Aalborg University Copenhagen
School of Information and Communication Technology
Medialogy

Engaging by Design: A Study on Engagement using Game Design Elements in an Informative First Aid Application

Hans Hansen, Mikkel Skovsmose Tewes, Nicolai Foldager

Submitted in part fulfilment of the requirements for the degree of
Bachelor of Science of Medialogy, Aalborg University Copenhagen, May 2016

Abstract

This study was aimed at developing an engaging application for first aid with the use of gamification. In Denmark, around 3.500 people suffer from out-of-hospital cardiac arrest each year, and it can therefore be imperative to keep your cardiopulmonary resuscitation (CPR) knowledge up to date. Gamification is defined as using game design elements in a non-game context in order to get users more engaged. This study will therefore focus on the following research question:

"How can game design elements be used to create an engaging and informative application about first aid?"

The attributes used for evaluating whether users were engaged were; positive affect, interest, challenge, perceived time, motivation, aesthetics, clear goal, and feedback.

Two prototypes were developed; one with game design elements and one without. A between-group design was used, where two different groups tested one prototype each. Data was gathered through data logging, interviews and observations. The interviews were semi-structured and contained a verbal numeric rating scale (VNRS). The observations were used for analysing facial expressions where two rating scales were developed for measuring enjoyment and concentration, which were based on the Facial Action Coding System (FACS).

The results showed that there was a minor increase in engagement when the prototype was gamified. Furthermore, the results indicated that the participants' motivation was affected by achievements and progression.

Contents

List of Tables

List of Figures

Chapter 1

Introduction & Research Question

Every year 3.500 people experience out-of-hospital cardiac arrest in Denmark. In 2011, only 57,9% out of these received cardiopulmonary resuscitation (CPR) before an ambulance arrived. The number of CPR recipients has been increasing every year, however it is not that high considering that receiving CPR triples the survival rate. As approximately 72,3% of out-of-hospital cardiac arrests occur in the home (Wissenberg, Hansen, Mortensen, Folke, Lippert, Christensen, Hansen, Lang-Jensen, Henriksen and Torp-Pedersen, 2013), it is important for the population to keep their CPR knowledge up to date, as chances are that there is not anyone else present to help them in an emergency situation.

A mobile application will be developed to help motivate people refresh their first aid knowledge. During the development of the application an engaging design will be emphasised, since users will look for another solution that they find more appealing if they are not engaged (O'Brien and Toms, 2008). We want to see if game design elements can provide a better experience and help increase user engagement. Several prior studies have researched how game design elements affect learning (Dickey, 2005; Faghihi, Brautigam, Jorgenson, Martin, Brown, Measures and Maldonado-Bouchard, 2014; Khaddage, Lattemann and Acosta-Díaz, 2014; Hwa, 2009; Pivec, Dziabenko and Schinnerl, 2004). Based on this the following research question was formed:

"How can game design elements be used to create an engaging and informative application about first aid?"

Two versions of the application will be developed to test the research question. The two will be nearly identical, except one will include game design elements. A between-groups test design will be used to see if there is a difference in user engagement between the two versions. The participants will either use the gamified or non-gamified version and the evaluation method will include data logging, semi-structured interviews and observations of the participants.

or

Chapter 2

Analysis

2.1 Engagement

Many scholars have come up with definitions of engagement. DeRose and Laurel (1993, p. 175) state that playfulness has to be emphasized, O'Brien and Toms (2008) propose a model of engagement in various platforms, Quesenbery (2003, pp. 82-83) suggests that engagement is a dimension of usability. Shernoff, Csikszentmihalyi, Shneider and Shernoff (2003, pp. 159-160) state that engagement is affected by several factors, some of them being perceived control, achievements and learning.

O'Brien and Toms' (2008) proposed model of engagement consists of three states (point of engagement, period of engagement, and disengagement) where each state consists of several attributes. We chose to focus on motivation, interest, aesthetics, challenge, feedback, perceived time and positive affect (enjoyment). Some of these attributes will be combined with Sweetser and Wyeth's (2005) criteria for enjoyment in games. Even though the criteria from Sweetser and Wyeth are meant for games, we suggest that they can still be used for gamified applications.

