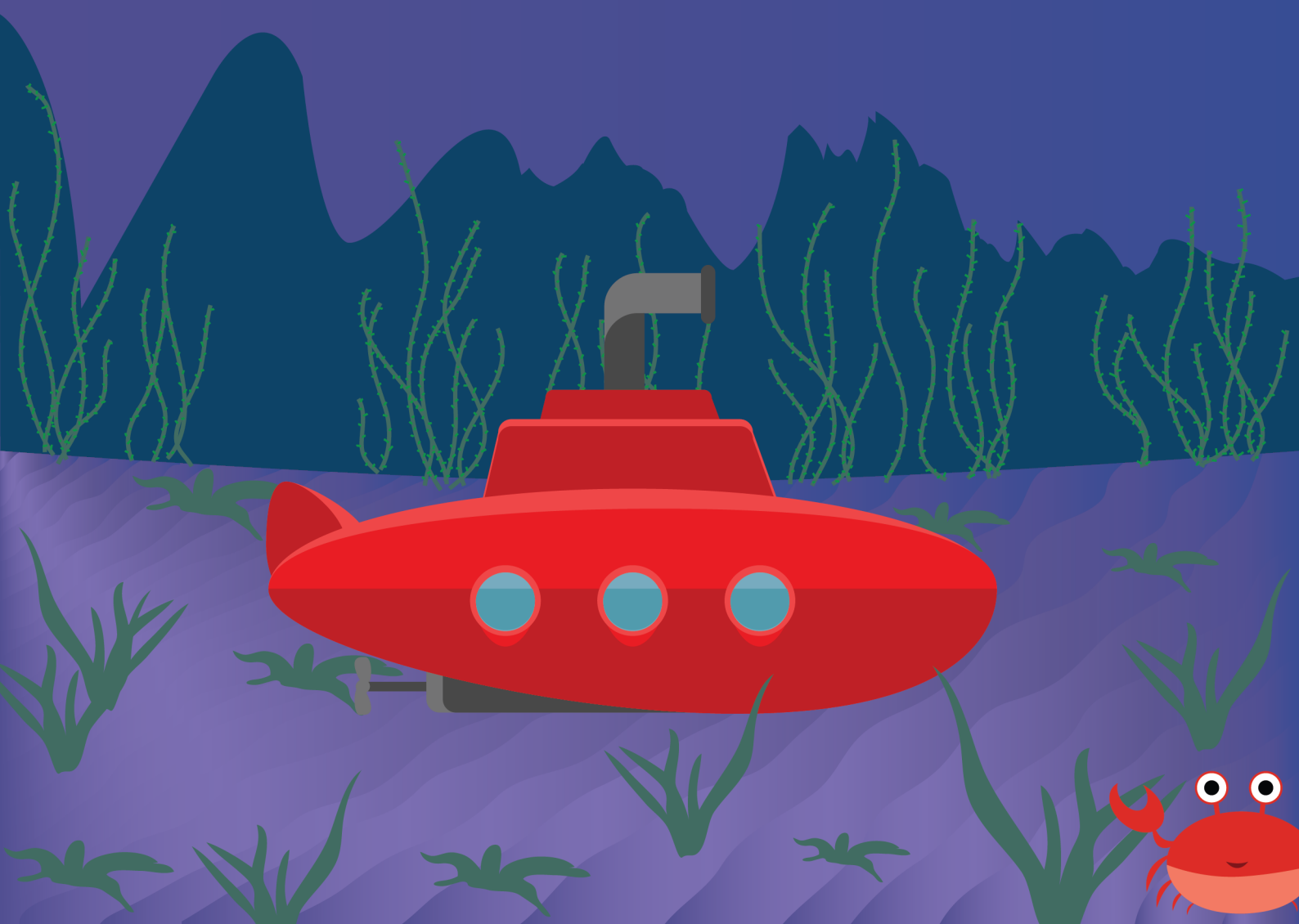
Investigating Game Experience and Engagement in Augmented Reality

P10 - Master Thesis

MED10-2-F19: Cecilie Kamp and Steffen Hedegaard Sørensen

Supervisors: Michael Boelstoft Holte and Anthony Brooks



[This page intentionally left blank]

Abstract

This study explored the game experience and engagement in a collaborataive, co-located augmented reality game, producet for Væbnermesterskabet 2019. The game was developed through an iterative process, where the requirements to the system were investigated and defined through context of use investigations. A prototype was then devoloped through iterations and user evaluations. At the event Væbnermesterskabet 2019, 183 teams tried the game and 642 questionnaire answers were collected. The study showed that gender affects the feeling of competence, and the age affects the reported levels of tiredness and behavioural involvement. The final version of the game was compared to the previous version of the game, which showed that the participants reported less tension/annoyance, challenge, negative experience and psychological involvement - negative feelings.

Resumé

Dette studie undersøgte brugeres spiloplevelse og engagement i et co-lokaliseret samarbejdsspil udviklet til Væbnermesterskabet 2019. Spillet blev udviklet igennem en iterativ process, hvor kravene til systemet blev undersøgt og defineret ved hjælp af brugerundersøgelser. Derefter blev en prototype produceret, som udviklede sig igennem flere iterationer og evalueringer. Til Væbnermesterskabet 2019 prøvede 183 grupper spillet og der blev indsamlet 642 spørgeskemabesvarelser. Spørgeskemaerne viste en signifikant sammenhæng mellem køn og følelse af kompetence, og mellem alder, træthedsfølelse og adfærdsmæssig involvering. Derudover blev den færdige version af spillet sammenlignet med den næstesidste version og der blev fundet signifikant lavere niveauer i deltagernes følelse af; anspændthed/irritation, udfordring, negativ oplevelse og psykologisk involvering - negative følelser.

Acknowledgements

We would not have been able to complete this study without the help and support of the following people:

- Christian Cannon and VMU, for letting us design and implement an exercise for Væbnermesterskabet 2019.
- Efterårshike/JMV-udvalget, for letting us collect questionnaire answers at their event.
- Majbrit Plougmann, FDF Vind and FDF Ørnhøj - for contributing with test participants for several iteration tests.

Table of Contents

1. Introduction

In the last decade smartphones have become ever more popular (Statista, 2019) and increasingly more capable. These devices that many people carry in their pockets have enabled new opportunities for people that cell phones did not. One such opportunity is to use Augmented Reality applications. Augmented Reality uses the camera in the smartphone in combination with the processing power of the smartphone to identify objects in images in real time. This allows the phones to show the images from the camera in combination with digital content.

Smartphone applications have started using the technology for instance to create interactive books where the images can be scanned to see digital content (Books & Magic, 2015). In some of these books the technology have been used to create new forms of games where scanning the images on the pages allow the user to access the games digital content (Jack Hunter Games, 2014). Another game, Pokémon GO, have skipped the scanning of image markers and uses other means to understand the camera images and combine them with the digital content, and became widely popular (Chamary, 2018, Feb. 10).

When considering new technologies it is, however, important to consider what value it adds rather than just being fascinated by the novelty of it. In order to gain an understanding of this it is important to get an understanding of what the game does for its players. Part of this can be to investigate what kind of user experience such a game provides. In relation to this it can be useful to get an understanding of whether the games provide an equivalent or different experience for players of different genders. It can also be interesting to see if the experience is different for players of different age groups.

This project aims to create such an Augmented Reality game and evaluate it with regards to the experience it creates for the players of different demographic groups.

1.1 Problem statement

Exploring the game experience and engagement in a collaborative co-located augmented reality game, designed for Væbnermesterskabet.2019.

2. Theoretical framework

The theory that is relevant to the study, will be defined and presented in the following section.

2.1 Defining collaboration in games

Before creating a collaborative game, it is important first to understand what defines such a game.

According to Zagal, Rick and Hsi (2006) traditional game theory divides games into two categories; competitive and cooperative. In competitive games, the players have opposing goals, which means that they are required to form strategies against the others players. In cooperative games on the other hand, the players are presented with opportunities to work together. The players are not guaranteed to benefit equally from their cooperation though, as their interests are not necessarily the same. Zagal et al. (2006) also state that a third category, collaboration, exists. In collaborative games, the participants are organized in teams. This means that any payoff or outcome, like winning or losing, is shared by the team. The participants on a collaborative team can have differing roles or information, but their interests and main goal is the same (Zagal et al., 2006).

2.2 Engagement

Engagement in video games is a concept that, according to Schoenau-Fog and Bjørner (2012), is related to several other concepts e.g. flow, fun, immersion, and motivation. In this study, engament will be investigated in line with Schoenau-Fog and Bjørner (2012), who state that continuation desire is a fundamental aspect of engagement. They further argue that the desire to continue playing is a prerequisite for other concepts related to engagement, like flow, fun, immersion and enjoyment.

2.3 Social presence

Social presence as defined by Bioacca, Burgoon, Harms and Stoner (2001), is the sense of being togehter with someone else. This applies to both mediated and non-mediated interactions. Though Biocca et al. (2001) is primarily concerned with mediated communication, according to de Kort, IJsselsteijn and Poels (2007), their definition also fits games in a co-located setting.

3. Design framework

The game that was developed as part of the project, was designed using a Human-Centred Design (HCD) apporach which should make the game system more likely to meet user requirements, make it easier til understand and play, and should improve the user experience of its players. Some of the key features of HCD includes understanding the users and context of use, including users in the design and development, and using an iterative process. The iterative process consists of four phases: (a) Context of Use, (b) Product Requirements, (c) Design Alternatives, and (d) Evaluation. The first phase, (a) Context of Use, addresses understanding the users and context of use, for instance by exploring the characteristics of the users and their physical environment, and can include analysis of similar systems to the one being developed. During the next phase, (b) Product Requirements, the user and stakeholder requirements of the product are specified. These should be based on the context of use intended for the product. The (c) Design Alternatives phase revolves around making one or more possible designs which should satisfy the specified requirements. This phase also includes making prototypes to allow the designs to be evaluated. The final phase, (d) Evaluation, focuses on involving users to establish whether the design satisfies the user requirements and how the design can be improved. This activity can also provide a better understanding of the user requirements prompting new requirements to emerge (Insert ISO 2010 reference).

4. Establishing requirements

A mixed method approach was used to gain a better understanding of the users of the product. The activity used a questionnaire method which will be elaborated upon with regards to how the questionnaire was designed, and how the data was gathered and treated. The results, which was used both to inform the development of the product requirements and as inspiration for design alternatives, will also be presented.

4.1 Questionnaire

The following section will describe how the questionnaire was created.

4.1.1 Questionnaire design theory

The questionnaire used a combination of open ended and closed questions and were designed in accordance with theory on questionnaires in general and for children and adolescents in particular.

According to Brace, the encoding of questions are highly important for the success of the communication process (as cited in Lietz, 2008). This means, that it is very important to consider how the questions are formulated. Lietz (2008) presents a range of best practice question design principles regarding the encoding of questions which can improve the response quality in terms of fewer non-response answers, and higher data accuracy caused by fewer misinterpretations of questions by the participants and less likelihood of lies as answers to questions.

The first principle concerns the length of questions which, according to Dillmann, Fink, and Foddy, should be as short as possible and should, according to Brislin, not exceed 16 words or, according to Oppenheim, not exceed 20 words (as cited in Lietz, 2008).

The second principle is about the grammar which Brislin and Dillmann argue should attempt to avoid complexity, i.e. it should use active voice, avoid use of pronouns and possessive forms, to avoid unnecessary use of the respondents' mental capacity, to instead allow them to focus on their response (as cited in Lietz, 2008).

The third principles is to make the questions specific and simple to reduce cognitive load on respondents. This implies, in line with Brislin, and Dillmann, that questions should use specific terms and avoid vagueness in the formulation, and, in line with Jobe and Mingay, complex questions should be broken into simpler ones. Additionally, Belson, and Foddy ad-

vice, that questions concerning future behaviour of the respondents should be asked through vignettes or alternative scenarios (as cited in Lietz, 2008). Another aspect which, according to Oksenberg and Cannell, Rockwood, Tourangeau et al., and Sudman and Bradburn, can increase the cognitive load on respondents is the requirement to recall past events that occurred more than a week ago, and the effect size of this depends on how important the event was (as cited in Lietz, 2008).

The fourth principle is that of social desirability. According to Foddy, the use of advanced vocabulary can cause respondents to feel stupid or uneducated which can make them more likely to give socially desirable responses or "don't know", even though this does not reflect reality (as cited in Lietz, 2008). Another approach to avoid such answers, as suggested by Brace, is to ask a direct question about the respondents ignorance of the given subject which might help the respondent acknowledge this (as cited in Lietz, 2008).

The fifth principle is to avoid "double-barrelled" questions, i.e. questions that contain several different concepts. Such questions has been shown by Brislin, Brace, Fink, Fowler, and van der Zouwen to reduce the accuracy of responses (as cited in Lietz, 2008).

The sixth principle is not to use negatively worded questions which is advised by Belson and Foddy. Wason and Weems et al. have found, that these questions take longer to process and, according to Dudycha and Carpenter, and Eifermann, respondents are more prone to make mistakes when answering such questions (as cited in Lietz, 2008).

The seventh principle is to be cautious with adverbs of frequency which, according to Simpson, can be interpreted differently by respondents (as cited in Lietz, 2008).

The eight principle concerns the question order which Schumann and Presser has found can affect the responses (as cited in Lietz, 2008). They describe three question combinations which can raise this issue; (1) part-whole combinations occur when the questions use a mixture of specific and general terms; (2) part-part combinations occur when questions are equally specific but the respondent alter their second answer to be normative consistent with the former; (3) salience can cause the question order to affect the answer when prior questions have touched the same topic.

Answering a question in a questionnaire requires use of cognitive abilities. The respont has to comprehend the question, retrieve information from their memory, judge which information is needed for the answer, and communicate the response (Schwarz & Sudman, 1996 as cited in Bell, 2007; Tourangea, 1984 as cited in Bell, 2007). If any of these aspects of the

answering process has not been performed, the quality of the answer may suffer (Krosnick, 1991 as cited in Bell, 2007; Vaillancourt, 1973 as cited in Bell, 2007). This can occur if the question formulation is too complex, either with regards to the vocabulary or syntax, if the respont does not possess the required cognitive capacities, or if the respont finds the interview boring or the topics uninteresting (Bell, 2007). For this reason, it is necessary to be aware of both how the questions are formulated and the cognitive development of the respondents.

The data gathering for the product requirements included participants aged 11 and above so the questionnaire needed to be suitable for that age group. According to Piaget (1929) as cited in Bell (2007), the cognitive abilities of children and adolescents are not yet fully developed. This meant, that the participants might not have had the abilities required to answer certain questions. However, according to Borgers et al. (2000) as cited in Bell (2007) and Scott (1997) as cited in Bell (2007), children aged 7 and above should be able to answer questionnaires as long as the questions are carefully adapted to their abilities, and once the children reach 11 years of age the level of adaptation required is much reduced.

A number of question design principles for questions aimed at children and adolescents has been suggested by Bell (2007). These include to; (1) be careful with the question length and wording; (2) avoid complex formulations; (3) be aware that children can have trouble recalling, and thus reporting, previous behaviour; and (4) avoid suggestion and connotation. When compared to the question design principles suggested by Lietz (2008), several similarities exists, e.g. both include question length, avoiding complex formulations, and troubles with recalling previous events. This is not surprising as Bell (2007) states, that the design principles are also applicable to adults, but are of particular importance for children and adolescents. These principles were emphasised during the development of the questionnaire.

4.1.2 Development of the questionnaire

The development process started with a specification of the overall themes the questionnaire should address. The themes revolved around the participants' previous experiences and preferences with activities similar to the one envisioned for the project. e.g. previous participation at Væbnermesterskabet, FDF activities, and game experiences. The questionnaire should also provide some relevant demographic information about the respondent.

