (i.e., Move, Push, Pull, etc.) and using stationary objects.

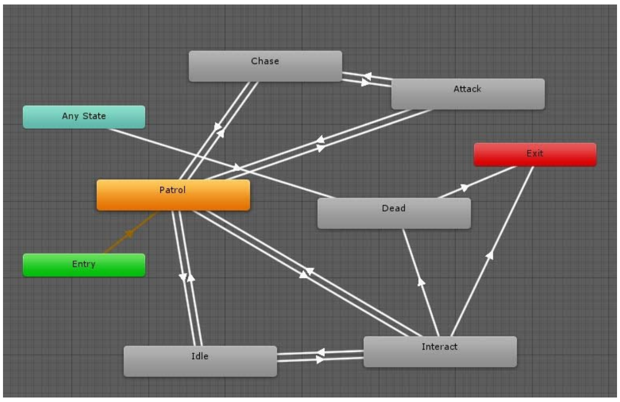



Fig. 2: In an attack scenario, chasing, attacking and interacting use cases are modeled.

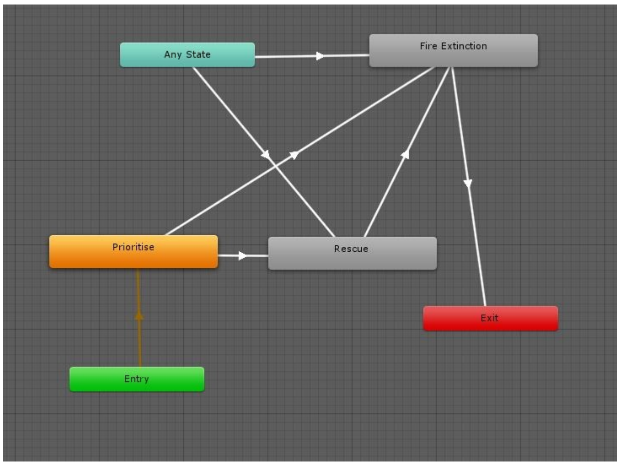



Fig. 3: Fire fighting scenario includes prioritizing the steps, extinguishing the fire and rescuing the affected people.

that the expectations were highly positive and the community was willing to adapt new technologies to their daily tasks. The participants' gamer profiles involved playing strategy games and multiplayer games to learn

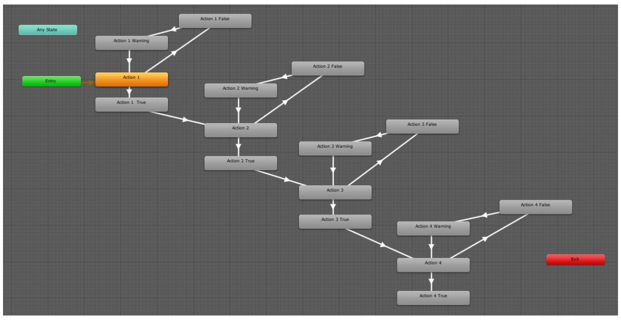



Fig. 4: A generic scenario including various different states.

new skills and to relieve stress. After the detailed analysis of the results, a tutorial mode was added to the initial game prototype.

### 3.2 Scenario-Based Game Generator

Scenario-based game generator is developed in combination with Unity Real Time Development Platform's Animator Controller tool and is composed of four different components: 1) Main Code, 2) Control Code, 3) Transition Code, and 4) User Interface. Main Code is where the state definitions and structures —the definitions of the final consequences of the actions— are initialized. The actions and consequences on this part are based on the task definitions of the scenarios which were created during the previous step. Control Code works as a mechanism to form and map action methods and their related states.This part of the code checks the current state and the condition if there is an action which creates any transition. Then, Transition Code is the link where the game generator works with Unity's Animator Controller. Transition Code maintains a connection between previous codes and the Animator Controller by triggering the animations. Finally, additional user interface mechanisms such as feedback, scores, and health points are added and grouped under the User Interface component (Figure 5). Due to the needs of VR environment, user interface of the virtual reality versions of the games have some differences in terms of locations and size of the elements, and camera location.

While developing the scenario-based game generator tool, the following steps are executed: 1) Creating an environment, 2) Defining the state diagrams, 3) Creating animations, 4) Adding basic artificial intelligence (AI) to states, 5) Resolving player and NPC interactions and 6) Adding basic AI for interactions.

In this scenario-based game generator, only linear scenarios, where the decision making of players do not modify the outcomes of the actions, are implemented.

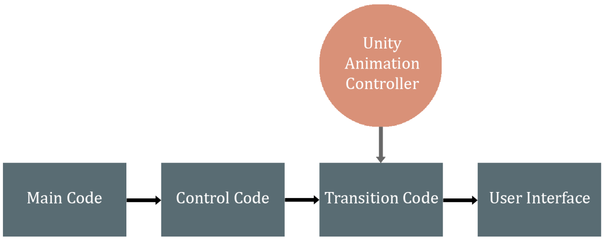



Fig. 5: Structure of the scenario-based game generator.

The scenario-based game generator is used to generate the duplicates of Hospital and BioGarden games. First, the game scenarios are tested in prototypes by using simple 3D game objects such as cubes, spheres, and capsules (Figures 6 and 7). Then, after checking that the scenario works correctly, initial 3D game objects are automatically replaced with game assets using tag information of the assets.

### 3.3 Development Environment

Unity3D has been used in games developed by both the game generator and game developers. Unity3D is a game engine that offers 2D and 3D game development environment for different platforms. Unity3D also has VR and augmented reality support. It has plug-in and SDK support for various VR glasses and headsets. The SteamVR SDK developed by Valve is one of the supported SDKs, and it enables playing and developing VR games by using the headsets such as Oculus Rift and HTC Vive. In Phase-II of the study, the games were switched to VR versions using the SteamVR SDK, and HTC Vive headset was used during the usability evaluation.

### 3.4 Design Patterns

In both games, players choose a role at the start of the game. The tasks that they have to perform according to the selected role and the interaction with the game world are different. For example, if the player is a nurse in the Hospital game, the doctor and other health personnel act as NPC following the requirements of the scenario. Likewise, if the player chooses to be a doctor, other characters in the scenario act as NPC. Thus, some characters can be both a playable character and NPC, which depends on the player's choice. The developers used the State Design Pattern to manage this interaction network in the game world.

Design Patterns [30] are general solutions used for common problems in software engineering. These solutions can be considered as templates that can be customized by the requirements of the project. State Design Pattern provides an answer to such a problem: In software design, if there are objects with different states and these objects behave differently depending on the state, these behavior transitions can be resolved using State Design Pattern.

This design pattern provides a solution to manage the interaction network in the games. The objects are characters and the state of a character is defined based on whether the character is NPC or not. If it is not NPC, the tasks are expected to be performed by the player. However, NPC automatically performs tasks in accordance with the player's progress in the scenario. The use of this design pattern in the game design has saved both from the workload and time loss that would require coding separately the playable and NPC versions of each role.

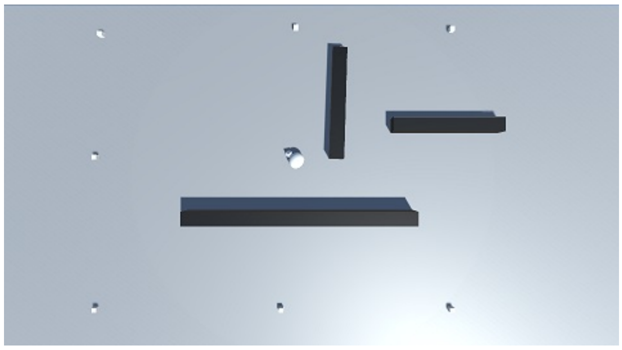


Fig. 6: Initial tests of the scenario-based game generator were performed on simple game objects.

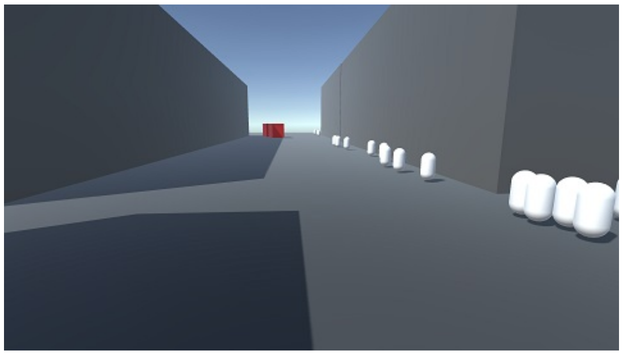


Fig. 7: Initial tests of the scenario-based game generator were performed on simple game objects such as cubes and capsules.