Based on O'Brien and Toms' (2008) model of engagement, Csikszentmihalyi's (1993) flow theory and Sweetser and Wyeth's (2005) GameFlow model, attributes for improving user engagement were formed. It was decided that these attributes were important for investigating user engagement in this study: feedback, clear goal, positive affect, aesthetics, challenge, interest, motivation, and perceived time.

Feedback and clear goals are often the reasons for flow to occur in the first place (Csikszentmihalyi, 1993), and clear goal is part of point of engagement while feedback preserves engagement (O'Brien and Toms, 2008). Both of these are also part of enjoyment (Sweetser and Wyeth, 2005). Aesthetics can also initialise and preserve engagement as the visual appeal of the system is important to catch the users interest (O'Brien and Toms, 2008).

Positive affect refers to fun and enjoyment (O'Brien and Toms, 2008), where concentration also

can affect enjoyment (Sweetser and Wyeth, 2005). It was deemed important as we wanted to motivate the users to refresh their first aid knowledge. O'Brien and Toms (2008) suggest that engagement shares attributes such as intrinsic motivation with flow. Motivation is needed to get users interested in the activity. Interest is therefore both in the point of engagement and period of engagement (O'Brien and Toms, 2008), and interest is what makes a user enjoy an activity (Csikszentmihalyi, 1993).

Challenge is an important attribute for engagement (O'Brien and Toms, 2008), enjoyment (Sweetser and Wyeth, 2005), and flow (Csikszentmihalyi, 1993), where the player skill and challenge have to be nearly matched. This means that it is important that the difficulty level of the content in the application is higher than the user's skill level.

O'Brien and Toms (2008) describe perceived time as not having sufficient time to finish a task, however, the attribute perceived time in this study is based on the altered sense of time, where time seems to pass faster when a user is enjoying a task (Csikszentmihalyi, 1993).

2.2 Gamification

The term "gamification" refers to the use of game elements in a non-game environment, with the intention of providing a better experience for the user and making them more engaged (Deterding, Sicart, Nacke, O'Hara and Dixon, 2011). The use of gamification can be seen in many applications today. Nike+, an exercise application for mobile devices, uses daily goals and achievements to help the users stay motivated (Nike, 2016). The Nissan Leaf car has a built-in gamified system, which helps drivers drive more eco-friendly (Nike, 2016).

2.2.1 Motivation and User Types

Users' motivation is what makes them play a game in the first place, and it can therefore be important to understand what motivates people to play games (Zichermann and Cunningham, 2011, p. 15). Charlier and De Fraine (2013) investigated the effectiveness of learning first aid through a board game. Their test was based on a traditional lecture and a board game, where the lecture showed to be more effective for learning, but the participants were more motivated to play the board game. A combination of the lecture and the game would have been more effective (Charlier and De Fraine, 2013).

All users are not motivated by the same things, some prefer hard fun while others prefer easy fun (Lazzaro, 2004), and it can therefore be useful to know what types of users exist. Bartle (1996) created four user types for the purpose of differentiating users in games. The same can be said about the types of fun defined by Lazzaro. Marczewski (2015) combines Lazzaro's (2004) types of fun with Bartle's (1996) user types and attempts to modify them for gamified systems.

There are certain elements that motivate these different user types, and if you are designing an application for specific user types, then it is valuable to know what attracts them.

2.2.2 Design Elements

Game design elements refers to elements that often are used in games. Examples of such elements are badges, leaderboards, levels, and challenges (Deterding, 2011; Robson, Plangger, Kietzmann, McCarthy and Pitt, 2016). Many of these are also related to what Sweetser and Wyeth (2005) present in their GameFlow table.