A number of questions were formulated to address each of these themes. The primary focus of the question formulations were to ensure they probed for information for the relevant

theme. Two additional questions were also included; one which allowed the participants to add any additional comments and another which asked the group leaders if they were interested in further involvement in the project. See appendix XX for the initial set of questions sorted by themes.

The questions were refined by considering whether or not they abided by the design principles outlined by Lietz (2008) and Bell (2007), with particular attention to the principles related to children and adolescents. This resulted in some changes, e.g. the wording of one question contained Prioriteret rækkefølge [Prioritised ranking] which was exchanged with an example to avoid complexity in the formulation, and two questions were asking hvilke poster kan du godt lide… [which exercises do you like] and hvilke aktiviteter kan du bedst lide... [which activities do you like best] which were made more specific by asking for a certain amount rather than leaving it unspecified. See appendix XX for the resulting questionnaire which was presented to the participants.

4.2 Data gathering

The following section will describe how the data was collected.

4.2.1 Setting and participants

The data gathering was carried out at an FDF event called JMV/Efterårshike. The event was for participants aged 11 years old and up. The event took place in Aulum and Vildbjerg over the course of a weekend. Most participants were members of FDF, but the event was also open for scouts to participate.

4.2.2 Procedure

The questionnaires were handed out to the participants on Saturday evening along with the ingredients for their dinner. The participants were asked to fill out the questionnaires and to hand over the questionnaires to the arrangers when they were done. The arrangers reminded the participants of this later the same evening. Thirty-three questionnaires, which had been filled out to varying degrees, had been handed in by the end of the event on Sunday.

4.3 Data treatment

The data were treated differently depending on the type, i.e. quantitative and qualitative.

4.3.1 Quantitative data treatment

The quantitative data were gathered in the first questionnaire item where participants rated which activities they prefered from a set of candidates. The answers from participants 1, 2, 3, 12, 20, 23, 24, 25, 26, and 27 were disregarded as they had not provided answers in line with the question formulation, i.e. they had only rated a few of the candidates, they had not used numbers to rate, or they had used wrong numbers. Participants 19 and 32 had only rated 10 of the 11 candidates but this was considered to be a mistake and their answers for the 10 candidates were still considered valid and used. The responses from the 23 remaining participants were compiled in a spreadsheet and used for the analysis and are presented in appendix ## which show the preference rating values of the participants.

4.3.2 Qualitative data treatment

The results for the qualitative data were gathered in the second, fourth, sixth, seventh, and eight questions which regarded the activities and games the participants had liked doing and playing in the past in various contexts. The questions both concerned the activities or games and reasons for why they were favored. These two aspects were separated during the data treatment and data were condensed by identifying keywords in the responses, e.g. one respondent answered Løs gåden, fordi det var udfordrende [Solve the riddle, because it was challenging] which was separated and condensed into the activity løs gåden [solve the riddle] and reason udfordring [challenge]. The activities and games were enumerated and sorted to give an overview of the data, see appendices ##a and ##b for the resulting data which was used for the analysis.

The reasons the participants had used to describe why they favored certain activities and games were further condensed into standardised categories, e.g. challenge and teamwork. The categories were found by considering each reason specified and adding a new category if a similar category did not already exist or otherwise add one to the count of the similar category. Each participant was only counted once for each category since one participant could use the same reason for several activities. This resulted in a count of how many participants had used a particular reason for why the liked an activity or game, see appendix ## for the list of standardised categories and counts.

The games which the participants liked playing were also treated in regards to their genres. First the particular genres of the games were determined using an online video game

database (https://videogamegeek.com/). Then the number of participants who had reported a given genre was counted to get an overview of which genres were most favored.

4.4 Results

The aim of the data gathering was both to develop a set of user requirements and to get inspiration for the design alternatives phase of the project. The results for both of these aspects will be presented separately.

4.4.1 User requirements

The user requirements were described as user requirement cards including a title, description, rationale, originator, and priority which was an approach inspired by the Vollere shell method, described by Sharp, Rogers and Preece (2015). The priorities values ranged from five, a must be satisfied requirement, to one, an optional requirement. Priority ratings in between could be excluded but with a reason to do so, e.g. trade offs with other requirements, and a four would require a better reason than a three which would require a better reason than a two.

The user requirements were developed from a combination of the standardised categories extracted from the qualitative data and the preference rating values extracted from the quantitative data. The count of the standardised categories showed a large variety of aspects the participants liked. However, many of the categories had low amounts of participants favoring them, i.e. one (3%), two (6%), or there (9%). These categories were not included in the user requirements. Categories with 4 (12%) or above participants favoring them were included. This resulted in the six user requirements fun (42%), thinking (24%), teamwork (21%), challenge (18%), learning (15%), and novelty (12%). The preference ratings supported that teamwork and thinking should be included in the user requirements, see figure 1 for the activity preference ratings.

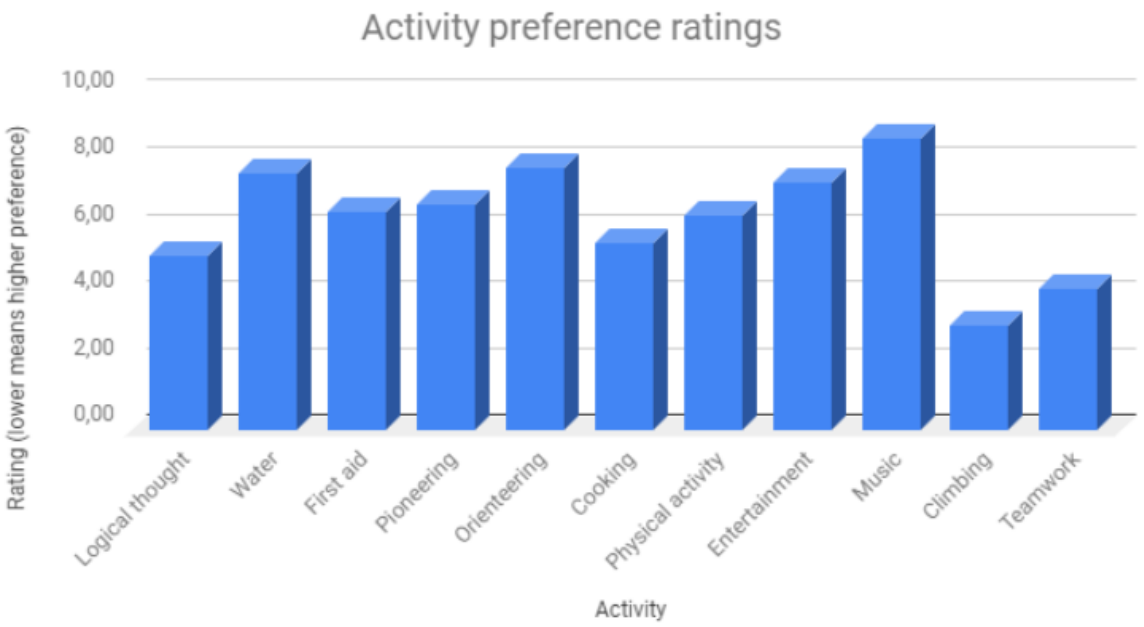


Figure 1. Activity preference ratings

It showed that participants on average placed teamwork high (4,20) and logical thought were in the midrange (5,16). Note that scores are in the range one, most prefered, to eleven, least prefered. Each user requirement were given a description and based on the scores they received a priority rating, see appendix ## for the resulting user requirement cards.

4.4.2 Design inspiration

The data regarding the genres of the video games played by the participants showed that they favored a wide variety of game genres. This meant, that it did not provide the intended inspiration for the design as there were no clear trends in the choice of games they played.

5. Design and implementation

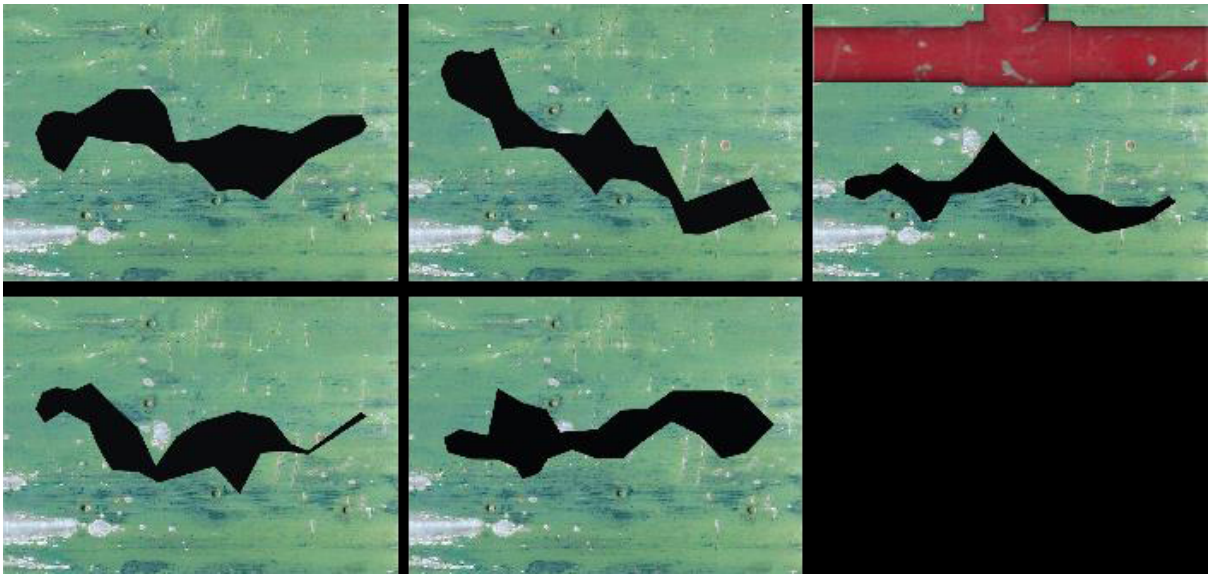
In the following section the game will be presented and then the process of designing and implementing the game will be described.

5.1 Game description

The premise of the game was that the players were trapped in a submarine that was malfunctioning under the sea. The submarine was running on a limited battery and the players needed to fix the power generator and call for help over the radio before time would run out. In order to achieve the goal, the players were required to complete a series of tasks. Failing to do so would mean a lower point score, ranging between 0 and 100, would be awarded.

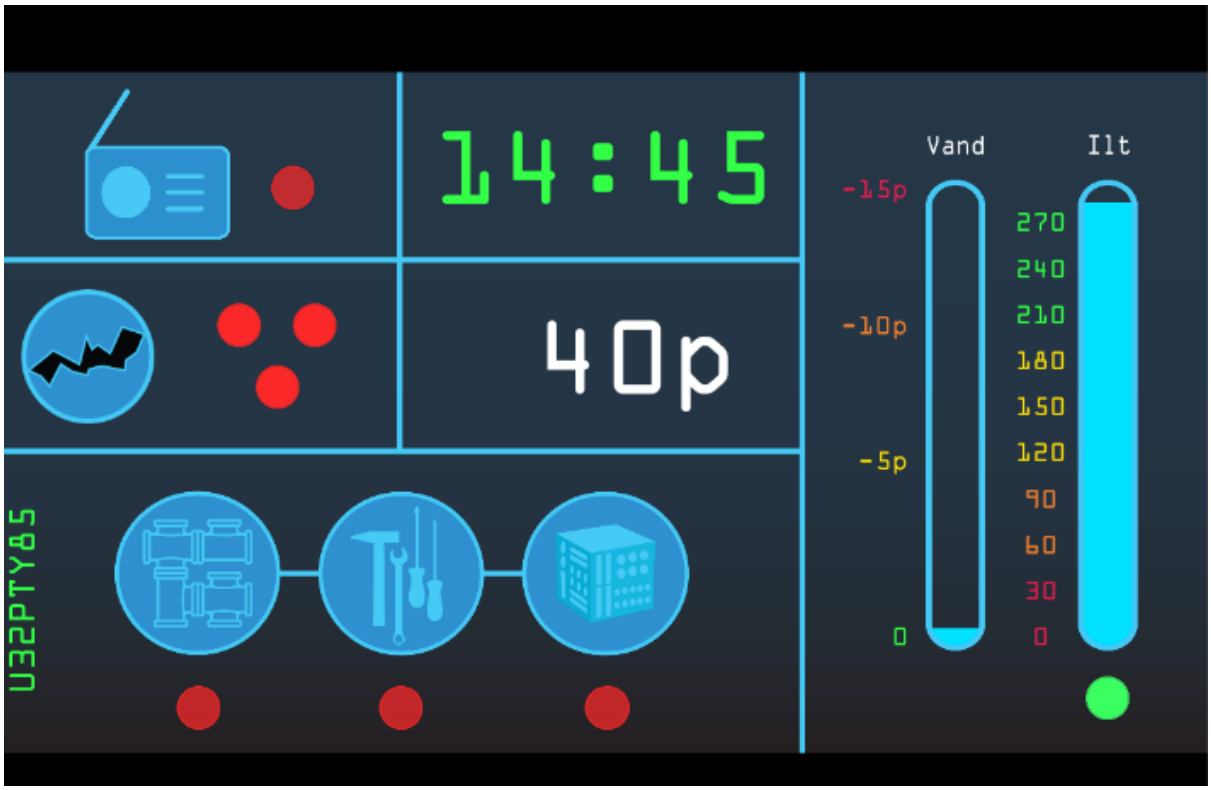
Upon starting the game the four players divided the four player colors between them. The chosen color would decide which tasks that player would need to do and which objects in the game world they could interact with.

One of the first challenges the players would be met by was a water leak in the hull which they would be warned about by a warning siren and blinking icon on the main console. The challenge would first require the players to search for, and find the image target which had the leakage, see figure 1 for the water leak image targets.



When the correct image target was found they could see the two colors of the players that would be required to participate to complete the challenge. The two players would then have to drag a gaffer tape object in the scene along an outline of a gaffer tape strip to apply the tape to stop the leakage. When both players had done so the task was solved, see figure 2 for the water leakage mini game containing gaffer tape objects and strip outlines.

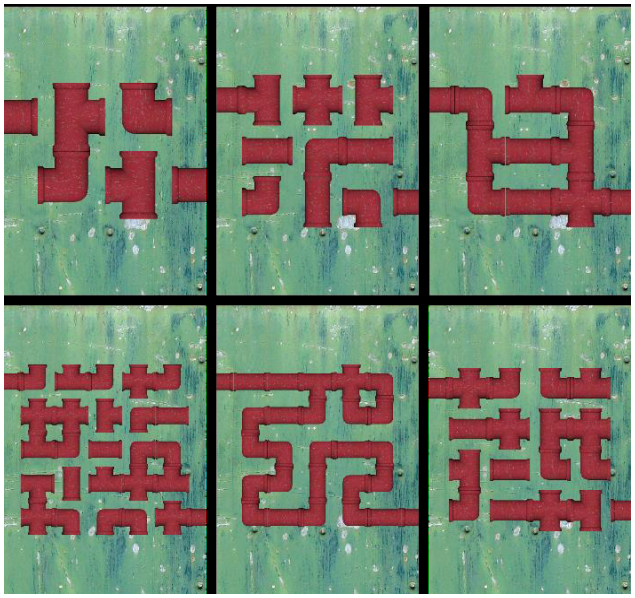
When the leakage was solved the red blinking icon on the console would stop blinking and turn green. See figure 3 for the main console, in this case with all three leakages active simultaneously.