### 3.5 Hospital Game

Hospital Game is based on the Nimes scenario, which was performed during the eNOTICE joint activity in

January 2018. The main purpose of this scenario is training the medical staff for CBRNe circumstances. In this game, players can play different roles like doctor and nurse to learn taking security measures sıch as using masks and gloves, blocking entrance of the hospital and applying decontamination procedures. In the game, player as a nurse meets the patients and learns their complaints. The nurse decides the room in which the patient will be located with the guidance of the doctor. As a doctor player also listens the complaints of the patients via the nurse. After checking the resources of the hospital such as rooms and beds, by the help of the secretary of the hospital, the doctor decides whether the patient will be located in a room. If the player chooses the right actions, she/he gets points. The game is based on a linear scenario where the player needs to follow strict hospital regulations (Figures 8, 9 and 10). The player learns taking security measures such as using gloves, using masks, and blocking the entrance of the hospital and applying decontamination procedures. Players can play different roles, such as a doctor, nurse, and secretary. It is based on a linear scenario, and when the player makes wrong choices, they lose game points (Figures 8, 9 and 10). This game is converted to the VR platform to make simulation more realistic for users which will help their learning process of these crucial steps.

This game was developed by a second-year Middle East Technical University (METU) Multimedia Informatics graduate program student, and the same game scenario was also given to the scenario-based game generator. The initial results of both environments were compared. In both versions of the games, Quadart's Hospital Lowply pack [31], which provides several realistic modular assets, was used.

Fig. 8: Screenshots from the Hospital game (Game Generator version).


Fig. 9: Screenshots from the Hospital game and a demo of Dialogue menu (Game Developer version).

Fig. 10: Screenshots from the Hospital game and a demo of the interaction mechanism (Game Generator version).

### 3.6 BioGarden Game

BioGarden game (Figures 11 and 12) is based on the eNOTICE joint activity, which was played in Belgium in June 2018. Although it had a nonlinear scenario, only linear parts of the scenario were implemented so that a comparison with the scenario-based game generator would be possible. In the scenario, there were different laboratories with different structures and responsibilities. The role-playing part was composed of decontamination, role assignment, and evaluation.

There are two player options in the game. Local Investigation Team and Decontamination Team. Local Investigation Team is responsible for detecting samples, collecting samples, sending samples with a drone to the laboratory, evaluating the sample results. Decontamination Team is responsible from the decontamination of the teams that are contamined and sending results to the laboratory. There are two scenes in the game: Menu scene and game scene. Menu scene (Figure 13) has options such as; player selection and scenario summary. UI of the game has several elements such as; back button for main menu, camera selection button which toggles the game camera as first person/third person, playing as panel which shows the current player in the

Fig. 11: The interior design of the Candestine lab from the BioGarden game and menu interactions (Game Generator version).

Fig. 12: The interior design of the Clandestine lab from the BioGarden game (Game Generator version).

game. The bottom bar has objective list and score of the player. Right Mouse click opens an action menu (Figure 14) which has all possible actions for the current team. Whenever the player selects an action, if the action is correct, character moves towards the objective area, information panel (Figure 15) shows the description of the action and points of the player increases. If the action is wrong, information panel indicates that this action is wrong, and the player loses points. Two different camera options are implemented for different platforms since the first person camera can provide a better immersion if the game is played on a VR headset.

As in the case of the Hospital game, BioGarden game was developed by a second-year METU Multimedia Informatics program student, and the same game scenario was also given to the scenario-based game generator, and the initial results were compared. In both versions of the games, 3LB Games' Low Poly laboratory pack [32], which provides several realistic models, textures, and diffuse maps, was used. Both of them prepared in the same version of the Unity Real Time Development Platform. Both of the games are played by

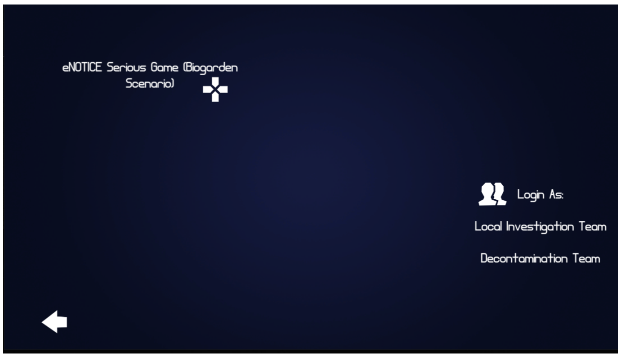



Fig. 13: Initial menu of the Biogarden game (Game Developer version).

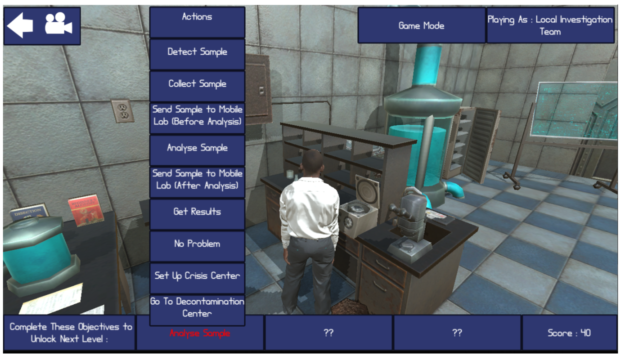



Fig. 14: Action menu of the Biogarden game (Game Developer version).


Fig. 15: User interface of the Biogarden game (Game Developer version).

the same gamers and their feedback is collected which will be discussed in the results part.

## 4 Results and Discussion

### 4.1 Comparative Performance Outcomes

In this section, the performance outcomes of the scenario-based game generator and the two serious games that

Table 1: CPU Usage of the Hospital Game Generated by the Scenario-Based Video Game Generator vs. Hospital Game

| CPU Usage | Hospital (Generator) | Hospital (Game) |
|---|---|---|
| CPU | 9 ms | 4 ms |

Table 2: CPU Usage of the BioGarden Game Generated by the Scenario-Based Video Game Generator vs. BioGarden Game

| CPU Usage | Biogarden (Generator) | Biogarden (Game) |
|---|---|---|
| CPU | 22.6 ms | 8.5 ms |

Table 3: Memory Usage of the Hospital Game Generated by the Scenario-Based Video Game Generator vs. Hospital Game

| Memory | Hospital (Generator) | Hospital (Game) |
|---|---|---|
| Used Total | 1.03 GB | 0.45 GB |
| Reserved Total | 1.32 GB | 0.63 GB |
| System Memory Usage | 1.96 GB | 1.38 GB |

Table 4: Memory Usage of the BioGarden Game Generated by the Scenario-Based Video Game Generator vs. BioGarden Game

| Memory | Biogarden (Generator) | Biogarden (Game) |
|---|---|---|
| Used Total | 0.62 GB | 0.28 GB |
| Reserved Total | 0.92 GB | 0.49 GB |
| System Memory Usage | 1.65 GB | 1.27 GB |

Table 5: Rendering Results of the Hospital Game Generated by the Scenario-Based Video Game Generator vs. Hospital Game

| Rendering | Hospital (Generator) | Hospital (Game) |
|---|---|---|
| SetPass Calls | 106 | 136 |
| Draw Calls | 252 | 298 |
| Total Batches | 217 | 243 |
| Triangles | 463.9K | 504.9K |
| Vertices | 319.5K | 361.1K |

Table 6: Rendering Results of the BioGarden Game Generated by the Scenario-Based Video Game Generator vs. BioGarden Game

| Rendering | Biogarden (Generator) | Biogarden (Game) |
|---|---|---|
| SetPass Calls | 1826 | 2200 |
| Draw Calls | 2584 | 3007 |
| Total Batches | 2584 | 3007 |
| Triangles | 3.1M | 3.2M |
| Vertices | 2.3M | 2.4M |

were developed by the game developers were compared in terms of CPU usage, rendering time, memory usage, and game development pipeline. All the tests were performed on a laptop having Intel Core i7 9750HQ CPU, 16GB RAM, and NVIDIA GeForce GTX 1660TI graphics.

Table 1 and Table 2 present the comparison on CPU usage (i.e., game generator's output vs. developer-based game), Table 3 and Table 4 on memory usage, and finally, Table 5 and Table 6 present a comparison using rendering parameters. The rendering profile uses the SetPass Calls, Draw Calls, Total Batches, Triangles, and Vertices as parameters. SetPass parameter is defined as "the number of rendering passes" [33], a Draw Call as a "call to the graphics API to draw objects" [33] and Batch as a "package with data that will be sent to the GPU" on the Unity's Renderer Profiler page [33].