Webster (1988) investigated playful activities, she concludes that game elements can have a positive effect on engagement. Adding a game design element into a non-game application does not automatically mean that it will increase the motivation of users to use the application (Deterding, 2011; Von Ahn and Dabbish, 2008). Malone's (1982) heuristics for enjoyable interfaces includes challenges, fantasy, and curiosity, and he also mentions that these elements are not for all users. It is therefore imperative to choose elements wisely for the intended user, else the elements might not have any effect on the user's motivation (Flatla, Gutwin, Nacke, Bateman and Mandryk, 2011). This leads back to Marczewski's (2015) user types that were discussed in the previous section, where some types are attracted by challenge and some by curiosity.

Applying gamification to applications needs to be considered wisely before the actual implementation, because aspects such as engagement, motivation and user types play a role in whether it makes sense to use game elements or not. With these aspects in mind, it seems evident to apply gamification to an informative application about CPR, which could make the application more engaging, and the outcome would be to learn a vital skill.

The intrinsic motivation is different between the various user types defined by Marczewski (2015), where we have chosen to use elements from some of them. Philanthropists seek serious fun, meaning, and care-taking (Marczewski, 2015), as first aid is meaningful and has a purpose, the application should be designed to be fun while still having a serious objective. We have chosen to include challenges and progression, which belong to the Achiever as we want them to feel challenged and motivate them to continue to use the application. The use of rewards should be included in the application as users are extrinsically motivated by them (Marczewski, 2015), where achievements and points should be implemented. The chosen game design elements can be seen in figure 2.1.

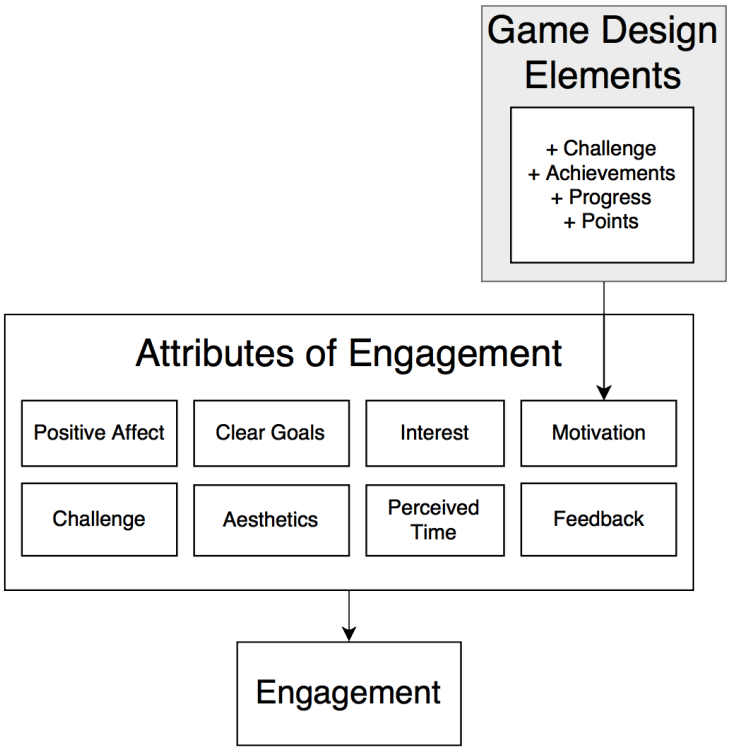


Figure 2.1: Visualisation of the attributes of engagement including the chosen game design elements.

2.3 Informative Application

To define an informative application a revised version of Bloom's Taxonomy of Educational Objects (Anderson, Krathwohl and Bloom, 2001) will be used. The benefit of the revised framework is that it works in two dimensions; the cognitive process dimension and the knowledge dimension (Anderson et al., 2001). Additionally, the revised framework puts emphasis on the subcategories in the cognitive dimension, where the original puts emphasis on the six main categories (Krathwohl, 2002). These factors will help define the learning outcome.