Besides solving the leakages the players were required to try to repair the emergency

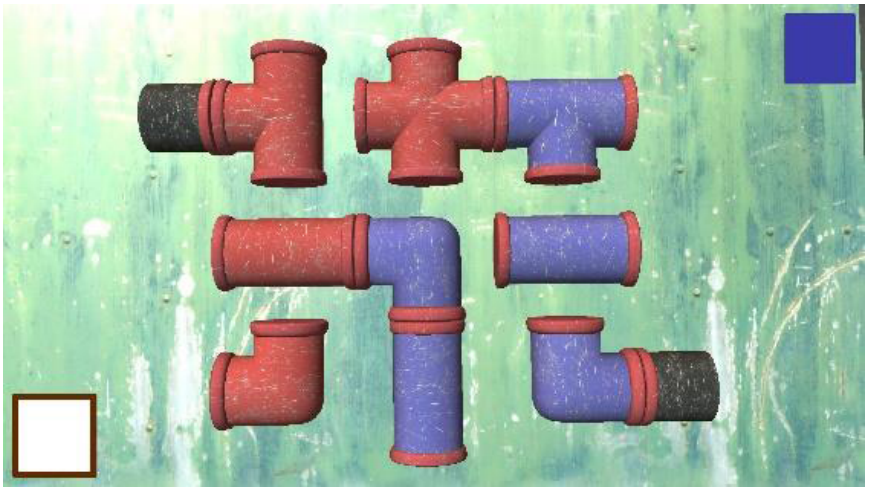
generator which was a cube in the center of the room with AR markers on all sides, see figure 4 for the generator.

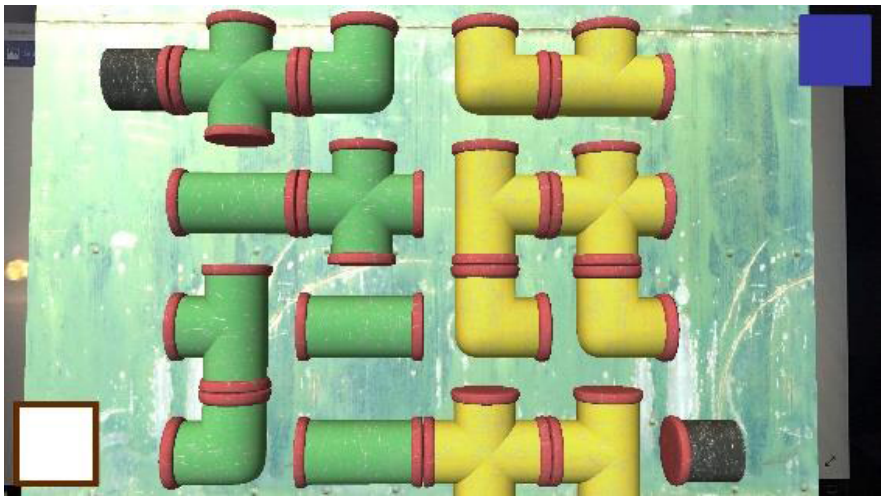
To do this the players needed to complete three tasks and their status was shown on the bottom left of the main console. The first of the tasks was to fix the broken fuel pipes in the room. This required finding the broken fuel pipe image targets, see figure 5 for fuel pipe image targets including some that are not broken which was meant to make the search more difficult.

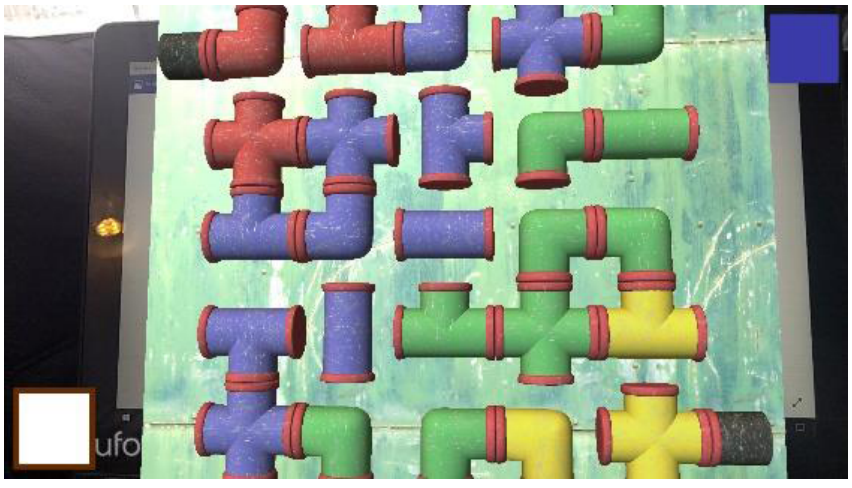


When the broken fuel pipes were located the players needed to solve a puzzle. They needed to connect the top left part to the bottom right by rotating the pipes by tapping on the to rotate them 90 degrees, see figures 6, 7, 8 for the three difficulties of fuel pipes requiring

connection of 3x3, 4x4 and 5x5 pipes.







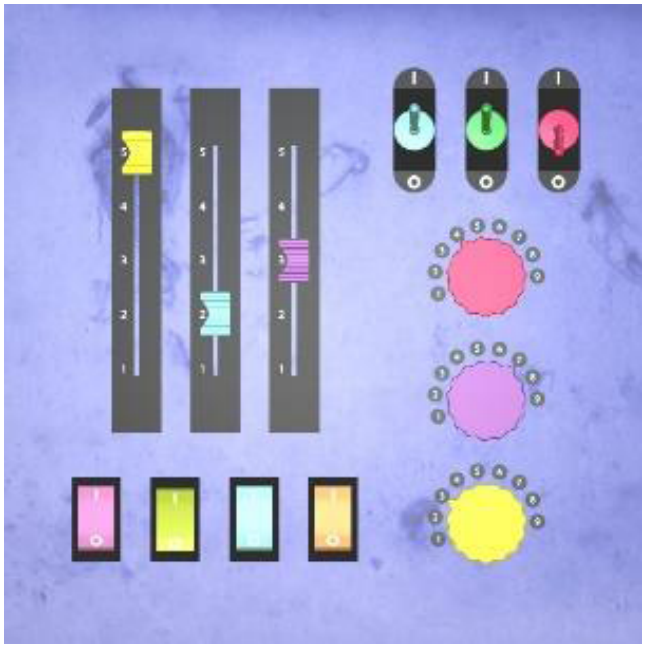
Besides the fuel pipes, the players also needed to find 4 pieces of tools that were

required for the generator repairs. They were located in cabinet image targets on the walls, see figure 9 for the cabinet image targets.

When one of the cabinet image targets was scanned the player could open the cabinet by tapping on the door, see figure 10 for the cabinet door. This revealed its contents.

By searching the cabinets they were required to find the four tools that were needed. The specific tools they needed could be found in the manual which they were encouraged to use while playing the game. The cabinets also contained a range of tools that were not needed to make the search more challenging. See figure 11 for an open cabinet and the tools it contained.

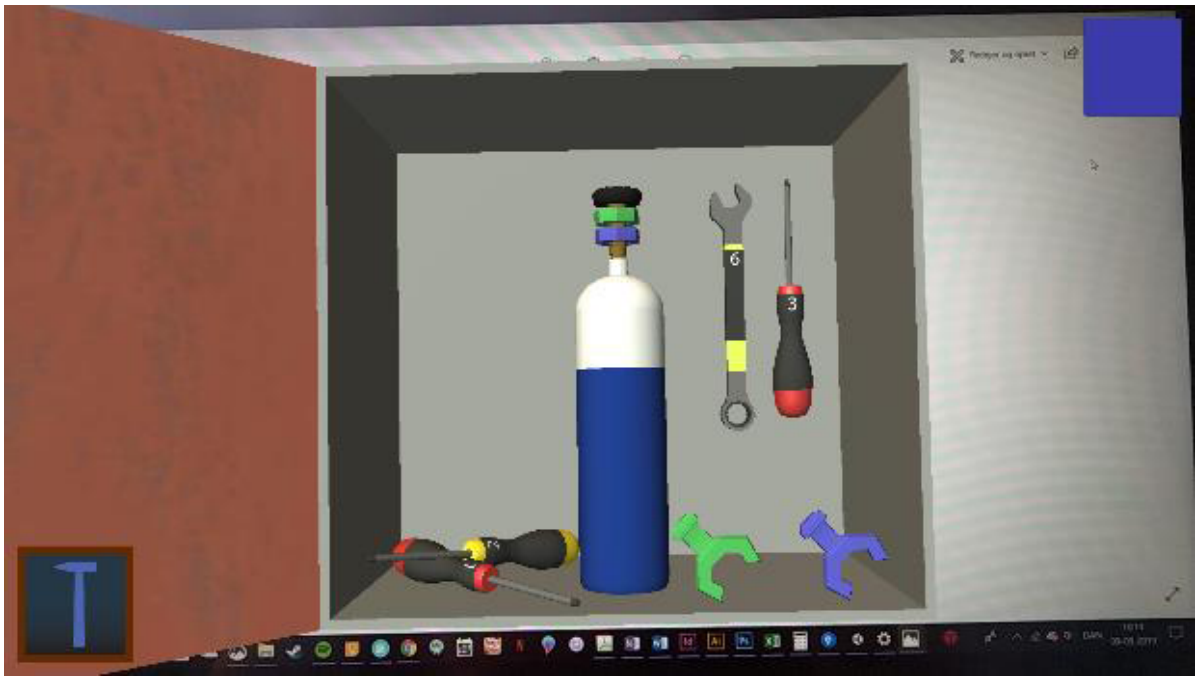
Once the tools had been found and fuel pipes fixed the players would need to scan the generator image targets and drag the tool onto the generator simultaneously to repair it. This would make it possible to interact with the generator which was indicated by a change of color. Each side had a different color corresponding to one of the player colors, see figure 11 for the blue players side of the generator.



The final step of the task was to set all of the flip switches, toggles, sliders, and dials, to their correct positions. These positions could be found in the manual.

The game contained one more task the players needed to attend to. They only had a limited amount of oxygen which they needed to keep track of on the main console. When

they were running low they would need to find oxygen bottles, which were located in some of the cabinets, and open them to replenish their oxygen supply, see figure 12 for a cabinet with an oxygen bottle.



At the start of the game the players would have 40 points. If they failed to stop the water leakages they would get a point penalty of maximum 15 points. Similarly, if they failed to replenish the oxygen supply the would get a point penalty of maximum 15 points as well. The other tasks of the game would award points, upon completion and if everything was solved, and no penalties were awarded they would get a maximum of 100 points.

5.2 First iteration

Going into the next phase of the project a basic idea for the game had already been formed. The game was going to be inspired by the escape room concept and the goal of the game would be for the players to survive in a submarine stuck at the bottom of the sea, by solving a number of different minigames. The following section will describe the process of developing the minigames.

5.2.1 Brainstorm

As stated by Fullerton (2014), ideas do not appear out of thin air. Therefore, a brainstorm was carried out in order to help the idea generation process along (p. 148). This was

done in line with Fullerton (2014) who describes brainstorming as a formalized system for idea generation (p. 150). As stated earlier the goal of the brainstorm was to come up with ideas for the minigames that ultimately was going to make up the gameplay of the game. Fullerton (2014) lists several different creativity methods that can be used for brainstorming (p. 153). For this session the method Idea Cards was used in combination with the method Cut It Up. Magazines written for the target group, and published by FDF, was used to find words for the Idea Cards, which were then randomly matched in groups 14 of four. Each group was then used to create an idea for a minigame. This resulted in 14 short descriptions of minigames that was used as inspiration for the final seven mini games. Next a concept document was written, putting the idea for the main game into words. A mind map with the formal elements of the main game was also created. The formal elements of the main game are shown in figure X. Afterwards concept documents and mind maps were created for each of the minigames.

5.2.2 Paper prototype

The next step of the iterative process of creating a game, as presented by Fullerton (2014), is to create a playable physical prototype (p. 15). The goal of this is to quickly implement a playable version of the game that can be used for testing the most fundamental game mechanics (p. 17 + 175).

Based on the above recommendations from Fullerton (2014), a paper prototype was created. The prototype consisted of a printed manual for the participants to look up information in, printed out markers for the participants to “scan”, cardboard mobile phones and printed screen images that was to be handed to the participants when they “scanned” a marker with the card board phones.

5.2.3 First iteration evaluation

Before proceeding to the second iteration, the paper prototype was evaluated. This evaluation will be described in the next section.

5.2.4 Setting, setup and participants

The paper prototype was evaluated with eight members of the target group at an FDF event.

The paper prototype was set up in an old classroom and the lights in the room were switched off during the test. The game was tested by two groups of four that represented the two age groups that participate at Væbnermesterskabet.

One moderator and an assistant were present during the evaluation. The moderator presented the game to the participants and both the moderator and the assistant were active during gameplay.

5.2.5 Evaluation

The goal of the evaluation was to do a proof of concept and find out whether the participants found the user experience fun and enjoyable, whether they were able to understand the concept and play the game. The evaluation would also look for usability deficiencies and test whether the manual worked as intended.

When a group entered the room they were given a short introduction to the game and were then given fifteen minutes to play. When participants “scanned” the markers with the cardboard phones, the moderator and an assistant would then give them the corresponding screen image. The participants used flashlights in order to be able to see in the dark room.

When the participants had finished the game, they answered a short questionnaire about the experience (Appendix X).

5.2.6 Results

The results showed that the participants thought that the game was fun to play, that the game had an appropriate level of challenge and that the game required logical thinking to some degree. The game only required a small amount of teamwork. The participants reported that the game did not remind them of other games, and that they found it hard to figure out how to play the game. Two participants stated that reading the manual was boring. One participant stated that the game was boring if not everyone had something to do.

It was observed by the moderator that the participants were confused by the serial numbers and that they did not notice that they had to repair leaks or find oxygen.

5.2.7 Suggested solutions

One way to introduce more collaboration to the game could be to introduce a role system, where each person has a specific set of abilities.

5.3 Second iteration

The implementation of the second iteration prototype will be described followed by the evaluation of the prototype.

5.3.1 Second iteration prototype

The goal of the second iteration prototype was to implement one of the mini-games to gain insight into how usable this part of the system, and subsequent similar parts of the system, would be. The game design was also refined to better support collaboration by the players. Zagal, Rick and Hsi (2006) outlined four lessons for collaborative games; (1) a collaborative game should introduce a tension between perceived individual utility and team utility; (2) individual players should be allowed to make decisions and take actions without the consent of the team; (3) players must be able to trace payoffs back to their decisions; (4) a collaborative game should bestow different abilities or responsibilities upon the players. The four lessons were met by allowing individual players themselves to decide what they wanted to do during the game (2), by providing feedback when the players completed a task which was needed to achieve points (3), by only awarding maximum points when all players participated as a team (1), and by introducing four roles each of which had to participate to solve certain tasks (4).

The prototype developed for the iteration was a software based implementation of the mini-game water leak envisioned for the final product. The game required the players to locate and scan augmented reality markers to find water leaks in the submarine. Upon scanning the markers, the superimposed digital content showed a wall of the submarine and possibly a water leak and one player object for each involved player. The goal of the game was for each player to tap on their player object and drag it to certain locations on the screen indicated by black rectangular objects depicting pieces of duct tape. When all players had completed their part of the task the water leakage would disappear. When the players moved their objects across the screen, the locations of the objects were synchronised across the network so each player could see the other players actions. Similarly, the players would all see the water leak disappear once the task was completed on all devices.

The prototype required the implementation of augmented reality, networking to allow several players to play the game together, roles to make the game a collaborative experience,

interaction by the players and tasks that needed to be solved. These aspects of the system were implemented as individual components that were intended to work independently of each other as much as possible. Each components were controlled by a managing script. The system also contained a user interface providing necessary tools for interaction and feedback, overlayed on the augmented reality content. The game was implemented using Version 2018.2.18f1 of Unity3D (2018) which provided tools for developing android games, augmented reality applications, and network based games.

5.3.1.1 Network implementation

When the game was first started the user was presented with a connection interface. This contained buttons to allow the user to set up a server or connect to a server hosted on a specified internet protocol address. Clicking on one of the buttons would call a function in the network manager script to perform the operation. The network manager was a specialisation of the unity network manager which provided much of the needed functionality, i.e. setting up a server or connecting to one, listening for networking events such as new connections being established or disconnections, and synchronisation of game state.