It took three weeks to develop and implement the scenario-based game generator. The most time-consuming part was the state transitions and handling the outcomes of the actions. After the game generator was built, it took three and a half hours to generate the Hospital game and four hours to generate the BioGarden game using the game generator.

The development of the original BioGarden game took 25 days in total: one week for the scenario clarification and role assignment; one week for the text-based decision mechanisms, nine days to finish the user interface and menus and two days to add the assets to the game.

The original Hospital game was first refactored from the initial prototype, which took one week. Then, it took another two weeks adapting the new scenario, merging different game modes, and dialogue generation.

In this study, real-life exercise scenarios of two eNOTICE joint activities were developed by game developers as well as a scenario-based game generator —developed during this study. All the frameworks used the same scenario-to-game mechanics mapping. All four versions of the games (i.e., developed by the game developer vs. generated by the game generator) were compared in terms of CPU usage, memory usage, rendering, and game development timeline perspectives. Scenario-based game generator's CPU usage, memory usage, and rendering time were higher when compared with the developer-based games. Due to the tag searching process and the need for more resources to trigger Unity's Animator Controller, game generator uses higher resources in terms of memory and CPU. The rendering performance results of both versions of the games were very similar because the working principle of the game generator was not dependent on the visual contents of the games. Although the game generator used higher memory, CPU, and rendering time, its game development timeline efficiency highly outperformed the game developers (i.e., four hours vs. three weeks). This is a highly promising outcome [34] that will enable further exercise scenarios to be mapped into games in a short period of time. This outcome can benefit the practitioners in two ways: 1) Visualizing the action-state diagrams of the exercise so that they can see the flaws or unassigned roles of their exercises, and 2) Having a rapid game prototype which becomes a fast, interactive testbed and training tool.

## 4.2 Usability Tests on PC versions - Phase I

Technology Acceptance Model (TAM) questionnaire covered descriptive questions to determine the level of technology acceptance representing the benefits obtained from the computer environment versions of the games together with profile questions to classify users. The age profile of the participants in the questionnaire is between 20 and 36 with an equal distribution for age ranges of 20-26 and 27-36. This enabled the researchers to investigate the relationship between age and technology acceptance with equal weights. Participants of the workshop hold a Bachelor of Science degree in different disciplines related to computer engineering, informatics, computer science, and others with varying experience in game development.

The experience profile regarding game development is dominated by the participants of the workshop on multi-modal interfaces with 70%. A similar profile of these participants could also be observed from 55% of the responses about having experience on VR environments. In a typical week, 25% of the participants spend four or more days playing at least 30 minutes of video games whereas the remaining spends 0-1 days or 2-3 days on a regular basis. The participants responded to a total of 28 questions in the survey classified as "system usability" and "technology acceptance". System Usability Scale (SUS) level is investigated by the first 10 questions and the remaining 18 questions are used to define the basis of the Technology Acceptance Model (TAM) of the two different games, Hospital and Biogarden (See Appendix).

Table 7: Basic statistical evaluation of the answers about SUS

| Game | Measurements | Q4 Q10 | Q2 Q6 Q8 | Q1 Q3 Q5 Q7 Q9 |
|---|---|---|---|---|
| HG | Mean | 1.81 | 1.71 | 4.23 |
| | Std. Dev. | 0.22 | 0.08 | 0.36 |
| HG-P | Mean | 2.13 | 1.92 | 3.36 |
| | Std. Dev. | 0.32 | 0.23 | 0.13 |
| HG-A | Mean | 1.94 | 1.69 | 4.00 |
| | Std. Dev. | 0.01 | 0.35 | 0.37 |

### *4.2.1 Phase I - Hospital Game*

The participants completed three different questionnaires after completing gaming sessions for the Hospital game as follows:

1. Questionnaire about hospital game (HG) 2. Questionnaire about hospital game with plain game generator (HG-P) 3. Questionnaire about hospital game with asset game generator (HG-A)

Questions are grouped into positive and negative concepts as even and odd numbered that are interpreted in different ways for the calculation of the SUS scores. These scores are rated between the values of 0-100 but should not be interpreted as percentage values. The SUS scores calculated are 81 for HG, 67 for HG-P and 78 for HG-A. Grade ranking of SUS scores are evaluated based on the scale developed by [35][36] and classified as "OK" for HG-P and "good" for HG and HG-A [37]. It is stated by Brooke [29][37] that the systems with higher SUS scores are better in terms of usability. This indicates that the respondents considered the Hospital Game with plain game generator as a comparably lower usable game in the dedicated game session.

Similar to the positive concept related questions, responses to the even numbered questions were evaluated for the learnability of the system and the usability of the system. The basic statistical evaluation of the questionnaire results was grouped based on this point of view is given in Table 7.

As seen in Table 7, questions related to learnability (Q4-Q10) indicate that the plain game generator had higher scores in the negative concept representing a lower perception by the respondents. Responses to the odd numbered questions in the positive concept for the Hospital Game were comparably higher than the game with plain and asset generator. This indicated that the respondents found the Hospital Game to have higher usability. This conclusion was also supported by the remaining eight questions divided into two as positive and negative questions for evaluation. The responses to the positive questions were found to be close to the highest scale value of five and the answers of the negative questions were found similarly to be close to the lowest scale value of one for the SUS model. The questionnaires were also evaluated for the TAM representing the three different versions of the Hospital Game.

There are various methodologies that can be followed to define the reasons of users to accept and prefer to adapt to a new technology. One of the most commonly utilized concepts is the Technology Acceptance Model (TAM) developed by Davis [28]. It is based on a theory of information systems and explains how people or society accepts a technology with a theoretical model. The evaluation with TAM evaluates the theories behind the application and acceptance of new technologies especially in information technologies (IT) based on the theory of reasoned action [38]. The theory is considered to be developed during studies on attitude in expectation value models [27] that explain perceived usefulness and perceived ease-of-use together with intention to use and user satisfaction/perceived enjoyment. The second part of the questionnaire is related with the TAM evaluated by a 10-point scale (1 - Strongly Disagree, 10 - Strongly Agree). The statistical evaluation of the TAM related responses started with a reliability analysis that aims to assess the consistency between the responses to questions with similar objective on an ordinal scale. According to [39], the Cronbach's Alpha ($\alpha$) value found for reliability should be greater than 0.7 as a threshold and $\alpha > 0.8$ indicates that the questionnaire has good reliability. The reliability test for the TAM related questions was calculated as 0.9631 indicating good reliability. Another evaluation of the TAM related part of the questionnaire was conducted by comparing the mean values of the responses.

The groups HG vs HG−P and HG vs HG−A are found to be statistically zero considering the responses to build the TAM by two-sample t-test (independent t-test). The null hypothesis (H0) is defined as "the difference of the mean values between HG and HG−P is equal to zero." In other words, the technology acceptance level of both groups is equal for these games (H0:$\mu$HG - $\mu$HG−P = 0). The alternative hypothesis (H1) is two-tailed as the assumption is that the difference is not equal to zero (H1:$\mu$HG - $\mu$HG−P$\neq$0). Table 9 summarizes the two sample t-test scores for the Hospital Game and the game with plain game generator.

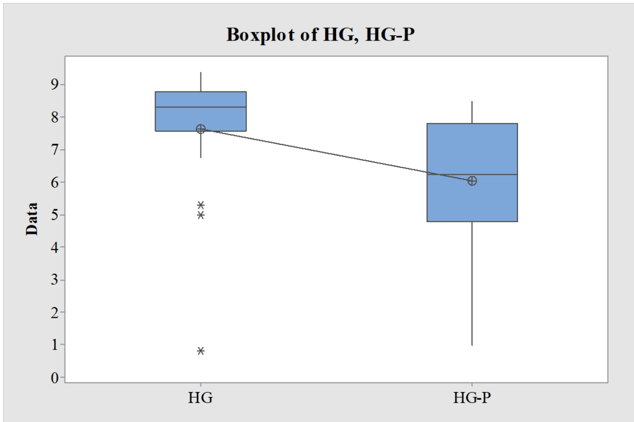
The null hypothesis H0 is rejected based on the p-value (0.026) lower than 0.05, the significance level (cut-off) value set for 95% confidence interval, and the t-two tailed value (2.032) is lower than t value (2.327). As a conclusion, the difference between HG and HG-P was found to be statistically significant and the hypothesis "H1: There is a statistically significant difference between HG and HG-P regarding technology acceptance level of this game." is accepted. The individual and box plots of are represented in Figure 16.