Both frameworks have 'apply' (category 3) as a prerequisite for 'analyze' (category 4). In this case, 'analyze' does not necessarily refer to the act of breaking down material in order to find interrelationships, but more so the act of analyzing a scenario and what the proper procedures are. Therefore, we suggest that 'apply' is not a prerequisite for 'analyze', but the other way around. In order for the users to perform the proper procedures in an emergency situation, they will have to be able to analyze a scene before they can decide on how to handle it.

As the users should already have gathered basic information and understanding (category 1 and 2) from taking their first aid certificate, it has been decided to put less emphasis on these categories. However, some emphasis will have to be put on them as repetition of knowledge can help the users' to recall and recognize previously acquired knowledge.

As performing CPR is a physical activity, it is vital that the users learn how to apply the theory into practice. In this context 'apply' might be the most important category, as the users will have to transfer their learning by answering questions about a real-life scenario. To increase the chances of transfer of learning, the audience should be surveyed to ensure that the instructional content is on par with their knowledge (Ford and Weissbein, 1997).

Ford and Weissbein (1997) also specifies that guided error-based learning can further transfer of learning. Error-based learning leans towards the hard fun and challenge section of Marczewski (2015). Furthermore, (O'Brien and Toms, 2008) also describes that the users should feel challenged in order to be engaged and Csikszentmihalyi (1993) describes challenge as a key aspect in getting users to be motivated.

Following the revised framework the learning outcome for the users has been set to procedural in the knowledge dimension and category 4 (apply) in the cognitive process dimensions. This should enable the users to analyze a scenario, know the proper procedures and apply them to casualties. The level of the content in the application should be higher than the user's level, as providing a challenge can further engagement (O'Brien and Toms, 2008; Sweetser and Wyeth, 2005; Csikszentmihalyi, 1993). However, it should not be too advanced as this can hinder the transfer of knowledge (Ford and Weissbein, 1997).

2.4 Previous Work

Two mobile applications, Røde Kors Norge (Red Cross Norway) and Lifesaver, will be presented, discussed and evaluated, since they both have interesting functionality which we could use as a source of inspiration.

Røde Kors is a mobile application, which can both help users' handle an emergency scene and test their knowledge by giving them questions regarding first aid (Røde Kors, 2014). The application provides simple steps for what a users should do in certain situations. These steps do not contain much text, which is seen as a quality, as too much text could become boring and disengage the user. Their knowledge is tested through challenging questions, since challenges can be motivating for some users (O'Brien and Toms, 2008; Sweetser and Wyeth, 2005; Marczewski, 2015). The Røde Kors application is not gamified and there is not much interactivity other than selecting a topic to get some information or selecting an answer to a question, however repetition helps people remember.

Lifesaver is a gamified mobile application, which teaches you CPR and first aid through interactive videos, medical questions and emergency scenarios (Unit9, 2013), you can see the lifesaver's interface in 2.2.