5.3.1.2 Role selection

The connection interface was substituted by a role selection interface if the connection was successfully established. In order to detect that this had occurred, the function being called on a new connection being established was overridden to wait for the connection to be ready, indicating that the connection attempt was successful. While the system was waiting for this to occur, a message was shown to provide feedback to the user of the waiting state.

When the user had selected a role, a request was sent to the server to assign that role to the client. The request was sent using a Command, which was a remote action enabled by Unity's network API allowing a client to invoke a function call on the server. The server would assign the role after having ensured that another client had not already acquired that role. The authority of a player role object was assigned to the client if the role request was granted. This would also cause a message to be sent from the server to all connected clients using a ClientRPC, i.e. another remote action that allowed the server to invoke a function call on all connected clients. The message informed that the role had been acquired and which connected client had acquired it. This allowed the client that had been granted the role to pro-

ceed to play the game and removed the option to select that role on the other clients.

5.3.1.3 Water leak task

The prototype only had a single task which was assigned by a task manager component, which was intended to keep track of tasks and their completion, when the server was started. This created an instance of the water leak prefab for two players and with a fixed number of targets for each player. The instantiated gameobjects were also spawned on the network using Unity's networking API, causing them to be instantiated on all clients. The prototype contained a single augmented reality marker, or an image target as they are called in Unity, to be scanned, which revealed the water leak and associated mini-game. The spawned objects' parents were set on both the server and clients to be the associated augmented reality image target which was present in the game scene to make it work with the augmented reality system. Additionally, the color and name of the objects were set. All of these parameters were stored in SyncVars, a type of variable in Unity's networking API that was automatically synchronised between server and clients on the network. This allowed the parents, colors, and names to be set on clients that would connect later, or had to reconnect, so game sessions would not have to be restarted in case of devices losing connection to the server.

5.3.1.4 Interation with the game

Users interacted with the game using the touchscreen of their device. If the touch was a new touch and they had selected a role this caused a ray to be cast perpendicular to the camera plane from the fingers location. If an object was intersected by the ray, the object allowed interaction by the player's role by having the corresponding tag applied, and was interactable, i.e. had a component which implemented the IInteractable interface, the select function in the interactable component was called. This caused a request to be sent to the server as a network command, to give the client network authority of the object, allowing the player to affect its state. The server would grant authority if no other player currently had authority over the object and sent a message through a ClientRPC call to change the color of the object as feedback of the authority change. The authority would be removed and the color would revert to its default color when the player ended their touch. If the player made a sustained touch the object position was changed allowing the player to move it. The new position was found by raycasting against the object depicting the wall of the submarine by placing it on a layer

and using a layermask for the raycast. Finally, the object position was synchronised using a SyncVar to apply the movement to all network entities. The objects the player could interact with were the tape objects, i.e. game objects shaped as a cube with the players color.

5.3.1.5 The player goal in the game prototype

The goal of the game was for each player to move their tape object to a number of locations in a specific order. The locations were created as target game objects showing a red sphere at the appropriate positions. The targets were stored in an array and the index of the next target the player needed to move towards was stored in a integer variable. When the player moved the tape object, the distance to the next target was calculated and if the distance dropped below a certain threshold, a new target would be set. To provide feedback to the player, a line gameobject object, i.e. an elongated cube with a black color, was placed between the two targets, rotated and scaled to make a line between the targets. This was meant to show that the player had applied the appropriate piece of tape. If the final target were reached a message was sent to the server using a network command to inform that the player had completed his part of the task. When all players had completed their parts, the server would remove the tasks by destroying the game object.

5.3.2 Second iteration evaluation

The second iteration prototype was evaluated before proceeding to the third iteration. This will be described in the following section.

5.3.2.1 Setting, setup and participants

The evaluation was carried out at one of the weekly FDF meetings of the FDF group that had also tested the paper prototype. Ten participants tried the game in pairs of two.

5.3.2.2 Evaluation

When the participants entered the room, they were given a short introduction to the game, where they were told that they had to close the whole in the wall in order to make the water stop. The participants were then handed an Android device each, to play the game on. The participants then started playing the game. The moderator were present during the evaluation and would help the participants if they did not understand how to solve the task.

5.3.2.3 Results

The participants were having a hard time understanding how to solve the task. Initially they were having trouble understanding that the icons they were moving was gaffer tape. This should be solved by making icons that look like gaffer tape.

The traceable outline of the gaffer tape should always be visible. The colour of the outlines should be changed to match the colour of the participant that is able to solve that part of the task. The participants should also be able to the outline that the other participant is supposed to trace with the tape in order to help them understand that they each have a task to do.

5.4 Third iteration

The changes made to the game prototype, new features implemented for the third iteration prototype, and evaluation of the prototype will be described.

5.4.1 Third iteration prototype

The third iteration prototype contained changes to deal with the issues encountered in the second iteration and two new components, the main console and oxygen task.

5.4.1.1 Changes based on results of second iteration

Certain changes were made based on the evaluation of the 2nd iteration prototype. The participants had found it hard to understand the meaning of the gaffa tape icon which was still only a placeholder piece. A 3D model of a gaffa tape roll was created in Autodesk Maya to replace the placeholder object.

Another issue encountered in the evaluation was that the participants did not understand the meaning of the sphere used to represent the starting point for applying the tape to solve the task. The sphere and ordering of the target locations were removed. In their place, the tape outline showing where the players had to apply tape was made permanently visible as semi-transparent rectangular cubes which was colored to match the color of the player who needed to apply that piece of tape. These outlines and the tape progression feedback objects were also made to work with the network system so they were shown synchronously on all devices.

5.4.1.2 Main console

The first of the new features implemented in this iteration was the main console component. It showed feedback on how the players was doing in the game and what tasks they currently needed to attend to. It was implemented using the User Interface component in Unity and the graphics were created in Adobe Illustrator. The main console contained square sections to provide feedback on certain aspects of the game. Some of these aspects related to features not yet implemented so they were left blank. See figure XX for the visuals of the main console.

An icon depicting a tear in the virtual submarine hull was placed on the left side of the main console accompanied by five circles. One of the circles would change color from green to red when a water leak occurred. This informed the players that they had a task to search for and solve. The water leaks were initiated by the task manager at certain time intervals using delayed function calls using the invoke function in Unity. This created the water leak and incremented a water leak count property, which described the current number of water leakages, and was passed as a function parameter to the main console script to update the water leakage display. The main console script had references to the five circle objects. When the update function was called, it used a loop, running from zero to the number of water leaks, to color the appropriate number of circles red and another loop running from the number of water leaks to the number of circles to color the remained of the circle objects green.

On the right hand side of the main console was presented two bars, one of which was labeled Vand [Water]. The water bar showed how much water had entered the virtual submarine which affected the participants score. The amount of water increased by a certain amount per second and per water leakage so by solving the water leak tasks fast the players would achieve a higher score. Next to the water bar was placed a bar labeled Ilt [Oxygen]. The oxygen bar showed the oxygen level in the virtual submarine. The oxygen level would deplete over time and a score penalty was applied when it reached zero while the level was reset to the maximum value. The oxygen level could also be reset to the maximum value by completing one of the oxygen task which was implemented for the iteration. This required participants to pay attention to the oxygen level and attend to it in time to avoid losing points. Both bars were implemented using colored image components. On startup the main console script would store the height of the image components as their maximum height, and in the case of the water bar set its current height to zero as it would fill up rather than deplete. When

the water or oxygen level changed, the updated level was sent to the main console script as function parameters. The script would then change the height of the image component to a percentage of the maximum height based on the updated value.

5.4.1.3 Oxygen task

The oxygen task was added in the third iteration of the product. This task required participants to search for oxygen bottles in the virtual submarine and complete a small task to refill their oxygen supply. The oxygen bottles were located in augmented reality image targets depicting cabinets located in the physical space. There were ten cabinet AR markers of which five contained oxygen bottles. This meant, that the game needed to differentiate between the different AR markers as the contents differed. Each cabinet AR marker included a numeric value to allow this differentiation.

The virtual cabinets were shown when the participants scanned the markers and included a cabinet door which opened by gradually rotating when tapped upon. The contents of the scanned cabinet was revealed once the door had opened. The cabinet contents were intended to be placed behind the AR marker to look as if the cabinet was recessed into the wall. This was not the default behaviour of the Augmented Reality API in Unity where contents were placed above the AR marker. However, a solution was found by using depth mask objects which was a tool included with the Unity Augmented Reality API to allow objects to be placed behind an image target which resulted in the desired behaviour.

The players needed to activate the oxygen bottles once they had been found to replenish their oxygen supply. In order to do so, they needed to first move an icon showing a wrench tool to the opening of the bottle. Only the player with the correct role could interact with the icon and the required role was shown by the color, i.e. a red colored tool required the red role player to interact. This was done to enforce collaboration, in line with Zagal et al. (2006) that states, that giving players different roles is a mean to enforce collaboration. The tool icons positions were synchronized across the network so the players could see each others actions and by communicating, they could assist each other. Once the player had moved their icon to the bottle a wrench tool object was instantiated and spawned on the network. When all players required for the task had done so, they could start to activate the bottle. This was done by tapping on the wrench tool and moving the finger to rotate the tool. The rotation of the tool was also synchronised on the network so the players again could help each other. The

synchronisation was also used to enforce a constraint to challenge the players by limiting the rotation to a certain degree of difference between the players. This meant all included players needed to rotate their tool simultaneously to complete the task.

Once a certain degree of rotation was achieved the task was completed which caused the oxygen supply to replenish. A text message appeared to inform the players of this to provide feedback on their success and the oxygen bottle object was from the game using the Destroy function in Unity.

5.4.2 Third iteration evaluation

The prototype that was developed during the third iteration was evaluated with members of the target group. The evaluation will be described in the following section.

5.4.2.1 Setting, setup and participants

The evaluation was carried out at an FDF meeting, where members of the target group were present. Three groups of participants tested the prototype. The first group had four participants from the ‘Væbner’-category and the two other groups consisted of two participants each, who were all from the ‘senior’-category. During the evaluation both researchers were present in the room.

Insert diagram showing the setup

5.4.2.2 Evaluation

The goal was to investigate whether the challenges that were seen in iteration two evaluation had been solved, and whether the new additions that had been made to the prototype worked as intended. When a group entered the room they were handed two mobile devices and two flashlights. They were then introduced to the game and afterwards started playing it. During the first group had finished the test it was clear that the game needed balancing with regards to when new obstacles occur. Therefore the three groups ended up playing the game at different difficulties.

5.4.2.3 Results

It is hard to hold both a phone and a flashlight at the same time. Therefore the participants who hold the mobile phones should preferably have headlamps.

If the light from the flashlights are too bright, it as hard for the mobile devices to scan the image targets. The participants should either be instructed in how to light up the image target or the flashlights should not shine too brightly.

The frame in which the gaffer tape is placed should be made bigger.

A red light underneath should be placed underneath the oxygen ‘måler’ and warn the participants the oxygen level is getting critical.

There should be different levels of oxygen in the oxygen tanks that are hidden in the cabinets.

The numbers by the oxygen ‘måler’ should indicate how much oxygen there is left in seconds instead of percent.

The water level indicator should be able to show how many points the participants have lost and when they lose them.

5.5 Fourth iteration

This section describes the changes made to the fourth iteration based on the third iteration evaluation, new features added to the prototype, and evaluation of the fourth iteration prototype.

5.5.1 Fourth iteration prototype

Several issues identified in the third iteration evaluation were dealt with for the fourth iteration prototype and the new tasks, Fuel pipe, Generator repair, and Generator startup was added.

5.5.1.1 Changes based on results of third iteration

The gaffer tape movement in the game was made by making a raycast from the finger position on the camera plane, and detecting the intersection with an invisible plane in the scene defining the gaffer tapes range of possible positions. This plane was scaled up to allow the players a larger area to move the gaffer tape inside so they would be less likely to try to move the gaffer tape to an illegal position which would stop the movement.

A warning was added to the main console to inform the players of when they needed to look for more oxygen. The warning was a circle that would change color when the level dropped below a certain threshold. The values on the oxygen bar was also changed to reflect

how much time the players had left before oxygen would run out, rather than being a percentage level of oxygen left. The water bar on the main console also had its values changed to reflect the point loss which would occur at specific water levels rather than just showing a percentage of water intake.

5.5.1.2 Fuel pipe task

One of the new tasks added to the system were the Fuel pipe task. When the participants scanned one of the Fuel pipe AR markers, a puzzle game would be presented. The game required the players to rotate a number of quadratic tiles with fuel pipe segments on them to make a connection between the two corners of the area. The different markers would contain tiles in one of the arrangements 2x2, 3x3, 4x4, 5x5, or 6x6. Some of the markers would contain a puzzle that was already solved to make it more difficult for the players to find the right markers. The solved puzzles could also server as an example for the players of how the puzzles should be solved.

The rotation of the tiles were synchronized across the network. When a player interacted with one of the tiles of that players color, it would begin its rotation, if it was not already rotating, and would continue until it had rotated 90 degrees. A command would be used to inform the server that the player had interacted and the server would also perform the rotation. The server would also tell them other connected clients of the interaction and they would make the rotation. The server would also keep a rotation state SyncVar updated to allow new connections or players that needed to reconnect to have their game state synchronized with the server.

When one of the tiles finished its rotation on the server, a path finding algorithm using breadth-first-search would be used to check if a solution to the puzzle had been found. The path finding worked by building a node tree with its root at one of the corners. The root would be expanded by looking at which type of pipe was on the tile and how it was rotated. The nodes that were found to be connected to the root would be checked for if they were the other end of the pipeline, in which case the game would be solved, otherwise the nodes would be added to the tree and added to an array of already expanded nodes. Then these nodes would be expanded, if they were not found in the array of already expanded nodes, in a similar fashion. This would repeat until either a solution was found or all branches had been fully expanded. In the case of a solution, the server would disable further interaction and

communicate to the connected devices to do the same through a ClientRPC call.

5.5.1.3 Generator repair task

The generator repair task was another new task added to the system. This task required players to locate specific tools in the Augmented Reality scene. Several types, sizes, and colors of tools were created in Autodesk Maya. These were placed as parts of the cabinet marker contents. The tools that were required for the repair were specified in the manual so the players would have to locate them there before searching for them in the cabinets.

When players interacted with a tool by tapping on it, the tool would disappear and a tool icon would be added to the GUI of the game to inform the players that they had acquired the it. The interaction would also be communicated to the server which would update its state and inform the connected devices to do the same, and store the fact that the tool had been picked up for new connections and players that would need to reconnect so their game states could be updated. If a new tool was interacted with the old tool would be returned to its original position which would also be synchronized with the server.

To repair the generator each player needed to acquire one of the necessary tools. They would then need to scan the generator marker, which would show a generator without colors, and drag their tool on to the generator simultaneously to make the repair. This required the players to collaborate as all players were required and they also had to time their actions to solve the task. The generator marker were different to the other markers as it was a physical box in the center of the room with four image markers, one on each side. The virtual generator object was a 3D model created in Maya and contained flip switches, toggles, sliders, and dials.

5.5.1.4 Generator startup task

When the generator had been repaired, the broken generator object would be destroyed and replaced with the four regular generator objects. These had colors and the background was colored by each of the four player colors to provide feedback that the generator could now be interacted with. The players then needed to interact with the input components on the side of the generator that had their player color, i.e. flip switches, toggles, sliders, and dials, of the generator to get it to start.

The interaction with the input components were created as simple movements of the players finger on top of the devices screen, e.g. a slider would require to move the finger in a specific direction, a toggle would require moving the finger across the toggle part, or a dial would require the player to make a circular motion on the dial. Movements of the components were also synchronized with the server and other devices. When one of the components had been interacted with the server would compare all of the components state with the target state to win the game. When that state was reached, further interaction would be inhibited on all connected devices.