Table 8: Two-sample t-test scores of HG and HG-P

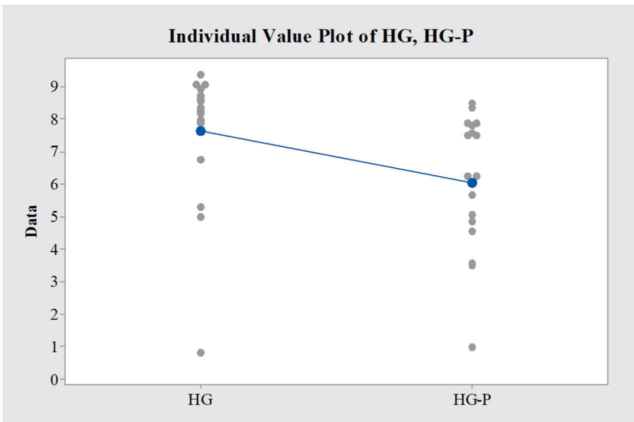
| Game | Mean | Std. Dev. | Std. Error Mean |
|---|---|---|---|
| HG | 7.64 | 2.09 | 0.49 |
| HG-P | 6.03 | 2.06 | 0.49 |

Table 9: Two-sample t-test scores of HG and HG-P

| Game | t | t-two tailed | df | p | 95% CI Lower | 95% CI Upper |
|---|---|---|---|---|---|---|
| HG HG-P | 2.327 | 2.032 | 34 | 0.026 | 0.202 | 3.015 |




(a) Box Plot of HG and HG-P.




(b) Individual Plot of HG and HG-P.

Fig. 16: (a) Box Plot of HG and HG-P and (b) Individual Plot of HG and HG-P.

Table 10: Two-sample t-test scores of HG and HG-A

| Game | Mean | Std. Dev. | Std. Error Mean |
|---|---|---|---|
| HG | 7.64 | 2.09 | 0.49 |
| HG-A | 7.26 | 2.19 | 0.52 |

Table 11: Two-sample t-test scores of HG and HG-A

| Game | t | t-two tailed | df | p | 95% CI Lower | 95% CI Upper |
|---|---|---|---|---|---|---|
| HG HG-A | 0.530 | 2.032 | 34 | 0.600 | -1.074 | 1.831 |

The potential outliers within the responses seen in Figure 16, represent the question that has a negative concept represented by a lower score compared to the remaining positive concept. Therefore, the data were considered to be suitable for performing t-test. Similarly, the Hospital Game was evaluated together with the game with asset generator.

The null hypothesis is failed to be rejected based on the statistical data given in Table 11 as the p-value (0.600) is greater than the significance value of 0.05 and the t-two tailed value (2.032) is greater than the t-value (0.530). This can be explained as the difference between HG and HG-A not being statistically significant where "H1: There is a statistically significant difference between HG and HG-A regarding technology acceptance level of this game" is rejected. The individual and box plots for HG and HG-A are given in Figure 17.

The question with the negative concept is representing the lower value and it was considered that the responses were suitable for applying t-test. As a conclusion, the respondents of the questionnaire indicated that the plain game generator was more adaptable than the asset game generator version of the Hospital Game. This was also observed in the similarity between the responses related to HG-A and HG, whereas the responses of HG-P are relatively different. The second game related to the Biogarden scenario was also evaluated in detail.

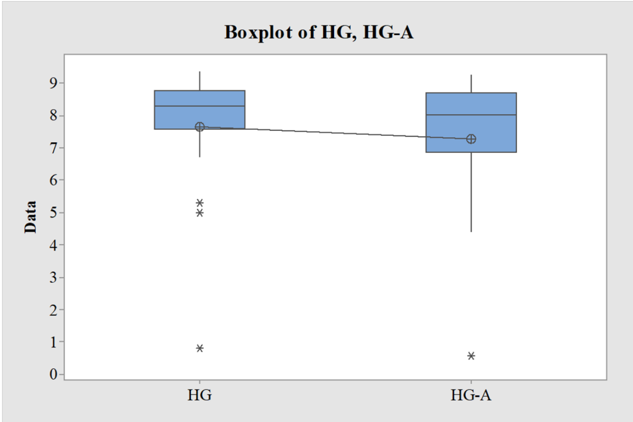
#### *4.2.2 Phase I - Biogarden Game*

The gaming sessions were finalized with the following questionnaires completed by the participants;

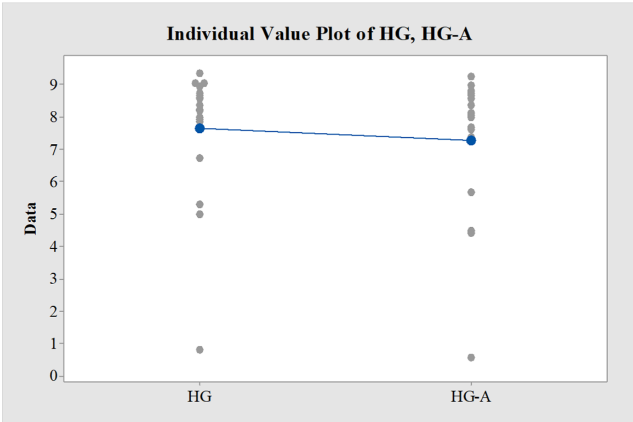
1. Questionnaire about Biogarden game (BG) 2. Questionnaire about Biogarden game with plain game generator (BG-P) 3. Questionnaire about Biogarden game with asset game generator (BG-A)

The SUS scores were calculated as 79 for BG, 72 for BG-P, and 80 for BG-A which indicated that the plain game generator was evaluated to have lower usability compared to the other versions. Table 12 summarizes the basic statistical evaluation of the questionnaire results for the Biogarden game.

The plain game generator was scored higher by the respondents for the questions related to learnability (Q4-Q10). The Biogarden game had higher scores in the questions defined in the positive concept indicating that the usability was higher than the versions with




(a) Box Plot of HG and HG-A.




(b) Individual Plot of HG and HG-A.

Fig. 17: (a) Box Plot of HG and HG-A and (b) Individual Plot of HG and HG-A.

Table 12: Basic statistical evaluation of the answers about SUS

| Game | Measurements | Q4 Q10 | Q2 Q6 Q8 | Q1 Q3 Q5 Q7 Q9 |
|---|---|---|---|---|
| BG | Mean | 1.78 | 1.69 | 4.08 |
| | St. Dev. | 0.08 | 0.11 | 0.36 |
| BG-P | Mean | 1.81 | 1.54 | 3.41 |
| | St. Dev. | 0.11 | 0.28 | 0.21 |
| BG-A | Mean | 1.69 | 1.65 | 4.06 |
| | St. Dev. | 0.09 | 0.06 | 0.30 |

Table 13: Two-sample t-test scores of BG and BG-P

| Game | Mean | Std. Dev. | Std. Error Mean |
|---|---|---|---|
| BG | 7.48 | 2.06 | 0.48 |
| BG-P | 5.60 | 2.06 | 0.48 |

Table 14: Two-sample t-test scores of HG and HG-A

| Game | t | t-two tailed | df | p | 95% CI Lower | 95% CI Upper |
|---|---|---|---|---|---|---|
| BG BG-P | 2.739 | 2.032 | 34 | 0.010 | 0.483 | 3.272 |

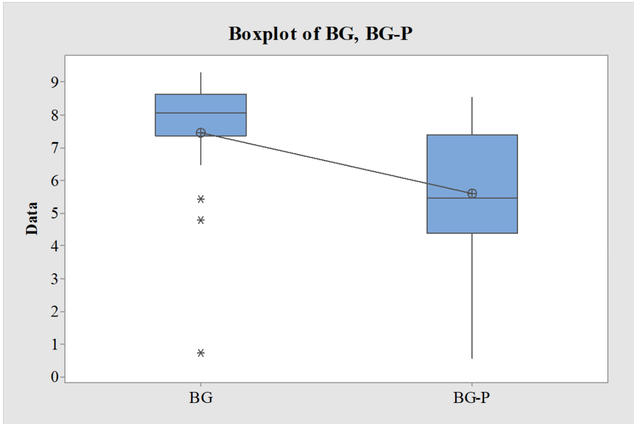
game generators. These results are similar to the responses of the Hospital Game and point out the consistency of the questionnaire results. The questions representing the usability of the system were found to be close to the highest scale value of 5 for the positive concept and similarly close to the lowest scale value of 1 for the negative concept questions. The perspective of the respondents was used as the basis of TAM for the Biogarden game.

The questionnaire part related to the TAM was first analyzed for the reliability of the results. The Cronbach's alpha value of TAM responses for the Biogarden game is calculated as 0.9705 representing a good level of accuracy. This evaluation was followed by a two-sample t-test for BG vs. BG-P and BG vs. BG-A. The null hypothesis (H0) was defined as the difference of the mean values of the responses related to TAM of BG and BG-P is equal to zero. Test scores of this comparison is summarized in Tables 13 and 14.

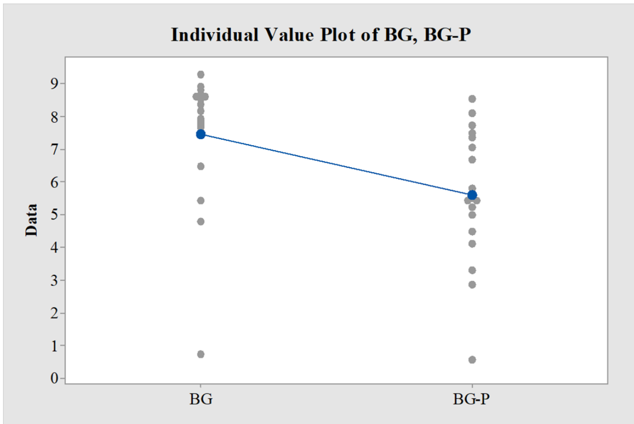
The null hypothesis is rejected as the calculated p-value (0.010) is lower than the cutoff value of 0.05 and the t-two tailed value (2.032) is lower than the t value (2.739). The difference between BG and BG-P was found as statistically significant and it was concluded that the alternative hypothesis, "There is a statistically significant difference between BG and BG-P regarding technology acceptance level of this game", is accepted. Figure 18 represents the individual and box plots of the BG and BG-P.

This distribution of the responses support the concept that the t-test is suitable for the evaluation of these responses. A similar assessment was performed for the TAM related responses of BG and BG-A with the same hypothesis definition and is summarized in Table 15.

Based on the data given in Table 16, the H0 hypothesis is accepted as the p-value (0.584) is greater than the cut-off 0.05 for 95% confidence interval and the t-two tailed value (2.032) is greater than the t value (0.584). As a result, the difference between BG and




(a) Box Plot of BG and BG-P.




(b) Individual Plot of BG and BG-P.

Fig. 18: (a) Box Plot of BG and BG-P and (b) Individual Plot of BG and BG-P.

Table 15: Two-sample t-test scores of BG and BG-A

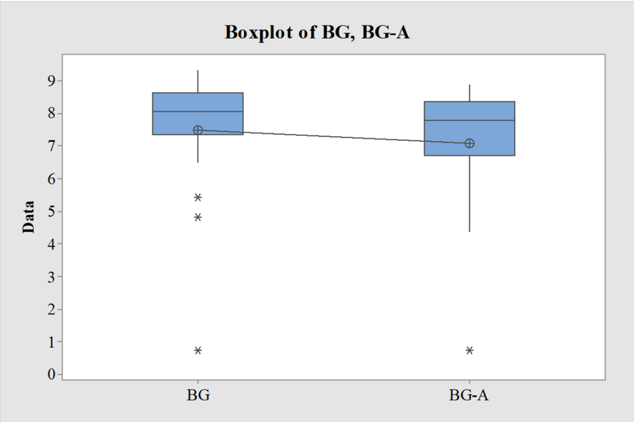
| Game | Mean | Std. Dev. | Std. Error Mean |
|---|---|---|---|
| BG | 7.48 | 2.04 | 0.48 |
| BG-A | 7.10 | 2.19 | 0.48 |

Table 16: Two-sample t-test scores of BG and BG-A

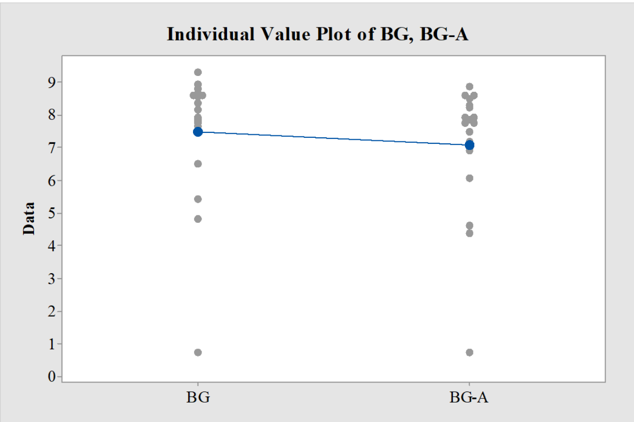
| Game | t | t-two tailed | df | p | 95% CI Lower | 95% CI Upper |
|---|---|---|---|---|---|---|
| BG BG-A | 0.553 | 2.032 | 34 | 0.584 | -1.011 | 1.766 |

BG-A was not statistically significant and the hypothesis that there is a statistically significant difference between BG and BG-A regarding technology acceptance level is rejected. The individual and box plots of the compared game versions are given in Figure 19.

As a conclusion of the Phase-I evaluation of the developed games, the plain game generator was found to




(a) Box Plot of BG and BG-A.




(b) Individual Plot of BG and BG-P.

Fig. 19: (a) Box Plot of BG and BG-A and (b) Individual Plot of BG and BG-A.

be more suitable than the asset game generator. This could be assessed from the similarity between the responses to BG and BG-A together with the differences of the responses given to questions related to BG-P. A similar result could also be observed for the Hospital game and its versions with the plain game generator and the asset game generator.

### 4.3 Usability Tests on VR versions - Phase II

The gaming sessions were finalized with the questionnaires that aimed to evaluate the experience of the participants in a systematic manner. The Phase-II of the questionnaire was related to the VR games developed for the Hospital and Biogarden scenarios. The participants were asked to complete questionnaires about the VR games and their versions with the asset game generator.

Similar to the Phase-I of the questionnaire, the VR version of the Hospital game (HG-VR) and the version with the asset game generator (HG-A-VR) were assessed according to their usability. The SUS scores were calculated as 79 for HG-VR and 80 for HG-A-VR which are both rated as "good" based on their performance in terms of usability. The basic statistical evaluation of the questionnaire results grouped based on this point of view is given in Table 17.

Table 17: Basic statistical evaluation of the answers about SUS

| Game | Measurements | Q4 Q10 | Q2 Q6 Q8 | Q1 Q3 Q5 Q7 Q9 |
|---|---|---|---|---|
| HG-VR | Mean | 1.75 | 1.81 | 3.94 |
| | St. Dev. | 0.60 | 0.28 | 0.28 |
| HG-A-VR | Mean | 1.57 | 1.69 | 3.91 |
| | St. Dev. | 0.39 | 0.39 | 0.54 |

Table 18: Two-sample t-test scores of HG-VR and HG-A-VR

| Game | Mean | Std. Dev. | Std. Error Mean |
|---|---|---|---|
| HG-VR | 7.36 | 1.48 | 0.37 |
| HG-A-VR | 7.31 | 1.63 | 0.41 |

The questions Q4-Q10 are related to learnability and indicate a comparably lower score for HG-A-VR representing a better end result as these questions are negative oriented. The odd numbered questions that have a positive concept can be evaluated as HG-VR having a better score. The remaining questions aim to consider the usability of the system where the positive concept questions were responded with a score close to the highest scale value of 5 and the responses to the negative concept questions were close to the lowest scale of 1. The technology acceptance model (TAM) for the HG-VR and HG-A-VR were also evaluated together with the reliability of this part of the questionnaire.

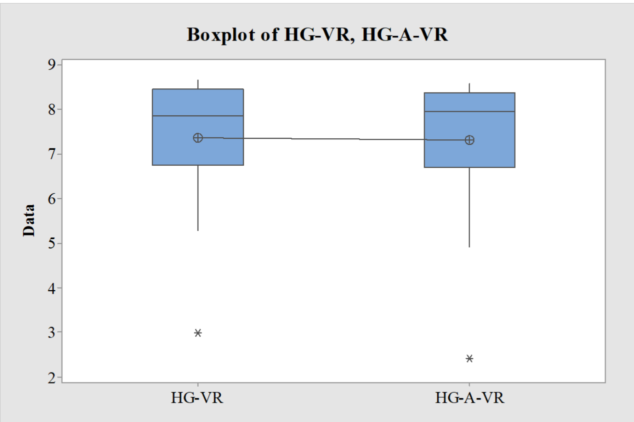
The Cronbach's alpha value of TAM responses for HG-VR and HG-A-VR are calculated as 0.9954. The difference between the mean values of the responses was evaluated by a two-sample t-test to consider whether it is statistically zero. The null hypothesis (H0) is defined for the condition of "the difference of the mean values between HG-VR and HG-A-VR is equal to zero". This could also be interpreted as the technology acceptance level of both groups being equal for the Hospital game. In this case, the alternative hypothesis (H1) is two-tailed because it assumes that the difference is not equal to zero. The results are summarized in Table 18 and Table 19.