5.5.2 Fourth iteration evaluation

The prototype that was developed during the fourth iteration was the most comprehensive prototype, which included several new additions to the game. This prototype was evaluated with a larger group of participants from the target group who had not tried the game before and therefore had no knowledge about the gameplay. The evaluation will be described in the following section.

5.5.2.1 Setting, setup and participants

The evaluation was carried out at a larger FDF meeting that was arranged with the purpose of testing the prototype. The participants were ‘seniorvæbnere’ from FDF Mejdal, FDF Aulum and FDF Haderup-Feldborg. Four groups tried the game. The first group consisted of four ‘seniorvæbnere’ and two leaders, the second of four ‘seniorvæbnere’, the third of four ‘seniorvæbnere’ and two leaders, and the fourth of seven ‘seniorvæbnere’ and one leader.

The evaluation was carried out in a room with no natural lights, which made it difficult to scan some of the image targets.

5.5.2.2 Evaluation

There were several goals for this evaluation. One goal was to test the difficulty of the game, so that it could be balanced before the final evaluation at the event VM19. It was also a goal to look into the user experience of the game. Therefore the participants were asked to fill out a questionnaire after finishing the game. The questionnaire consisted of five parts. The first part consisted of demographic questions, the next three parts were from the GEQ and the last part was the IPQ. The last group did not have time to fill out the questionnaires, therefore

they got a link for the questionnaire, to fill out after they got home. The last goal of the evaluation was to look for usability problems in the game. Mainly in the new game elements that had not been tested before.

5.5.2.3 Results

The players found it confusing to get messages during play that were not directly related to themselves, e.g. if a player completed an oxygen or water task in the other end of the room, the completion feedback would be shown on all devices screens. This was also observed to cause confusion where players thought they had completed a task, when it was actually another player completing another task.

The results of the questionnaire showed that 93.3% of the participants answered that they would like to try the game again and 6.7% that they would not. All the male participants chose that they would like to try again compared to 80% of the females. For the age groups,

The 15-18 group and 19+ group wanted to try the game again in all cases. The 11-14 group would like to try again in 93.3% of the cases and would not in 6.7%

The questionnaire items related to Game Experience were used to calculate the component scores in accordance with the guidelines provided by IJsselsteijn, W., A., de Kort, Y., A., W., and Poels, K. (2013). The descriptive statistics for the core module can be found in figure x, for the Post-game Module in figure y, and for the Social-presence Module in figure z.

	Competence	Sensory and Imaginative Immersion	Flow	Tension/ Annoyance	Challenge	Negative Affect	Positive Affect
Count	16	16	16	16	16	16	16
Mean	2.15	2.89	2.71	1.35	2.93	0.44	3.34
Standard Error	0.19	0.13	0.18	0.18	0.16	0.13	0.13
Standard Deviation	0.77	0.51	0.70	0.70	0.65	0.54	0.51

Table 1. Descriptive statistics for the Game Experience Questionnaire Core Module in the fourth iteration evaluation.

The scores show that the participants experienced a fair to extreme level of Positive Affect and between moderately and fairly, towards fairly, for Sensory and Imaginative Immersion, and Flow. These scores were considered desirable aspects of a game experience so they did not suggest any changes were needed. The challenge component was scored a bit below fairly. The game was observed to be very challenging for the players and the points they obtained where very low. This suggested that the challenge of the game should be lowered

which was done by removing some of the tasks the players needed to complete and giving the players more time to do the tasks. The player competence component was experienced to be slightly above moderately. The players were however observed to have trouble playing the game. In order to support the players in developing the competencies needed for the game, more information was added to the game in the form of a description the players would read before playing the game and hints to help the player during play. The Tension / Annoyance component was scored between slightly and moderately which was considered to be an undesired component. During play, the players encountered some bugs that made the game difficult to play. These were assumed to be one of the factors causing Tension / Annoyance so the game was improved by solving the bugs that were identified before the final game test to reduce this component. Finally, the Negative Component scored between Not at all and Slightly. This was also considered to be an undesirable component so the low score did not give rise to any changes.

	Positive Experience	Negative Experience	Tiredness	Returning to Reality
Count	16	16	16	16
Mean	2.36	0.69	0.38	1.23
Standard Error	0.19	0.15	0.12	0.17
Standard Deviation	0.77	0.61	0.47	0.66

Table 2.Descriptive statistics for the Post-game Module component scores in the fourth iteration test.

The results for the Post-game Module showed that the participants had a moderately to fairly Positive Experience which was also considered a desirable aspect of the game. Negative Experience and Tiredness was found to be between not at all and slightly and were considered undesirable aspects. Returning to Reality was a bit higher than slightly but were considered neither desirable or undesirable. These components did not give rise to any changes to the game design.

	Psychological Involvement - Empathy	Psychological Involvement - Negative Feelings	Behavioural Involvement
Count	16	16	16
Mean	2.89	1.29	3.28
Standard Error	0.15	0.14	0.15
Standard Deviation	0.60	0.57	0.57

Table 3.Descriptive statistics for the social-presence component in the fourth iteration evaluation.

Psychological Involvement and Behavioural Involvement were both desirable component and scored close to fairly and higher than fairly. Psychological Involvement on the other hand were not desirable but did score higher than slightly. No specific changes were made to reduce this score.

5.5.3 Changes for final game

The game was found to be too challenging so the number of tasks the players needed to complete and the amount of time they had to complete them were changed to reduce the time pressure on the players. A warning siren sound was also added to the main console to make critical levels of oxygen and water leakages more noticeable for the players. After the evaluation a bug was found that made it impossible to interact with the generator, which was not noticed in the test as no participants got to that stage of the game. This bug was also fixed and the interactions with the generator was changed to be more in line with similar objects in the real world. The previous implementation was based on the direction the player moved the finger on the device screen but if the player tilted the screen sideways, this movement would no longer fit with the AR contents which had been tilted. The interactions were changed to instead be based on differences between the point of interaction in the scene using raycasting to find that position. More feedback and hints was also added to certain parts of the game to make it easier for the players to figure out how to play and easier to defeat the challenges.

6. Method

The methods that were used in this study, will be described in the following sections. A combination of quantitative and qualitative methods were used to collect the data. Four questionnaires were used to collect quantitative data about different aspects of the participant's user experience playing the game. Video recordings were used on a limited number of the participants to collect qualitative data. The independent variables of the data were the participant's age and gender.

6.1 Data Gathering

The following will describe how the data was collected.

6.1.1 Setting and participants

The data gathering took place at the headquarters of FDF in Copenhagen, at the event VM19. The participants were all members of FDF, who participated in teams of two to six participants in either the age group 11 to 14 or 15 to 18. There were also one or two leaders on each team, who were older than 18.

The data gathering was carried out from 8 am to 5 pm over the course of three days, where the participating teams travelled between different exercises in Copenhagen and the surrounding suburbs. Over the course of three days, 184 teams tried out the game.

6.1.2 Setup

In order to allows as many teams as possible to play the game, it was set up in three rooms, all located in the basement of the building. In each room 22 image markers were taped to the walls, a manual was placed in the room, and the emergency generator marker was placed on a tall round table in the middle of the room. A laptop with a loud speaker attached and an external screen for the main console was also placed in the room. In each room there also were four Android devices for the participants to play the game on. Before the participants entered the room, they were given the possibility of installing the game on their own Android devices in order to save battery power on the devices in the room. Finally, in each room, there also were an extra mobile phone which provided the rooms with mobile hotspots, for the game to run on. The lighting in the rooms would change over the course of the day, according to the weather or the time of the day. This meant that it would sometimes get harder for the devices to scan the image targets, and they would have to be moved to another

location on the walls.

Figure 2. Representation of the game setup

On day two, video recordings were made on a limited number of teams. These teams all played the game in the same room, where two cameras were placed in the corners of the room.

6.1.3 Briefing

When the participants arrived at the location of the exercise, they used an online app provided by VM19, to add themselves to the que of teams who were waiting to play the game. Two exercise facilitators welcomed the teams to the location, handed them a description of the exercise and a manual, in order for them to prepare before playing the game. The description informed the teams about their main tasks during the exercise and about how they would be scored during the game. It also informed them about the rules, gave them a few advices for playing the game and informed them that the game was a master project, and that video recordings might be made, with their consent. Finally the description provided the teams with a QR code, that enabled them to download the application to their own Android devices. See appendix X for a copy of the exercise description.

When a team entered the room where they would play the game, they were welcomed by another exercise facilitator. The team would then be handed the devices they needed to play the game. Before starting the game, the exercise facilitator asked the team whether they

had any questions. When the game was started, the exercise facilitator would stay in the room. If the facilitator noticed any technical difficulties they would try to solve them. When the team had finished playing the game they would receive their score and were then asked to fill out the questionnaires when they left the room. Outside the three rooms, where the teams would play the game, a table was set up with the different questionnaires spread out on the table. The participants were then asked to fill out a questionnaire when they left the room.

6.1.4 Questionnaires

Three different questionnaires were used for the data gathering. The three questionnaires were different modules from the Game Experience Questionnaire (GEQ) by IJsselsteijn, de Kort and Poels (2013). The modules that were used, were the GEQ Core Module, GEQ Social Presence Module, and the GEQ Post-game Module. The three questionnaires had been divided into 14 smaller questionnaires, that each only held questions related to one component. The fourteen components were; Competence, Sensory and Imaginative Immersion, Flow, Tension/Annoyance, Challenge, Negative affect, Positive affect, Psychological Involvement - Empathy, Psychological Involvement - Negative Feelings, Behavioural Involvement, Positive Experience, Negative Experience, Tiredness, and Returning to Reality. Each questionnaire held between two to six questions relating to it's component. At the top of each questionnaire, the participants were asked to fill out their team-id, age and gender. They were also asked the question "Kunne du tænke dig at prøve spillet igen en anden gang?". This question was used in accordance with Schønau-Fog and Bjørner (2013), who argue that continuation desire can be used as an indicator of engagement. The fourteen questionnaires were laid out on a table outside the rooms where the teams played the game. Answering the questionnaires was optional and the participants were told that they could fill out as few or as many as they would like. The participants chose the questionnaires randomly.

6.1.5 Data treatment

The questionnaire data was manually compiled in a digital spreadsheet. The individual component scores for the GEQ were calculated in accordance with the scoring guidelines provided by IJsselsteijn, de Kort and Poels (2013). Scores that were left blank were disregarded and the component scores were calculated with the remaining values. The raw data and component scores for each questionnaire can be found in appendices, x, x, x, x, x. The data

was separated into individual sheets one for the entire data set, one for gender comparison, and one for age comparison. The age scores were changed to the groups 11-14, 15-18 and at least 19 which corresponds to the age groups the participants were divided into for the event. Scores where the age was left blank were removed for the age group.

7. Results

The test results are divided into three parts, first results for the engagement scores will be presented, then the results of the Game Experience Questionnaire, and last the points to the players awarded. Scores where an invalid age had been answered, e.g. age above 100, were disregarded for the results based on age differences. The scores where the gender male and female had been selected were used for the analysis but for the third gender option Other, too few answers had been recorded so they were disregarded for the results based on gender differences.

7.1 Engagement

The question of whether the participants wanted to try the game again at a later occasion was answered by 647 participants. Yes was the response in 94.7% of the cases, No in 1.5% of the cases and Don't know in 3.7% of the cases. Compared to the previous iteration test where 15 participants had answered the question, the Yes response was chosen 93.3% of the cases and No in 6.7% of the cases.

The males responded Yes in 96.7% of the cases which was more than the females who did so in 93.5% of the cases. They also selected No in 1.8% for the males against 1.4% for the females while the females selected Don't know in 5.1% of the cases against 1.5% of the cases for the males. Compared to the previous iteration test the males responded Yes in all cases and the females in 80% of the cases, 20% responded No.

For the different age groups, the 11-14 group selected Yes in 93.8% of the cases, No in 1.1% of the cases, and 5.1% of the cases, while the 15-18 group selected Yes in 93.3% of the cases, No in 3.0% of the cases, and 3.7% of the cases. Lastly, the 19+ group selected Yes in 98.3% of the cases No in 1.7% of the cases and they did not select Don't know in any of the cases. Compared to the previous iteration test, the 11-14 group selected yes in 93.3% of the cases and No in 6.7% while the 15-18 group and 19+ group selected Yes in all cases.

7.2 Game Experience Questionnaire

The results for the component scores of each module of the Game Experience Questionnaire will be presented. The results contain descriptive statistics, comparison to the previous iteration, comparison between genders and comparison between the different age groups, for each module.

7.2.1 Core Module scores descriptive statistics

The scores for the components of the GEQ Core Module were tested for outliers defined as scores that deviated more than 2.5 times the standard deviation from the mean. One outlier were removed from each of the components Competence, Sensory and Imaginative Immersion, Flow, Tension / Annoyance, and Negative Affect while two outliers were removed from the components Challenge and Positive Affect. The remaining scores were used to calculate the descriptive statistics for the components, see table x1 for the descriptive statistics.

	Competence	Sensory and Imaginative Immersion	Flow	Tension/ Annoyance	Challenge	Negative Affect	Positive Affect
Count	44	43	48	52	49	50	57
Mean	2.41	2.88	2.59	0.77	2.49	0.39	3.44
Standard Error	0.13	0.10	0.12	0.11	0.09	0.08	0.08
Standard Deviation	0.87	0.68	0.85	0.81	0.62	0.55	0.59

Table 4. GEQ Core Module component scores after outlier removal

The participants scored Positive Affect between fairly and extremely while Negative Affect and Tension / Annoyance scored between not at all and slightly. Competence, Sensory and Imaginative Immersion, Challenge, and Flow scored between moderately and fairly.

7.2.2 Core Module scores comparison to previous iteration

Each of the component scores of the GEQ Core Module were compared to the corresponding component score from the previous iteration test to see if the experience changed in the final game. Shapiro-Wilk tests were performed to test if the scores from both tests were normally distributed which was the case for Competence and Flow ($p > 0.05$). These two scores were also tested for homogeneity of variance using Levene's tests that showed the scores did not violate this assumption either ($p > 0.05$) and were thus applicable for parametric analysis. Student t-tests were performed to test if the scores were significantly different and showed this not to be the case for either Competence ($t(58) = 1.06$, $p = 0.29$) or Flow

($t(62) = 0.53$, $p = 0.60$). Although the differences were not found to be significant, the difference between the mean of the scores for Competence for the previous iteration ($M = 2.15$, $SD = 0.77$) and the final game ($M = 2.41$, $SD = 0.87$) showed an increase by 0.26 points, and the mean of the scores for Flow for the previous iteration ($M = 2.71$, $SD = 0.70$) and the final game ($M = 2.59$, $SD = 0.85$) showed a decrease by 0.12 points which indicated that a difference might exist.

The remaining five components were tested for significant differences using non-parametric Mann-Whitney tests. The Tension / Annoyance component mean difference between the previous iteration test ($M = 1.35$, $SD = 0.70$) and the final game ($M = 0.77$, $SD = 0.81$) was found to be a significant reduction ($U = 233$, $p < 0.01$) with a medium sized effect ($r = 0.32$). Similarly, tthe mean difference for the challenge component between the previous iteration test ($M = 2.93$, $SD = 0.65$) and the final game ($M = 2.49$, $SD = 0.62$) was also found to be a significant reduction ($U = 247$, $p = 0.03$) with a medium effect size ($r = 0.28$).

The differences between scores from the tests for the last three scores were found not to be significant; Sensory and Imaginative Immersion ($U = 323.5$, $p = 0.36$), Negative Affect ($U = 362$, $p = 0.28$), and Positive Affect ($U = 386.5$, $p = 0.17$). The Negative Affect component mean score difference for the previous iteration test ($M = 0.44$, $SD = 0.54$) and the final game ($M = 0.39$, $SD = 0.55$) showed a decrease by 0.05 points while the difference for Positive Affect for the previous iteration test ($M = 3.34$, $SD = 0.51$) and the final game ($M = 3.44$, $SD = 0.59$) showed an increase by 0.10 points. The means of the Sensory and Imaginative Immersion component showed a decrease by 0.01 points from the last iteration test ($M = 2.89$, $SD = 0.51$) to the final game test ($M = 2.88$, $SD = 0.68$).