The null hypothesis is failed to be rejected as the p-value (0.929) is greater than 0.05 and the t-two tailed

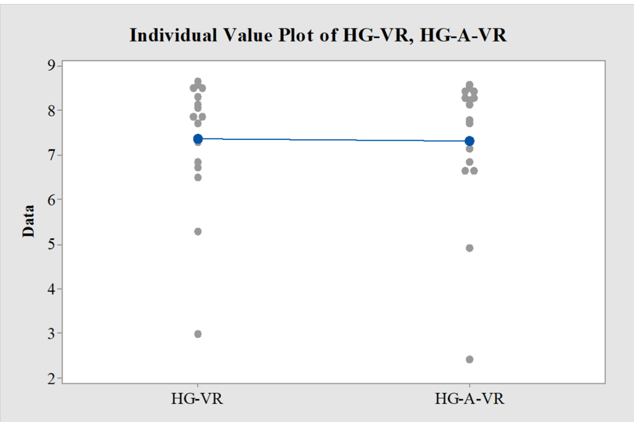
Table 19: Two-sample t-test scores of HG-VR and HG-A-VR

| Game | t | t-two tailed | df | p | 95% CI Lower | 95% CI Upper |
|---|---|---|---|---|---|---|
| HG-VR HG-A-VR | 0.090 | 2.042 | 30 | 0.929 | -1.079 | 1.178 |

value (2.042) is greater than t value (0.090). As a result of the two-sample t-test, the difference between HG-VR and HG-A-VR was not statistically significant. Therefore, “H1: There is a statistically significant difference between HG-VR and HG-A-VR regarding technology acceptance level of this game” is rejected. The individual and box plots are given in Figure 20.




(a) Box Plot of HG-VR and HG-A-VR.




(b) Individual Plot of HG-VR and HG-A-VR.

Fig. 20: (a) Box Plot of HG-VR and HG-A-VR and (b) Individual Plot of HG-VR and HG-A-VR.

The responses provided for the questionnaire related to HG-VR and HG-A-VR were considered as suitable for the implementation of the two-sample t-test. A similar methodology was also used for the evaluation of the VR version of the Biogarden game (BG-VR) and the version with the asset game generator (BG-A-VR).

Table 20: Basic statistical evaluation of the answers about SUS

| Game | Measurements | Q4 Q10 | Q2 Q6 Q8 | Q1 Q3 Q5 Q7 Q9 |
|---|---|---|---|---|
| BG-VR | Mean | 1.75 | 1.74 | 4.09 |
| | St. Dev. | 0.60 | 0.08 | 0.30 |
| BG-A-VR | Mean | 1.57 | 1.71 | 4.12 |
| | St. Dev. | 0.39 | 0.27 | 0.42 |

Table 21: Two-sample t-test scores of HG-VR and HG-A-VR

| Game | Mean | Std. Dev. | Std. Error Mean |
|---|---|---|---|
| BG-VR | 7.56 | 1.66 | 0.42 |
| BG-A-VR | 7.61 | 1.77 | 0.43 |

Table 22: Two-sample t-test scores of HG-VR and HG-A-VR

| Game | t | t-two tailed | df | p | 95% CI Lower | 95% CI Upper |
|---|---|---|---|---|---|---|
| BG-VR BG-A-VR | 0.097 | 2.042 | 30 | 0.924 | -1.276 | 1.161 |

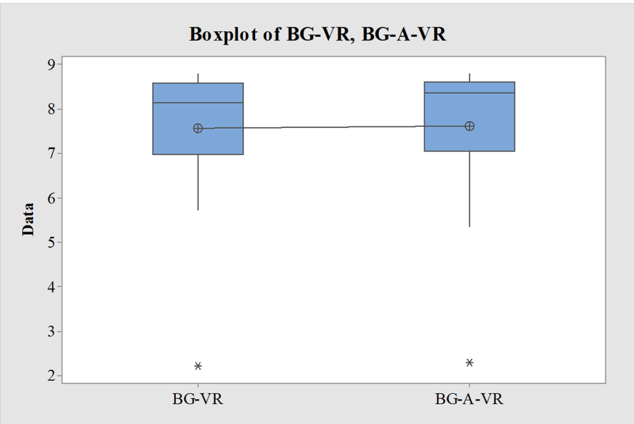
The SUS scores were calculated as 81 for BG-VR and 81 for BG-A-VR, classified as having a “good” ranking. The learnability of the system was assessed by the remaining questions and the results are summarized in Table 20.

BG-A-VR was considered as having a better experience compared to BG-VR whereas the positive concept questions indicated that the participants classified BG-VR as having a higher score. This could be interpreted as the fact that the VR versions of the Biogarden game were highly similar by structure and the participants could not differentiate the usability of the system distinctly. The Technology Acceptance Model (TAM) was also assessed for BG-VR and BG-A-VR.

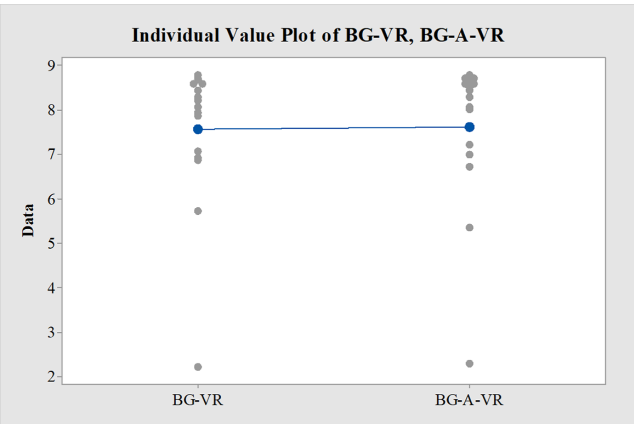
The Cronbach’s alpha value of TAM responses are calculated as 0.9977 indicating a high level of reliability. The evaluation was continued with an independent t-test to compare the two versions of the Biogarden game based on the null hypothesis (H0) , “the difference of the mean values between BG-VR and BG-A-VR is equal to zero.” The results of the t-test are presented in Table 21 and Table 22.

The null hypothesis defined for the t-test was failed to be rejected as the p-value (0.924) is greater than

0.05 and the t-two tailed value (2.042) is greater than t value (0.097). As a result, the two-sample t-test concluded that the difference between BG-VR and BG-A-VR was not statistically significant. The individual and box plots are given in Figure 21.




(a) Box Plot of BG-VR and BG-A-VR.




(b) Individual Plot of BG-VR and BG-A-VR.

Fig. 21: (a) Box Plot of BG-VR and BG-A-VR and (b) Individual Plot of BG-VR and BG-A-VR.

The assessment of the developed games was continued with a presence evaluation of the virtual environments. The Immersive Tendencies Questionnaire (ITQ) was originally introduced to measure the capability or tendency of participant involvement. The Presence Questionnaire (PQ) aims to measure the experience of the participants by means of presence in a virtual environment together with the possible contributing factors. The questionnaire has a seven-point scale based on the semantic differential principle [40]. Participants are commonly asked to place a mark in the appropriate box of the scale in accordance with the question content and descriptive labels. 28 questions were available in the questionnaire from which 14 are

Table 23: Two-sample t-test scores of HG-VR and HG-A-VR

| Category | Hospital | | Hospital Asset Generator | |
|---|---|---|---|---|
| PQ Main Factor | Score | Std. Dev. | Score | Std. Dev. |
| Control | 64.64 | 1.58 | 62.86 | 1.72 |
| Sensory | 32.71 | 1.91 | 33.29 | 1.95 |
| Distraction | 19.86 | 1.82 | 21.00 | 1.58 |
| Realism | 4.43 | 1.60 | 4.43 | 1.83 |
| Total Score | 121.64 | | 121.57 | |

Table 24: Two-sample t-test scores of HG-VR and HG-A-VR

| Category | Biogarden | | Biogarden Asset Generator | |
|---|---|---|---|---|
| PQ Main Factor | Score | Std. Dev. | Score | Std. Dev. |
| Control | 66.21 | 1.54 | 65.71 | 1.69 |
| Sensory | 34.50 | 1.81 | 34.86 | 1.88 |
| Distraction | 20.71 | 1.68 | 21.07 | 1.63 |
| Realism | 4.93 | 1.21 | 5.14 | 1.10 |
| Total Score | 126.36 | | 126.79 | |

Table 25: Two-sample t-test scores of HG-VR and HG-A-VR

| Category | Hospital | | Hospital Asset Generator | |
|---|---|---|---|---|
| ITQ Subscale | Score | Std. Dev. | Score | Std. Dev. |
| Focus | 34.21 | 1.52 | 34.36 | 1.53 |
| Involvement | 22.71 | 1.71 | 22.71 | 1.71 |
| Game | 8.29 | 2.10 | 8.29 | 2.10 |
| Total Score | 65.21 | | 65.36 | |

related to control factors (CF), 8 are related sensory factors (SF), 5 are about distraction factors (DF) and 1 question represents the realism factors (RF). The results are summarized in Table 23 and Table 24.

The results indicated that both of the developed games have no significant difference by means of presence based on the participant responses. This is interpreted as the outcome of similar VR development methodology for both games and the almost identical level of experience of the participants related to VR. The last stage of the evaluation used ITQ items with 14 questions with a distribution of 7 for focus, 5 for involvement, and the remaining 2 for the game itself. The results are summarized in Table 25 and 26.

Similar to the PQ, the results of the ITQ represented a trend where the scores calculated for the re-

Table 26: Two-sample t-test scores of HG-VR and HG-A-VR

| Category | Biogarden | | Biogarden Asset Generator | |
|---|---|---|---|---|
| ITQ Subscale | Score | Std. Dev. | Score | Std. Dev. |
| Focus | 34.29 | 1.52 | 34.29 | 1.52 |
| Involvement | 22.71 | 1.71 | 22.71 | 1.71 |
| Game | 8.29 | 2.10 | 8.29 | 2.10 |
| Total Score | 65.29 | | 65.36 | |

sponses of the participants have no significant difference. The tendency to become involved in activities and the tendency to maintain focus on current activities were evaluated identically by the participants. This indicates the similarity between the development stages of the VR versions of the game.

### 4.4 User Comments and Suggestions

Participants played the game generator's games, Biogarden game and Hospital game both in computer and VR environments. Besides the abovementioned questionnaires, open-ended questions regarding the general comments and suggestions, were also asked to the users. The users could not spot differences between the outcomes of the game generator- and game developer- based games. One user particularly mentioned:

*"Besides the graphics, the main concept is completely the same. This is the strength of the game generator as it can easily provide a framework to prepare mini games."*

There were mixed feelings about the plain game generator. While some of them found it very difficult to follow with cubes and simple shapes, several users underlined its potential:

*"I was very impressed while playing plain games. They were almost abstract and they show the potential of the generator - when I was playing the game generator with assets, I think that this game is just one of the possible solutions."*

Some of the users found the assets to be more explanatory:

*"Assets are beautiful, it definitely helps to create a flow state and presence feeling. Assets also help me to understand what is happening in the games and what should I do. They make it easier to play the games."*

The menu items on the UI mechanism were found to be useful, but the number of the menu items on the screen was suggested to be reduced. Different actions were suggested to be triggered by different menu icons to simplify the UI. Most of the users preferred the VR version more and they said that it was more fun and immersive.

*"VR version is better. When character moves, looking around is good. It can be better when needing of UI components in order to interact with the environment is decreased."*

*"I enjoyed the VR version. It feels like you are much more involved in the environment."*

## 5 Conclusion

In this study, a scenario-based video game generator, which targets the scenarios in CBRNe domain, was developed. This initial version of the game generator used linear scenarios that were based on the joint activities of the eNOTICE project. The effectiveness of the game generator was tested in comparison with two serious games, which were developed by the game developers. Even though the performance of the game generator lacked on the rendering, memory usage, and CPU usage aspects, it highly outperformed the game development pipeline of the game developers. Besides, computer version and VR version of the games and the game generator outcomes were produced and tested by 15 participatns. System usability, technology acceptance, immersion and presence aspects of the outcomes were analyzed thoroughly. The results show that the proposed game generator produces usable and immersive games in a short amount of time. This is a highly promising result that will enable the practitioners to visualize their scenarios while also generating prototype games rapidly both for computer and VR environments so that the training of CBRNe personnel will become thorougly enriching and immersive. As a future work, the proposed game generator framework will be adapted to nonlinear scenarios.

**Acknowledgments**

**Conflict of interest**

The authors declare that they have no conflict of interest.

## References


1. R. Santos, C. Magalhaes, L. F. Capretz, J. C. Neto, F. Q. B. da Silva, A. Saher, *Computer games are serious business and so is their quality: particularities of software testing in game development from the perspective of practitioners*, 2018. arXiv preprint arXiv:1812.05164 `https://arxiv.org/abs/1812.05164`.
2. D. R. Michael and S. L. Chen, *Serious games: Games that educate, train, and inform*, Muska & Lipman/Premier-Trade, 2005.
3. M. Pirovano, E. Surer, R. Mainetti, P. L. Lanzi, N. A. Borghese, *Exergaming and rehabilitation: A methodology for the design of effective and safe therapeutic exergames*, Entertainment Computing, 14, 55-65, 2016.
4. T. Susi, M. Johannesson, P. Backlund *Serious Games – An Overview*, University of Skövde, 2007.
5. D. Crookall, *Serious games, debriefing, and simulation/gaming as a discipline*, Simulation & gaming, 41(6), 898-920, 2010.
6. R. Tinati, M. Luczak-Roesch, W. Hall, *An investigation of player motivations in Eyewire, a gamified citizen science project*, Computers in Human Behavior, 73, 527-540, 2017.
7. V. Curtis, *Motivation to participate in an online citizen science game: A study of Foldit*, Science Communication, 37(6), 723-746, 2015.
8. M. Dorval and M. Pepin *Effect of Playing a Video Game on a Measure of Spatial Visualization*, Perceptual and Motor Skills, vol. 62, no. 1, pp. 159-162, 1986.
9. R. Nelson and I. Strachan *Action and Puzzle Video Games Prime Different Speed/Accuracy Tradeoffs*, Perception, vol. 38, no. 11, pp. 1678-1687, 2009.
10. B. C. Tobler-Ammann, E. Surer, E. D. de Bruin, M. Rabuffetti, N. A. Borghese, R. Mainetti, M. Pirovano, L. Wittwer, R. H. Knols *Exergames encouraging exploration of hemineglected space in stroke patients with visuospatial neglect: a feasibility study*, JMIR serious games, vol. 5(3), p. e17, 2017.
11. B. C. Tobler-Ammann, E. Surer, R. H. Knols, N. A. Borghese, E. D. de Bruin, *User perspectives on exergames designed to explore the hemineglected space for stroke patients with visuospatial neglect: usability study*, JMIR serious games, vol. 5(3), p. e18, 2017.
12. P. Milgram, H. Takemura, A. Utsumi, F. Kishino *Augmented reality: a class of displays on the reality-virtuality continuum*, Telemanipulator and Telepresence Technologies, 1995.
13. XVR Simulation — Incident command training tool for safety and security, (September, 2019). Retrieved from `https://www.xvrsim.com/en-cn/`.
14. D. Djaouti, J. Alvarez, J. Jessel, O. Rampnoux *Origins of Serious Games*, Serious Games and Edutainment Applications, 2011.
15. C. Abt *Serious games*, Lanham: University Press of America, 1970.
16. F. Laamarti, M. Eid and A. El Saddik *An Overview of Serious Games*, International Journal of Computer Games Technology, vol. 2014, pp. 1-15, 2014.
17. D. Michael, S. Chen *Serious Games: Games That Educate, Train, and Inform*, Boston, MA: Thomson Course Technology PTR, 2006.
18. I. Heldal *Simulation and serious games in emergency management: experiences from two case studies*, 22nd International Conference on Virtual System & Multimedia (VSMM), pp. 1-9. IEEE, 2016.
19. H. Lukosch, T. van Ruijven, and A. Verbraeck, *The participatory design of a simulation training game*, Proceedings of the Winter Simulation Conference, Winter Simulation Conference, 2012.
20. P.L. Ingrassia, L. Ragazzoni, L. Carenzo, F.L. Barra, D. Colombo, G. Gugliotta, F. Della Corte *Virtual reality and live scenario simulation: options for training medical students in mass casualty incident triage*, Critical Care, 16(1), p.P479, 2012.
21. L. Ragazzoni, P.L. Ingrassia, L. Echeverri, F. Maccapani, L. Berryman, F.M. Burkle, F. Della Corte *Virtual reality simulation training for Ebola deployment*, Disaster medicine and public health preparedness, 9(5), pp.543-546, 2015.
22. User Expectations Questionnaire for eNotice Serious Game (2019, November). Retrieved from `https://ec.europa.eu/eusurvey/runner/eNoticeSeriousGamePreDevelopment`.
23. European Network Of CBRN TraIning CEnters – Official Website (2019, November). Retrieved from `https://www.h2020-enotice.eu`.
24. E. Surer, T. B. Atalay, D. Ç. Demirkan, H. Ş. Düzgün *Serious Gaming in CBRNe Domain: A Survey on User Expectations, Concerns and Suggestions*, 3rd International Conference CBRNE - Research & Innovation, Nantes, France, p.43, 2019.
25. J. Garrett *The elements of user experience*, Berkeley, CA: New Riders, 2011.
26. B. Witmer and M. Singer *Measuring Presence in Virtual Environments: A Presence Questionnaire*, Presence: Teleoperators and Virtual Environments, vol. 7, no. 3, pp. 225-240, 1998.
27. F. Davis *A technology acceptance model for empirically testing new end-user information systems: theory and results*, Doctoral dissertation, MIT Sloan School of Management, Cambridge, MA, 1986.
28. V. Venkatesh and F. Davis *A Model of the Antecedents of Perceived Ease of Use: Development and Test*, Decision Sciences, vol. 27, no. 3, pp. 451-481, 1996.
29. J. Brooke *SUS-A quick and dirty usability scale*, Usability evaluation in industry. CRC Press, 1996.
30. S. Bjork and J. Holopainen *Patterns in game design*, Hingham: Charles River Media, 2005.
31. Unity Asset Store - Hospital LowPoly, (2019, September). Retrieved from `https://assetstore.unity.com/packages/3d/environments/hospital-lowpoly-82552`.
32. Unity Asset Store - Low Poly Laboratory Pack, (2019, September). Retrieved from `https://assetstore.unity.com/packages/3d/environments/low-poly-laboratory-pack-47677`.
33. Rendering Profiler - Unity Documentation (2019, November). Retrieved from `https://docs.unity3d.com/560/Documentation/Manual/ProfilerRendering.html`.
34. E. Surer, M. Erkayaoğlu, Z. N. Öztürk, F. Yücel, E. A. Bıyık, B. Altan, B. Şenderin, Z. Oğuz, S. Gürer, H Ş. Düzgün *Developing a Scenario-Based Video Game Generation Framework: Preliminary Results*, arXiv preprint arXiv:1911.07380, 2019.
35. A. Bangor, P. T. Kortum, J. T. Miller *An empirical evaluation of the system usability scale. Intl. Journal of Human–Computer Interaction*, 24(6):574-94, 2008.
36. A. Bangor, P. T. Kortum, J. T. Miller *Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of usability studies*, 4(3):114-23, 2009.
37. J. Brooke *SUS: a retrospective. Journal of usability studies*, 8(2):29-40, 2013.