7.2.3 Core Module scores comparison between genders

The scores from each gender for each component of the GEQ Core Module were tested for outliers defined as values deviating by more than 2.5 times the standard deviation. One outlier were removed from each of the female's scores for Flow, Annoyance / Tension, Challenge, and Negative Affect and two outliers were removed from the females scores for Positive Affect. For the male's scores one outlier were removed from each of the components challenge and Negative Affect.

The remaining scores were tested for normal distributions for each component for each gender which showed that the data were normally distributed for Competence, Sensory and

Imaginative Immersion, Flow, and Challenge components ($p > 0.05$) for both genders. These components were also tested for homogeneity of variance which was confirmed in each case ($p > 0.05$). This meant that the data satisfied the assumptions for parametric analysis and the component scores were tested for differences using Student t-tests. The Competence component score was significantly higher for the males ($t(42) = 2.23$, $p = 0.03$) with a medium sized effect ($r = 0.33$). The scores for the Sensory and Imaginative Immersion component ($t(40) = 1.12$, $p = 0.27$), Flow component ($t(46) = 0.49$, $p = 0.63$), and Challenge component ($t(46) = 0.62$, $p = 0.54$) were found not to be significantly different. The difference of means between the genders for the components showed, that the males ($M = 2.95$, $SD = 0.61$) had a 0.26 point higher mean than the females ($M = 2.69$, $SD = 0.87$) for the Sensory and Imaginative Immersion component while the females ($M = 2.64$, $SD = 0.82$; $M = 2.54$, $SD = 0.67$) had a higher mean that the males ($M = 2.52$, $SD = 0.90$; $M = 2.43$, $SD = 0.52$) by 0.12 and 0.11 for the Flow component and Challenge component respectively. These differences indicated that a difference between the genders might exist.

The last three components were tested for differences using Mann-Whitney tests. In all three cases the differences were found not to be significant; Tension / Annoyance ($U = 270.5$, $p = 0.86$), Negative Affect ($U = 209.5$, $p = 0.09$), and Positive Affect ($U = 375.5$, $p = 0.63$). The difference between the means for the Tension / Annoyance component were found to be 0.04 higher for the the males ($M = 0.81$, $SD = 0.95$) than the females ($M = 0.77$, $SD = 0.77$). For the Negative Affect component, the females mean ($M = 0.46$, $SD = 0.54$) were found to be 0.24 points higher than the mean for the males mean ($M = 0.22$, $SD = 0.37$). Last, the mean difference for the Positive Affect component between the males ($M = 3.39$, $SD = 0.61$) and the females ($M = 3.48$, $SD = 0.57$) showed a 0.09 points larger mean for the females. The differences between the means suggested that the game experience of participants from the two genders might be different for some of the components.

7.2.4 Core Module scores comparison between age groups

The data for the age groups were tested for outliers defined as 2.5 times the standard deviation and 2 outliers were removed from the Negative Affect component scores for the males, while 1 outlier were removed from the same component for the females and one outlier were removed from the Positive Affect component for the males.

The resulting scores were tested for normality of their distributions and homogeneity

of variances to see if the data could be analysed using parametric methods. The scores for the components Competence, Flow, and Challenge were found to be normally distributed for all three age groups, using Shapiro-Wilk tests ($p > 0.05$), and Levene's tests showed the same scores to satisfy the homogeneity of variance assumption ($p > 0.05$). The data were compared for differences using ANOVA tests and the Competence component ($F(2, 39) = 2.46$, $p = 0.10$), Flow component ($F(2,45) = 0.09$, $p = 0.91$) and Challenge component ($F(2, 45) = 0.71$, $p = 0.50$) were found not to show significant differences. Comparison between the means for the Competence component for the 11-14 group ($M = 2.14$, $SD = 0.99$), 15-18 group ($M = 2.61$, $SD = 0.58$), and 19+ group ($M = 2.94$, $SD = 0.93$) showed that the 11-14 group scored 0.47 points lower than the 15-18 group and 0.80 points lower than the 19+ group while the 15-18 group scored 0.33 points lower than the 19+ group. A mean comparison for the Flow component showed that the 11-14 group ($M = 2.52$, $SD = 0.94$) scored the component 0.14 points lower than the 15-18 ($M = 2.66$, $SD = 1.11$) group and 0.01 points higher than the 19+ group ($M = 2.51$, $SD = 0.71$) while the 15-18 group scored the component 0.15 points higher than the 19+ group. The mean comparison for the challenge component between the 11-14 group ($M = 2.32$, $SD = 0.78$), 15-18 group ($M = 2.47$, $SD = 0.49$), and 19+ group ($M = 2.59$, $SD = 0.71$) showed that the 11-14 group scored the component 0.15 points lower than the 15-18 group and 0.27 points lower than the 19+ group while the 15-18 group scored the component 0.12 points lower than the 19+ group.

The remaining scores were tested for differences using non-parametric Kruskal-Wallis tests which showed no significant difference for Sensory and Imaginative Immersion ($H(2) = 1.27$, $p = 0.53$), Tension / Annoyance ($H(2) = 1.34$, $p = 0.51$), Negative Affect ($H(2) = 0.82$, $p = 0.66$), and Positive Affect ($H(2) = 1.91$, $p = 0.35$). A comparison of the mean differences for the Sensory and Imaginative Immersion component between the 11-14 group ($M = 2.67$, $SD = 0.84$) and 15-18 group ($M = 2.83$, $SD = 0.80$) showed a difference of 0.16 points in favor of the 15-18 group, between the 11-14 and 19+ group ($M = 3.11$, $SD = 0.45$) a 0.44 points larger value for the 19+ group, and between the 15-18 to 19+ group a difference of 0.28 points larger for the 19+ group. The mean differences for the Tension / Annoyance component between the 11-14 group ($M = 0.76$, $SD = 0.78$), 15-18 group ($M = 0.88$, $SD = 1.31$), and 19+ group ($M = 1.18$, $SD = 1.00$) showed a 0.12 points larger mean for the 15-18 group compared to the 11-14 group, a 0.42 points larger mean for the 19+ group compared to the 11-14 group, and a 0.30 points larger mean for the 19+ group compared to the 15-18 group. Comparing the

means for the Negative Affect component scores showed a 0.09 points larger mean for the 11-14 group (M = 0.32, SD = 0.35) than the 15-18 group (M = 0.23, SD = 0.44), a 0.22 points larger mean for the 19+ group (M = 0.54, SD = 0.71) compared to the 11-14 group, and a 0.31 points larger mean for the 19+ group compared to the 15-18 group. A comparison of the means for the Positive Affect component scores showed a difference between the 11-14 group (M = 3.23, SD = 0.67) to 15-18 group (M = 3.42, SD = 0.91) of 0.19 points in favor of the 15-18 group, a difference between the 11-14 group and the 19+ group (M = 3.58, SD = 0.62) of 0.35 points higher for the 19+ group, and a 0.16 points higher score for the 19+ group compared to the 15-18 group.

7.2.5 Post-game Module scores descriptive statistics

The component scores for the Post-game Module were tested for outliers defined as 2.5 times the standard deviation. One and two outliers were removed from Positive Experience and Negative Experience respectively. The remaining component scores were used to calculate the descriptive statistics, see table y1 for descriptive statistics of the Post-game Module scores.

	Positive Experience	Negative Experience	Tiredness	Returning to Reality
Count	36	49	34	48
Mean	2.70	0.27	0.90	1.05
Standard Error	0.14	0.04	0.15	0.12
Standard Deviation	0.83	0.26	0.90	0.81

Table 5. Descriptive statistics for the Post-game Module scores after outlier removal

The scores showed that the positive experience for the participants were rated between moderately and fairly while the negative experience component score were rated towards not at all. Tiredness and Returning to Reality were rated close to and just above slightly, respectively.

7.2.6 Post-game Module comparison to previous iteration

The scores for the Post-game Module from the previous iteration test and the final game test were compared. First the scores were tested for normal distribution using Shapiro-Wilk normality tests which showed that the scores for the Positive Experience component were normally distributed ($p > 0.05$) in both cases. The scores for this component were also test-

ed for homogeneity of variance using a Levene's test which showed the data variance to be homogenous ($1 > 0.05$). This meant the data was applicable for parametric statistical analysis. A Student t-test was used to test if the data was significantly different which was found not to be the case ($t(50) = 1.36$, $p = 0.18$). Looking just at the two means of the scores, an increase of 0.34 points between the previous iteration mean ($M = 2.36$, $SD = 0.77$) and the final game mean ($M = 2.70$, $SD = 0.83$) was found, suggesting a difference might exist.

The other three component scores, which failed the assumption of being normally distributed, were compared for differences using non-parametric Mann-Whitney tests. The result for the Negative Experience component difference between the previous iteration test ($M = 0.69$, $SD = 0.61$) and final game test ($M = 0.27$, $SD = 0.26$) were found to be a significant reduction ($U = 236.5$, $p = 0.02$) and the effect size was found to be medium ($r = 0.30$). The scores of the Tiredness component for the previous iteration test ($M = 0.38$, $SD = 0.47$) and final game test ($M = 0.90$, $SD = 0.90$) were found not to be significantly different ($U = 184$, $p = 0.06$) but the difference between the means was found to be 0.52 points higher for the final game test which suggested that a different might exist. The scores for the Returning to Reality component at the previous iteration test ($M = 1.23$, $SD = 0.66$) and final game test ($M = 1.05$, $SD = 0.81$) were also found not to be significantly different ($U = 329.5$, $p = 0.39$) but the difference between the means was 0.18 points which indicated that the scores might still be different.

7.2.7 Post-game Module comparison between genders

The data for the Post-game Module were tested for outliers defined as scores deviating by more than 2.5 times the standard deviation. A single outlier was found and removed from the Negative Experience component.

The remaining scores were tested for normality using Shapiro-Wilk tests and homogeneity of variance using Levene's tests to establish if they satisfied the necessary assumptions for parametric analysis. This was found to be the case for Positive Experience where the data was normal ($p > 0.05$) and the variance was homogenous ($p > 0.05$). This component was tested for significant difference between the genders using Student t-test which showed that the difference was not significant ($t(35) = 0.03$, $p = 0.98$). The difference between the means of the males ($M = 2.64$, $SD = 1.10$) and females ($M = 2.63$, $SD = 0.83$) was 0.01 points and suggested that a difference might not be present.

The three components which failed to satisfy the parametric assumptions were tested for differences using Mann-Whitney tests. These showed no significant difference between the genders for the Negative Experience Component ($U = 294.5$, $p = 0.84$), the Tiredness component ($U = 135$, $p = 0.75$), or the Returning to Reality component ($U = 268.5$, $p = 0.88$). The means for the Negative Experience component for the males ($M = 0.32$, $SD = 0.93$) and the females ($M = 0.28$, $SD = 0.25$) showed a slightly higher value for the males by 0.04 points. For the Tiredness component the mean different between the males ($M = 0.89$, $SD = 0.98$) and the females ($M = 0.91$, $SD = 0.84$) were 0.02 points higher for the females and for the Returning to Reality component the difference between the mean for the males ($M = 1.08$, $SD = 0.85$) and the females ($M = 1.03$, $SD = 0.79$) were 0.05 higher for the males. These values suggested that a difference might not present.

7.2.8 Post-game Module comparison between age groups

The scores for the GEQ Post-game Module for the age groups were tested for outliers defined as scores differing more than 2.5 the standard deviation. One outlier were removed from the female scores for Negative Experience and one from the male scores for Tiredness.

The remaining scores were tested for normal distribution using Shapiro-Wilk tests and homogeneity of variance using Levene's tests to establish if they were applicable to parametric tests. The scores for the Positive Experience and Returning to Reality component were found to be normally distributed ($p > 0.05$) and have homogenous variances ($p > 0.05$). They were tested for differences using ANOVA tests and showed no significant difference for the Positive Experience component ($F(2,32) = 0.73$, $p = 0.49$) or the Returning to Reality component ($F(2,43) = 1.53$, $p = 0.23$). A comparison of the means for the Positive Experience component showed that the 15-18 group ($M = 2.85$, $SD = 0.69$) had a 0.28 points higher mean than the 11-14 group ($M = 2.57$, $SD = 1.02$), the 11-14 group had a 0.46 points higher value than the 19+ group ($M = 2.11$, $SD = 0.42$), and the 15-18 group had a 0.74 points higher mean than the 19+ group. The same comparison for the Returning to Reality component between then 11-14 group ($M = 1.07$, $SD = 0.86$), 15-18 group ($M = 1.44$, $SD = 0.71$), and 19+ group ($M = 0.82$, $SD = 0.69$) showed a 0.25 points higher mean for the 11-14 group compared to the 19+ group, a 0.62 points higher mean for the 15-18 group compared to the 19+ group, and a 0.37 points higher mean for the 15-18 group compared to the 11-14 group.

The component scores for Negative Experience and Tiredness were tested for differenc-

es using Kruskal-Wallis tests which showed no significant differences for Negative Experience ($H(2) = 0.99$, $p = 0.60$) or Tiredness ($H(2) = 5.86$, $p > 0.05$). The Tiredness component scores were further tested using Mann-Whitney tests between pairs of the groups. These showed a significant difference between the 11-14 group and 15-18 group ($U = 23.5$, $p = 0.02$) with a medium to large effect size ($r = 0.45$), and between then 15-18 group and 19+ group ($U = 4$, $p = 0.04$) with a large effect size ($r = 0.61$) where the 15-18 group scored higher than the others. No significant difference was found for the 11-14 group compared to the 19+ group ($U = 46.5$, $p = 0.69$) but a comparison of the means showed a 0.14 points higher mean for the 11-14 group ($M = 0.64$, $SD = 0.76$) than the 19+ group ($M = 0.50$, $SD = 0.70$). Comparing the means for the Negative Experience component showed a 0.04 points higher mean for the 19+ group ($M = 0.33$, $SD = 0.30$) compared to the 11-14 group ($M = 0.29$, $SD = 0.24$) and a 0.03 points higher mean than the 15-18 group ($M = 0.30$, $SD = 0.41$), while the 15-18 group had a 0.01 points higher mean than the 11-14 group.

7.2.9 Social-presence Module descriptive statistics

The Social-presence Module scores were tested for outliers defined as 2.5 times the standard deviation. One outlier were removed from each component and the descriptive statistics were calculated with the remaining scores, see table 6 for the descriptive statistics.

	Psychological Involvement - Empathy	Psychological Involvement - Negative Feelings	Behavioural Involvement
Count	41	45	43
Mean	3.07	0.85	3.01
Standard Error	0.09	0.09	0.12
Standard Deviation	0.57	0.63	0.81

Table 6. Descriptive statistics for the component scores from the Social-presence Module after outlier removal

The participants scored Psychological Involvement - Empathy and Behavioural Involvement close to fairly while Psychological Involvement - Negative Feelings scored a bit lower than slightly.

7.2.10 Social-presence Module comparison to previous iteration

The scores for the Social-presence Module at the final game test were compared to the scores from the previous iteration test. The scores for each component were found to violate the assumption of normal distribution ($p < 0.05$) in at least one case using Shapiro-Wilk tests

which meant they were not applicable for parametric analysis. Non-parametric Mann-Whitney tests were performed to analyse the components.