38. M. Fishbein and I. Ajzen *Understanding attitudes and predicting social behavior*, 1980.
39. M. Tavakol and R. Dennick *Making sense of Cronbach's alpha*, International journal of medical education, 2:53, 2011.
40. R. Dyer, J. J. Matthews, J. F. Stulac, C. E. Wright, K. Yudowitch *Questionnaire construction manual, annex literature survey and bibliography*, Palo Alto, CA: Operations Research Associates, 1976.

**Supplementary Materials**

| | **Presence (Answers from 1 to 7, 1 being most negative)** |
|---|---|
| Q1 | How much were you able to control events? |
| Q2 | How responsive was the environment to actions that you initiated (or performed)? |
| Q3 | How natural did your interactions with the environment seem? |
| Q4 | How much did the visual aspects of the environment involve you? |
| Q5 | How natural was the mechanism which controlled movement through the environment? |
| Q6 | How compelling was your sense of objects moving through space? |
| Q7 | How much did your experiences in the virtual environment seem consistent with your real world experiences? |
| Q8 | Were you able to anticipate what would happen next in response to the actions that you performed? |
| Q9 | How completely were you able to actively survey or search the environment using vision? |
| Q10 | How well could you actively survey or search the virtual environment using touch? |
| Q11 | How compelling was your sense of moving around inside the virtual environment? |
| Q12 | How closely were you able to examine objects? |
| Q13 | How well could you examine objects from multiple viewpoints? |
| Q14 | How well could you move or manipulate objects in the virtual environment? |
| Q15 | How involved were you in the virtual environment experience? |
| Q16 | How much delay did you experience between your actions and expected outcomes? |
| Q17 | How quickly did you adjust to the virtual environment experience? |
| Q18 | How proficient in moving and interacting with the virtual environment did you feel at the end of the experience? |
| Q19 | How much did the visual display quality interfere or distract you from performing assigned tasks or required activities? |

| | **Presence (cont'd) (Answers from 1 to 7, 1 being most negative)** |
|---|---|
| Q20 | How much did the control devices interfere with the performance of assigned tasks or with other activities? |
| Q21 | How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities? |
| Q22 | How completely were your senses engaged in this experience? |
| Q23 | To what extent did events occurring outside the virtual environment distract from your experience in the virtual environment? |
| Q24 | Overall, how much did you focus on using the display and control devices instead of the virtual experience and experimental tasks? |
| Q25 | Were you involved in the experimental task to the extent that you lost track of time? |
| Q26 | How easy was it to identify objects through physical interaction, like touching an object, walking over a surface, or bumping into a wall or object? |
| Q27 | Were there moments during the virtual environment experience when you felt completely focused on the task or environment? |
| Q28 | How easily did you adjust to the control devices used to interact with the virtual environment? |

| | **System Usability Scale (1: Strongly Disagree 5: Strongly Agree)** |
|---|---|
| Q1 | I think that I would like to use this system frequently. |
| Q2 | I found the system unnecessarily complex. |
| Q3 | I thought the system was easy to use. |
| Q4 | I think that I would need the support of a technical person to be able to use this system. |
| Q5 | I found the various functions in this system were well integrated. |
| Q6 | I thought there was too much inconsistency in this system. |
| Q7 | I would imagine that most people would learn to use this system very quickly. |
| Q8 | I found the system very cumbersome to use. |
| Q9 | I felt very confident using the system. |
| Q10 | I needed to learn a lot of things before I could get going with this system. |

| | **Immersive Tendency Questionnaire Item (Answers from 1 to 7, 1 being most negative)** |
|---|---|
| Q1 | Do you ever get extremely involved in projects that are assigned to you by your boss or your instructor, to the exclusion of other tasks? |
| Q2 | How easily can you switch your attention from the task in which you are currently involved to a new task? |
| Q3 | How frequently do you get emotionally involved (angry, sad, or happy) in the news stories that you read or hear? |
| Q4 | How well do you feel today? |
| Q5 | Do you easily become deeply involved in movies or TV dramas? |
| Q6 | Do you ever become so involved in a television program or book that people have problems getting your attention? |
| Q7 | How mentally alert do you feel at the present time? |
| Q8 | Do you ever become so involved in a movie that you are not aware of things happening around you? |
| Q9 | How frequently do you find yourself closely identifying with the characters in a story line? |
| Q10 | Do you ever become so involved in a video game that it is as if you are inside the game rather than moving a joystick and watching the screen? |
| Q11 | Are you easily disturbed when working on a task? |
| Q12 | How well do you concentrate on enjoyable activities? |
| Q13 | How often do you play arcade or video games? (OFTEN should be taken to mean every day or every two days, on average.) |
| Q14 | Do you ever become so involved in doing something that you lose all track of time? |

| | **Technology Acceptance Model (Answers from 0 to 10)** |
|---|---|
| Q1 | I liked the idea that creating a game to teach and improve user interactions with new devices. |
| Q2 | The game helped me to getting used to VR environment and its systems interaction methods. |
| Q3 | Was playing the game fun in VR? (What did you like more? What would you want to change?) |
| Q4 | During the game, I felt pain and/or discomfort. |
| Q5 | The game is challenging and exciting in VR. |
| Q6 | It's easy to learn to play the proposed game. |
| Q7 | The game reacts readily to my movements in the VR environment. |
| Q8 | I did not find it hard to interact with the virtual world using gestures/controllers. |
| Q9 | I found the graphical interface clear and explanatory. |
| Q10 | The interface is simple to use. |
| Q11 | I liked the interface design. (Any advice on how to improve it?) |
| Q12 | The instructions of the game are clear. I understood what to do in the game and how. |
| Q13 | I found it easy to reach targets. |
| Q14 | I did not have problems finding targets in the screen. |
| Q15 | I would like to learn new interaction methods of different devices with games in the future. |
| Q16 | If I had the option to keep using the system at home, I would play with them often. |