The scores for the Psychological Involvement - Negative Feelings component showed a significant reduction ($U = 211.5$, $p = 0.01$) between the previous iteration test ($M = 1.29$, $SD = 0.57$) and the final game test ($M = 0.85$, $SD = 0.63$). The effect size of the difference was medium ($r = 0.31$). The scores for the Psychological Involvement - Empathy component did not show a significant difference ($U = 263$, $p = 0.25$) between the previous iteration test ($M = 2.89$, $SD = 0.60$) and final game test ($M = 3.07$, $SD = 0.57$) but the difference between the means were a 0.18 points increase, indicating a difference might be present. Similarly, the Behavioural Involvement component showed no significant difference ($U = 312.5$, $p = 0.59$) between the previous iteration test ($M = 3.13$, $SD = 0.82$) and the final game test ($M = 3.01$, $SD = 0.81$). The difference between the means for the Behavioural Involvement component were a 0.12 points reduction, suggesting a difference might be present.

7.2.11 Social-presence Module comparison between genders

The GEQ Social-presence scores for the genders were tested for outliers defined as a deviation of at least 2.5 times the standard deviation. One outlier was removed for the males for each of the components Psychological Involvement - Empathy and Psychological Involvement - Negative Feelings while one outlier were removed for the females from the Behavioral Involvement component scores.

The component scores were tested for normality using Shapiro-Wilk and homogeneity of variance using Levene's tests to establish if they were applicable for parametric analysis. This was found to be the case for the Behavioural Involvement component scores where the data was normally distributed ($p > 0.05$) and satisfied the assumption of homogeneity of variance ($p > 0.05$). A Student t-test showed that the difference between the genders for the Behavioural Involvement component was not significant ($t(41) = 0.97$, $p = 0.34$). Comparing the means for the component showed a higher value for the males ($M = 3.17$, $SD = 0.76$) than the females ($M = 2.92$, $SD = 0.84$) by 0.25 points indicating a difference might exist.

The two components which failed to satisfy the necessary assumptions to be applicable for parametric analysis were compared using Mann-Whitney tests. These showed no significant difference for Psychological Involvement - Empathy ($U = 188$, $p = 0.58$) or Psychological Involvement - Negative Feelings ($U = 223.5$, $p = 0.50$). Comparing the means for the

Psychological Involvement - Empathy component between the males (M = 3.04, SD = 0.45) and females (M = 3.09, SD = 0.67) showed a larger value for the females by 0.05 points and for the Psychological Involvement - Negative Feelings component for the males (M = 0.88, SD = 0.46) and females (M = 0.85, SD = 0.81) showed a larger value for the males by 0.03 points. These differences suggested that a difference might not be present.

7.2.12 Social-presence Module comparison between age groups

The scores for the Social-presence Module components for the age groups were tested for outliers defined as scores that deviated more than 2.5 times the standard deviation. One outlier were removed from each of the component scores for the males.

The remaining scores were tested to see if they satisfied the assumptions for parametric analysis. Shapiro-Wilk tests showed that the scores for the Psychological Involvement - Empathy component were normally distributed but the two other component were not. This score were further tested for homogeneity of variance using a Levene's test and were found not to violate the assumption. An ANOVA test showed that there were no significant differences between the groups for the Psychological Involvement - Empathy component ($F(2,36) = 1.24$, $p = 0.30$). Comparing the means showed that the 15-18 group (M = 3.43, SD = 0.43) had a 0.42 points higher mean than the 11-14 group (M = 3.01, SD = 0.61) and 0.33 points higher mean than the 19+ group (M = 3.10, SD = 0.43), while the 19+ group had a 0.09 points higher mean than the 11-14 group.

The other two components were compared using Kruskal-Wallis tests and showed that the Behavioral Involvement component scores were significantly different ($H(2) = 8.27$, $p = 0.02$) while the Psychological Involvement - Negative Feelings were not ($H(2) = 1.95$, $p = 0.37$). Mann-Whitney tests were performed as post-hoc tests between the groups for the Behavioural Involvement component and showed a significantly higher mean for the 15-18 group than the 11-14 group ($U = 35.5$, $p < 0.01$) with a large effect size ($r = 0.51$) and 15-18 group than the 19+ group ($U = 12.5$, $p = 0.01$) with a large effect size ($r = 0.59$), but no significant difference between the 11-14 group 19+ group ($U = 87.5$, $p = 0.50$). Comparing the means between the 11-14 group (M = 2.75, SD = 0.84) and 19+ group (M = 3.00, SD = 0.58) showed a 0.25 points higher mean for the 19+ group. A comparison of the means for they Psychological Involvement - Negative Feelings component between the 11-14 group (M = 0.77, SD = 0.64), 15-18 group (M = 0.87, SD = 0.68) and 19+ group (M = 1.14, SD = 0.51)

showed a 0.10 higher mean for the 15-18 group than the 11-14 group, a 0.37 points higher mean for the 19+ group than the 11-14 group, and a 0.27 points higher mean for the 19+ group than the 15-18 group.

7.3 Points awarded to players

Points were only recorded for the players on one of the event days. A total of 60 scores were recorded and resulted in 78.8 points being awarded on average.

8. Discussion

8.1 Engagement

In general, the results indicate that most of the participants were engaged in the game, with 94.7% of the participants wanting to play the game again.

8.1.1 The effect of Age on Engagement

The level of engagement was similar across all age groups, though mean differences showed that the 15-18 group and the 19+ group reported slightly higher levels than the 11-14 group. This could be explained by the lower amount of participants within these groups, making the results unreliable. The 19+ group being more likely to want to play the game again could be explained by the fact that some of them were more focused on the experience of the children, rather than their own participation in the game. The 19+ group would also be likely to participate in the event again with another group of children, thereby making the game a less repetitive experience.

8.1.2 The effect of Gender on Engagement

The level of engagement was similar across the genders, though the mean differences showed that the males reported slightly higher levels of engagement compared to the girls. This difference could be explained by gaming habits of Danish children in the age of 10-19 years old. Research by Thorhauge and Gregersen (2019), show that males in general play more video games than females. It also shows that 70% of male players are competitive team players, compared to 20% female team players. This research could explain why the game appealed slightly more to males than to females, as the game was played in teams, in a competitive setting.

8.1.3 Engagement compared to previous iteration

The game caused a large degree of engagement in both of the iterations where engagement was measured. The mean difference between genders was greater in the previous iteration compared to the final iteration. This could be explained by the low amount of participants in the previous iteration, causing the results to be less reliable.

8.2 Competence

A part of the game's intended objectives for the player were to explore and figure how the game system worked to develop their competences at playing the game. The game also offered hints and assistance during play which was intended to support the players in acquiring the competences they needed to reach the games objectives. So, the players were not intended to feel competent from the beginning of the game, but rather to develop their competences. The players of the game experienced themselves as being moderately to fairly competent at playing the game. This might suggest that the players did develop their competences as they played.

8.2.1 The effect of Age on Competence

The effect of age on competences did not show significant results but the differences between the means suggested that the 11-14 group felt least competent while the 19+ group felt the most competent. This difference could be explained by the differences in previous experiences. The older player may have found themselves in more situations where they needed to solve unfamiliar tasks, e.g. in other games they have played or in various situation in their longer life spans. This might mean they already had some of the required competences or found it easier to figure out how the game worked and develop the needed competences. Some of the information intended to support the players through the game were also text based, so the older players may have had an advantage as they may have been better readers, especially compared to the youngest participants, which again would make it easier for them to figure the game out and become more competent.

8.2.2 The effect of Gender on Competence

The male players were found to experience themselves as being significantly more competent by a medium amount. According to Thorhauge and Gregersen (2019), a larger ratio of the male population in the age group 10-19 plays video games. This suggests, that the

male gender are more experienced with video games which might explain why they feel more competent when playing such games. Thorhauge and Gregersen (2019) also found, that the males are more competitively inclined than the females. Even though the game was a collaboratively based game, the groups were competing in a larger scale competition where their team score were in competition with other groups. As the males would be more competitively minded, they might be expected to work harder towards winning, which would require them to develop their competences causing them to feel more competent.

8.2.3 Competence compared to previous iteration

The players experience of competences did not change significantly between the previous iteration and the final game. The difference between the means, however, suggest that a difference might exist where the players experienced themselves as being more competent in the final game. Between the two iterations, the game was changed to allow the players to read the game instructions and more hints were added to the game to guide the players. These changes might explain the difference in means.

8.3 Sensory and Imaginative Immersion

The results indicate that the participants felt sensory and imaginative immersion to a fair degree. This could indicate that the game was successful in immersing the participants in the storyworld of the game.

8.3.1 The effect of Age on Sensory and Imaginative Immersion

The results did not show a significant difference between the age groups. However the mean difference showed that the 15-18 group and the 19+ group did report slightly higher levels of sensory and imaginative immersion, compared to the 11-14 group. The 19+ group also showed higher levels of sensory and imaginative immersion than the 15-18 group.

8.3.2 The effect of Gender on Sensory and Imaginative Immersion

The results did not show a significant difference between the genders, the mean difference however showed that the males reported a slightly higher degree of sensory and imaginative immersion compared to the females. This slight difference could be a result of males being more used to playing video games, as seen in the research by Thorhauge and Gregersen

(2019).

8.3.3 Sensory and Imaginative Immersion compared to previous iteration

The results did not show a significant difference in sensory and imaginative immersion compared to the previous iteration.

8.4 Flow

The results showed that the participants felt a moderate to fair degree of flow. This indicates that the difficulty of the game matched the participants skill level, without being neither too easy nor too hard. Factors that might have prevented the participants from reaching a higher level of flow could be distractions in the game environment such as; noise, like yelling or alarms from the neighbouring rooms, people entering the room to talk to the exercise facilitator, photographers taking pictures, and technical difficulties causing them to need help from the play facilitator.

8.4.1 The effect of Age on Flow

The results did not show a significant difference in the levels of flow between the age groups. The mean difference however showed that the 15-18 group reported a slightly higher level of flow compared to both the other groups.

8.4.2 The effect of Gender on Flow

The results did not show a significant difference in the levels of flow when comparing the genders, though the females did show slightly higher levels of flow compared to the males, when looking at the mean difference.

8.4.3 Flow compared to previous iteration

The results did not show a significant difference in the levels of flow compared to the previous iteration.

8.5 Tension/Annoyance

The results showed that the participants only felt a very slight degree of tension/annoyance.

8.5.1 The effect of Age on Tension/Annoyance

The results did not show a significant difference between the age groups. However, when looking at the mean difference the 15-18 group did show a slightly higher level of tension/annoyance compared to the 11-14 group. The 19+ group showed an even higher level of tension/annoyance compared to the other age groups.

8.5.2 The effect of Gender on Tension/Annoyance

The results did not show a significant difference in the level of tension/annoyance when comparing the genders.

8.5.3 Tension/Annoyance compared to previous iteration

The results showed a significant reduction at a medium sized effect in the level of tension/annoyance when comparing the final game to the previous iteration. This indicates that the changes that were implemented from the previous version to the final version were successful in causing the participants to feel less tension/annoyance.

8.6 Challenge

The challenge of the game were experienced as being moderately to fairly challenging. One of the user requirements for the game was that it should be a challenge which it seemed to achieve. According to Sweetzer and Wyeth (2005), there should be a balance between the players skill level and the challenge of the game. Comparing the challenge level to the players' competence level, the two were rated similarly which suggests a good balance was found. The recorded average points on the other hand, might suggest the opposite. The players average points were fairly high which would suggest the game was not challenging enough.

8.6.1 The effect of Age on Challenge

The age did not show to have a significant effect on the challenge. The mean differences, however, showed that the 19+ group found the game to be more challenging than the others and the 15-18 group to find it more challenging than the 11-14 year olds. According to Statistikbanken (2019) and Thorhauge and Gregersen (2019), a lower ratio of players in the older age groups play computer games. This might explain why they find computer games to be more of a challenge as they may be less experienced in dealing with such games.

8.6.2 The effect of Gender on Challenge

The effect of gender on challenge was also not significant but the means for the genders suggested that the females experienced the game as a bigger challenge than the males. Thorhauge and Gregersen (2019) showed, that the females play computer games less than the males, which might suggest that the find games to be more challenging and they are less used to dealing with such games.

8.6.3 Challenge compared to previous iteration

The challenge level was found to be significantly smaller in the final game than the previous iteration with a medium effect. Between the two iterations several modifications were made to balance the game difficulty which was found to be too high in the previous iteration game. Factors of challenge, according to GEQ (2013), include time pressure, which was reduced by allowing the players more time to complete the challenges, and the amount of effort required, which was reduced by removing some of the tasks that needed to be completed. These changes seemed to have had the desired effect that the game achieved a more appropriate level of challenge.

8.7 Negative Affect

The results showed that the participants only experienced very low levels of negative affect. This indicates that the participants did not see the game as a negative experience. According to Vorderer, Hartmann and Klimmt (2003), players find exploring the possibilities in a game enjoyable. While the participants were in the exploratory state of the game, they were met by no competition from the game, except from the timer counting down. Vorderer et al. state that when a game element like the alarm started, it adds a competitive element to the game. When an alarm sounded, the participants were required to become active and try to resolve the task at hand. When the thread was no longer there, the game would inform the players of whether they were successful in solving the problem or not. Vorderer et al. (2003), state that either result will affect the emotional state of the player; a positive outcome will result in positive affect and enjoyment, while a negative outcome will lead to emotions like anger or frustration. This might explain why the participants reported low levels of negative affect, as most of the teams had a positive outcome from the game, with a mean score of approximately 79 points out of a hundred.

8.7.1 The effect of Age on Negative Affect

The results did not show a significant difference in the levels of negative affect between the age groups. However, when looking at the mean difference, the 19+ group did show a slightly higher level of negative affect compared to the other groups.

8.7.2 The effect of Gender on Negative Affect

The results did not show a significant difference in the levels of negative affect between the genders. When looking at the mean difference however, the females reported at slightly higher level of negative affect compared to the males. This coincides with the fact that more females than males answered negatively when asked whether they would like to try the game again.

8.7.3 Negative Affect compared to previous iteration

The results did not show a significant difference in the levels of negative affect when comparing the final game to the previous version.

8.8 Positive Affect

The results showed that the participants felt a a fair to high level of positive affect. This indicates that the participants found the game to be a positive experience, which coincides with 94.7% of the participants reporting that they would like to try the game again. This also confirms that the user requirement of the game being fun was fulfilled. The high levels of positive affect could also be explained by the theory of Vorderer et al. (2003), described in the section about negative affect.

8.8.1 The effect of Age on Positive Affect

The results did not show a significant difference in the level of positive affect between the age groups. The mean difference however shows that the 15-18 group reported a slightly higher level of positive affect, while the 19+ group reported a higher level of positive affect compared to both the other age groups.

8.8.2 The effect of Gender on Positive Affect

The results did not show a significant difference in the level of positive affect between the gender groups. The game was not designed to create a different experience between the

genders so that finding was in line with expectations.

8.8.3 Positive Affect compared to previous iteration

The results did not show a significant difference in the level of positive affect when comparing the final game to the previous version. However, when looking at the mean difference, the participants reported a slight increase in positive affect, indicating that the changes that had been implemented in the game, had a positive effect on the participants experience.

8.9 Positive Experience

The game was found to be a moderately to fairly Positive Experience. This may have been affected by the fact that the players managed to get a large amount of points which would make them feel most of the factors of this component, i.e. the experience felt like a victory, they felt proud, and they felt satisfied.

8.9.1 The effect of Age on Positive Experience

Age was not found to have a significant effect on the Positiveness of the experience. The mean differences, however, suggested that both the 11-14 and 15-18 group felt a more Positive Experience than the 19+ group. The 19+ group of participants consisted of the group leaders which in some groups acted more as facilitators, helping the children, rather than as players of the game. This might explain why they felt lesser degrees of victory, felt less strong, or felt less proud, as they participated in a different way where they were not the ones interacting with the game.

8.9.2 The effect of Gender on Positive Experience

No significant difference was found between the genders and the mean difference was also very small. This suggested, that both genders had a similarly Positive Experience. The game was not designed in a way to create a different experience between the genders so that finding was in line with expectations.

8.9.3 Positive Experience compared to previous iteration

The results for the previous iteration comparison showed no significant difference but the mean difference suggested, that the final game might be a more Positive Experience that

the previous iteration. The games Challenge level was also lowered between the two iterations while the players Competence level increased which seemed to have resulted in creating a more Positive Experience.

8.10 Negative Experience

The players scored the Negative Experience component close to not at all. The game was not designed to give them a Negative Experience so that was in line with expectation. When looking at the scores for the individual factors for this component, the question “I felt that i could have done more useful things” accounted for the majority of the score. This particular question was interesting as some of the participants did not have a device during play so they were very limited in their range of actions during the game session. They might have accounted for a large part of the, already low, score for Negative Experience.

8.10.1 The effect of Age on Negative Experience

The effect of age on the Negative Experience component was found not to be significant. The mean differences between the groups were also very small suggesting that the Negative Experience was not different between players of different ages.

8.10.2 The effect of Gender on Negative Experience

The gender differences between players were also found not to have a significant effect on the Negative Experience component. Likewise was the case for the mean differences which were small indicating that the Negative Experience component was not different for the differing gender groups.

8.10.3 Negative Experience compared to previous iteration

The Negative Experience component was significantly reduced by a medium amount between the previous iteration and the final game. This might be explained by the fact, that the final game was tested as part of the larger VM event, so it may have felt like less of a waste of time which was one of the factors for the component. The previous iteration test also suffered from more difficulties with AR marker scanning causing players to waste time during play. This may have caused them to feel that they could have done more useful things which was another factor.

8.11 Tiredness

The results showed that the participants reported very low levels of tiredness.

8.11.1 The effect of Age on Tiredness

The results showed a significant difference with a large effect between the 15-18 group and two other age groups. This 15-18 group covers the participants that falls under the Senior category at VM19. This is a very competitive category, which might explain why they reported higher levels of tiredness than the other age groups, as they might have used more energy solving other exercises before playing the game.

8.11.2 The effect of Gender on Tiredness

The results did not show a significant difference in the level of tiredness between the genders. The game was not designed to create a different experience between the genders so that finding was in line with expectations.

8.11.3 Tiredness compared to previous iteration

The results did not show a significant difference in the level of tiredness between final game and the previous iteration. However, when looking at the mean difference, the participants reported an increase in tiredness after playing the final version of the game. This result could be explained by the circumstances of the event VM19, where the participants traveled between several different exercises.

8.12 Returning to Reality

The players only had a slight experience of Returning to Reality after the game session had ended. This might be explained by the fact that the game was an augmented reality game. So even though the game combined the real world with a digitally created world it may not have been experienced as a departure from reality causing a sensation of returning upon finishing the session.

8.12.1 The effect of Age on Returning to Reality

The age differences for the Retning to Reality did not show a significant difference. The differences between the means did however indicate that the different age groups might have

experienced this somewhat differently. The differences suggested that both the 11-14 group and 15-19 group had more of an experience of Returning to Reality. As the 19+ group contained leaders, some of which did not act as players in the game, they may have not had this experience since they did not experience the digital world.

8.12.2 The effect of Gender on Returning to Reality

The gender differences did not show a significant difference. The difference between the means of the gender groups were also small suggesting that the genders did not have differing experiences of Returning to Reality.

8.12.3 Returning to Reality compared to previous iteration

The component difference between the previous iteration and final game was not significant but the mean differences indicated a decreased sensation of Returning to Reality for the final game. This might be explained by the fact that the game was part of the overarching competition rather than being a thing in itself, as was the case for the previous iteration. As the participants may have had the VM competition in mind they may have been less inclined to become detached from reality.

8.13 Psychological Involvement - Empathy

The results showed that the participants reported a fair amount of psychological involvement - empathy. This might be explained by the fact that the participants have entered the competition in teams, with people that they know are used to solving exercises with. Good teamwork is one of the main focus points at VM19, which could also be one of the explanations as to why the teams reported high levels of psychological involvement - empathy, because they are very focused on their teamwork.

8.13.1 The effect of Age on Psychological Involvement - Empathy

The results did not show a significant difference in the level of psychological involvement - empathy between the age groups. However, the mean difference showed that the 15-18 group reported a slightly higher level of psychological involvement - empathy.

8.13.2 The effect of Gender on Psychological Involvement - Empathy

The results did not show a significant difference in the level of psychological involvement - empathy between the genders. The game was not designed to create a different experience between the genders so that finding was in line with expectations.

8.13.3 Psychological Involvement - Empathy compared to previous iteration

The results did not show a significant difference in the level of psychological involvement - empathy when comparing the final game with the previous version. However, the mean difference indicates that the participants reported a slight increase in psychological involvement - empathy.

8.14 Psychological Involvement - Negative Feelings

The players only reported a slight experience of Psychological Involvement - Negative Feelings. This might be due to the collaborative aspect of the game which made the players work together rather than against each other while playing the game. This might have made them less likely to feel jaloux or revengeful.

8.14.1 The effect of Age on Psychological Involvement - Negative Feelings

The age difference did not have a significant effect on the Psychological Involvement - Negative Feelings. The differences between the means did however show that the 19+ group had a higher score. One of the factors of this component was, “I influenced the mood of the others”. Since most of the players from this group were group leaders, they may have been more inclined to experience themselves as influencing the mood of the group.

8.14.2 The effect of Gender on Psychological Involvement - Negative Feelings

The gender differences did not show a significant result for the Psychological Involvement - Negative Feelings nor did the small mean difference suggest that the experience were different between the genders.

8.14.3 Psychological Involvement - Negative Feelings compared to previous iteration

The Psychological Involvement - Negative Feelings showed a significant reduction by a medium amount between the previous iteration test and final game test. When looking at the

factor scores rather than the component score, the two factors “I influenced the mood of the other(s)” and “I was influenced by the other(s) moods” scored higher in the previous iteration test than the final games test. During the previous iteration test, more technical problems were encountered. This might have affected the players moods, as they would be more frustrated by the problems, and may also have caused them to affect each others moods. This would explain why the component was scored lower in the final test where most of the technical problems had been solved.

8.15 Behavioural Involvement

The results showed that the participants reported a fair level of behavioural involvement. This could be explained by the focus on teamwork, that was also mentioned in psychological involvement - empathy.

8.15.1 The effect of Age on Behavioural Involvement

The results showed a significant difference between the age groups. The 15-18 group reported significantly higher behavioural involvement with a large effect, compared to the two other age groups. When looking at the mean difference the 19+ reported a slightly higher level of behavioural involvement than the 11-14 group.

8.15.2 The effect of Gender on Behavioural Involvement

The results did not show a significant difference in the level of behavioural involvement when comparing the genders. However, when looking at the mean difference, the males reported a slightly higher level of behavioural involvement than the females.

8.15.3 Behavioural Involvement compared to previous iteration

The results did not show a significant difference in the level of behavioural involvement when comparing the final game and the previous version. The mean difference however, shows that the participants reported a slightly lower level of behavioural involvement, when comparing the final game to the previous version.

8.16 Sources of error

The sources of error for the project, will be described in the following section and their

influence on the results will be discussed.

8.16.1 Difference between test setups

Even though the three rooms that the game was played in at Rysensteen was very similar, there were some differences between them two. The rooms were different sizes and they held different kinds of furniture. Before setting up the game, most of the furniture had been moved out, but some of the furniture was too big to be moved. This resulted in the game being set up in three different ways in the rooms, as the image targets could not be hung in the same places in the different rooms. This may have resulted in different experiences for the participants.

8.16.2 Issues with image targets

There were some problems with the image targets that affected the teams very differently. The fuel pipe image targets were quite difficult to read for some devices. This problem was especially pronounced when the light conditions were problematic. Because the rooms were not the same size, and the windows were not placed in the same way, the amount of natural light varied a lot from room to room and would also change depending of the weather and the time of the day. This resulted in fuel pipe image targets in one of the rooms, being very difficult to scan. only certain devices were able to scan the image targets, and the image targets had to be moved to a place where they were easier to scan for the participants. This resulted in the three image targets with the fuel pipe puzzles being collected in one place in one of the rooms. They were more difficult to spot in this place, causing some teams to look for the fuel pipe targets for quite a long time. The rest of the fuel pipe image targets, were near impossible to scan, making it necessary for the exercise facilitator in the room to inform these teams of what they would have seen, if their devices had been able to scan the image target. These problems may have caused the participants to experience the game in very different ways, depending on what room they played the game in.

8.16.3 Differences in the moderation

The exercise facilitator would try to give the teams the same introduction, but because there were five different exercise facilitators over the course of the three days, it was very difficult to give all teams the exact same information. The teams would all be given the same

information when they arrived at the exercise, but as the teams had to read the exercise description and the manual themselves before starting the exercise, it was clear that they did not read this information to the same extent. It was also different how many questions the teams would ask the facilitator while they were playing the game. These issues may all have led to the teams having experienced the game in different ways.

8.16.4 Disturbances

While the teams were playing the game, there were different situations that may have disturbed them. The rooms that the game was played in, were not soundproof. This meant that the teams were able to hear the other teams that were playing in the neighbouring rooms, but in some instances they were also able to hear the alarms sounding from the other rooms. Other disturbances could be people watching from outside the windows or other people entering the room in order to speak to the exercise facilitator. Some teams also experiences that photographers from VM19 would enter the room while the were playing, either in order to take photos or to make video recordings while the teams were playing.

8.16.5 Technical issues

Several technical issues were encountered over the course of the three days. One issue was the difference in quality of devices in the different rooms. The camera quality differed a lot from device to device. This had an impact on how easy it was to scan the image targets. The devices also had very different battery capacities. This resulted in some of the participants having to carry around a power bank that was attached to the device they were playing with during the game. This made it more difficult for the participants to move freely. A few times it also happened that a device would run out of battery while a team was using it. If this happened the team would be given another device to play with, and if they had lost points due to these problems, they would be added to the score at the end of the game.

8.16.6 Concerns with video recordings

The main concern with video recording the teams while they were playing the game, is the risk of the Hawthorne effect. According to Preece et al. (2002) video recordings, compared to other recording methods, can be used to capture the most complete amount of data (p. 376). As mentioned, the problem with video recordings is that the presence of cameras in

the room can feel obtrusive to the participants. This can cause the Hawthorne effect, which is related to participants changing their behaviour because of the perceived extra attention (Preece et al., 2002, p. 376; Onwuegbuzie & Leech, 2007). Recordings were only made of four groups, but the video cameras were also present in the room when participants who were not being recorded played the game. Therefore the presence of the cameras may have caused the Hawthorne effect, and thereby altered behavior, for all the teams that played in this specific setup.

8.16.7 Issues with the questionnaires

There were several possible issues with the questionnaires. One issue is that there were no titles on the questionnaires, putting the questions into context. This could mean that the participants have understood the questions differently from how they were intended. Another issue is that the participants answered the questionnaires under very casual circumstances, where they were able to talk to each other about the questions and the answers. The participants also answered the questions based on different prerequisites, because only four participants on each team had a device to play with. The results and discussion does not take this into consideration. This could have been solved by asking the participants whether they had played the with a device or not on the questionnaires. This might have provided the discussion with different results. This could also be an explanation for why the data is not normally distributed in a lot of cases. Another issue could be that the gender category includes all age groups, including the leaders. The leaders may have played the game differently than the other participants, because their focus is on ensuring that the team works well together.

8.17 Conclusion

This study set out to explore the game experience and engagement in a collaborative co-located augmented reality game for Væbnermesterskabet 2019.

The study found that gender influence the feeling of competence in a collaborative co-located augmented reality game. The study also found that age influence the level of tiredness and behavioural involvement.

A comparison between the final version of the game and the previous version, showed that the reported levels of tension/annoyance, challenge, negative experience and psychological involvement - negative feelings had been reduced.

9. References

Bell, A. (2007). Designing and testing questionnaires for children. Journal of Research in Nursing, 12(5), 461-469.

Biocca, F., Burgoon, J., Harms, C., & Stoner, M. (2001). Criteria and scope conditions for a theory and measure of social presence. Presence: Teleoperators and virtual environments.

Books & Magic. (2015, June 11). The Little Mermain teaser [Video File]. Retrieved from https://www.youtube.com/watch?v=pqsFx_CHv44

Chamary, J. V. (2018, Feb. 10). Why 'Pokémon GO' Is The World's Most Important Game. Forbes. Retrieved from https://www.forbes.com/

Danish Standards Association. (2002). Ergonomics of human-system interaction - Usability methods supporting human-centred design (DS/ISO/TR Standard No. 16982).
Retrieved from https://sd-ds-dk.zorac.aub.aau.dk/Viewer?ProjectNr=41409

Danish Standards Association. (2010). Ergonomics of human-system interaction - Human-centred design for interactive systems (DS/EN ISO Standard No. 9241-210).
Retrieved from https://sd-ds-dk.zorac.aub.aau.dk/Viewer?ProjectNr=M246855

Danmarks Statistik. (2019, K1). Danmarks Statistik: Befolkningens brug af kulturaktiviteter indenfor de seneste tre måneder efter kulturaktivitet, tid og alder og køn
Retrieved from: https://www.statistikbanken.dk/KVUHOVED

de Kort, Y. A. W., IJsselsteijn, W. A, & Poels, K. (2007). Digital Games as Social Presence Technology: Development of the Social Presence in Gaming Questionnaire (SPGQ). Proceedings of the 10th Annual International Workshop on Presence.

Fullerton, T. (2014). *Game design workshop: a playcentric approach to creating innovative games*. CRC press.

IJsselsteijn, W. A., de Kort, Y. A. W., & Poels, K. (2013). The Game Experience Questionnaire. Unpublished manuscript.
Retrieved from https://pure.tue.nl/ws/files/21666907/Game_Experience_Question-naire_English.pdf

Jack Hunter Games. (2014, February 18). Jack Hunter: The French Connection - PR

Video & Interview with martin King [Video file]. Retrieved from https://www.youtube.com/watch?v=DxiWR5j9bQM

Lietz, P. (2008). Questionnaire Design in Attitude and Opinion Research: Current State of an Art. Lokaliseret på/fra http://www.priorisierung-in-der-medizin.de/documents/FOR655_Nr13_Lietz.pdf

Onwuegbuzie, A. J., & Leech, N. L. (2007). Validity and qualitative research: An oxymoron?. Quality & Quantity, 41(2), 233-249. doi:10.1007/s11135-006-9000-3

Preece, J., Rogers, Y., & Sharp, H. (2002). Interaction Design beyond human-computer interaction. West Sussex, UK: John Wiley & Sons Ltd.

Preece, J., Rogers, Y., & Sharp, H. (2015). Interaction design: beyond human-computer interaction (4th ed.). Glasgow: Bell & Bain Ltd.

Schønau-Fog, H., & Bjørner, T. (2012). "Sure, I Would Like to Continue": A Method for Mapping the Experience of Engagement in Video Games. Bulletin of Science Technology & Society, 32(5), 405-412.
Doi: https://doi.org/10.1177/0270467612469068

Statista (2019). Number of smartphone users worldwide from 2014 to 2020 (in billions). Retrieved from https://www.statista.com/

Sweetser, P. & Wyeth, P., (2005). GameFlow: a model for evaluating player enjoyment in games. ACM Computers in Entertainment, 3(3), 1-24.

Thorhauge, A. M., & Gregersen, A. L. (2019). Individual pastime or focused social interaction: Gendered gaming practices among Danish youth. New Media & Society.

Unity3D [Computer Software]. (2018). Retrieved from https://unity3d.com

Vorderer, P., Hartmann, T., & Klimmt, C. (2003). Explaining the enjoyment of playing video games: The role of competition. ICEC '03 Proceedings of the second international conference on Entertainment computing (pp. 1-9). Pittsburg, PA: Carnegie Mellon University.

Zagal, J. P., Rick, J., & Hsi, I. (2006). Collaborative games: Lessons learned from board games. Simulation & Gaming, 37(1), 24-40.