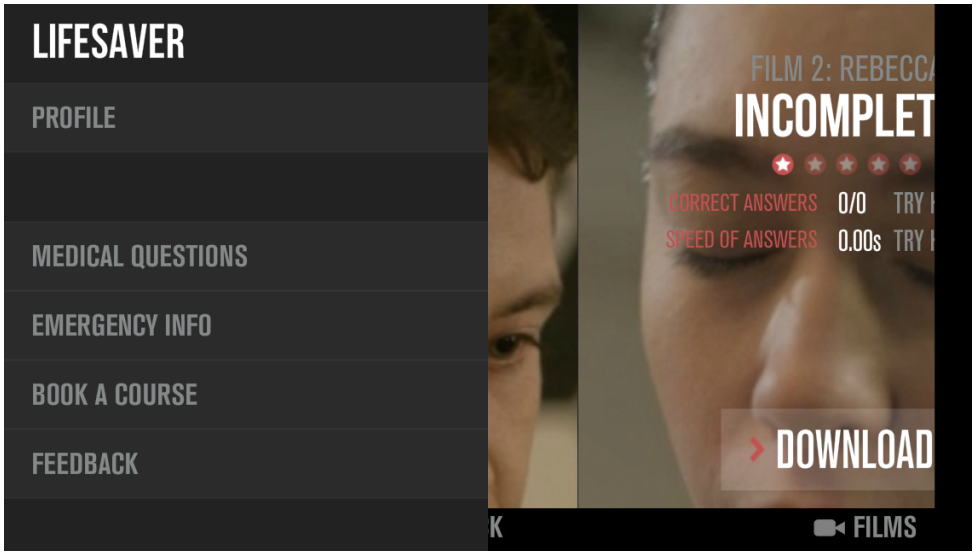


Figure 2.2: Lifesaver's user interface with progression displayed

In the video challenges you have to react as fast as possible if you want the highest score, however this is not our desire, since we want them to analyze (Anderson et al., 2001) the situation, and then apply (Anderson et al., 2001) their knowledge. Occasionally, the interactive videos only had one answering option which is not optimal if you want to challenge and engage the user (O'Brien and Toms, 2008). One of the scenarios also features physical activation of the user. In order to simulate performing chest compressions on a casualty in the scenario, the user has to move their phone up and down.

Both Røde Kors and Lifesaver contains features that would be beneficial to have in our prototype. Elements such as clear feedback and challenging aspects are featured in both applications and are supported by O'Brien and Toms (2008) model of engagement and Sweetser and Wyeth (2005) attributes for enjoyment in games. Furthermore, as CPR is a physical activity it would be beneficial if the application physically activates to user as Lifesaver does.

While Lifesaver physically activates the user it failed on challenging the user at times as multiple questions only had one answering option. Furthermore, the application did not have the interactive scenarios installed, this could disengage the users as they had to wait for them to download. Table 2.1 shows the advantages and disadvantages of the two applications.

Table 2.1: Advantage and disadvantage for the two applications presented.

Application	**Advantages**	**Disadvantages**
Røde Kors	+ Multiple topics + Illustrations for each topic + Short description for teaching first aid + Multiple-choice questions	- Ordinary interface - No physical activation
Lifesaver	+ Interactive scenarios + Challenging questions + Physical activation + Guiding narrative + Clear feedback + Short description for teaching CPR	- Download videos - Only one answering option sometimes - No subtitles in scenarios

Some advantages derived from these two examples are repeated by several studies on engagement. Interactivity and clear feedback are described by O'Brien and Toms (2008) as attributes of engagement and clear feedback is a part of Sweetser and Wyeth's (2005) GameFlow model. In addition, Csikszentmihalyi (1993) states that the state of flow usually begins when clear feedback and clear goals are present. These attributes will be emphasised in the design in order to hopefully increase engagement, however the others will still be taken into consideration.

Chapter 3

Methods

The test was primarily conducted at Aalborg University Copenhagen in Copenhagen on Tuesday May 9, 2016 through May 12, 2016 . These test sessions were conducted between 09.00 and 16.00 o'clock. However, four tests were conducted at a test conductor's home in Copenhagen North West between 20.00 and 22.00 o'clock on May 10, 2016. Furthermore, as the test was a between-groups design it consisted of two groups with ten participants in each. The male to female ratio was 1:1 in both groups, five males and five females in each group. Table 3.1 shows four steps that were followed during the test and in which order they were followed.

Table 3.1: Steps followed in the test.

No.	**Activity**	**Data gathered**
1	Using the application	Video recording Logged data
2	Interview	Verbal Numeric Rating Scale Unstructured interview Perceived Time
3	Questions about a situation	Their knowledge of first aid
4	Questions about the application	How much do they remember?

3.1 Test Design, Target Group & Sampling

Two different versions of the application were used for the test; one with game design elements (gamified version) and one without (non-gamified version). A between-group test design was used, where each participant was using either of the two versions. The participants were not told what group they belonged to or what the test was evaluating in order to avoid the participants' expectations possibly biasing the test. Had a participant known that they were testing the non-gamified version of the application, they may have decided on their expectations prior to

the test and in turn not use the application for the entire duration of time they normally would have.

A major factor for deciding upon a between-groups design was the learning aspect, as one of the measurements was the participants learning outcome. This data was gathered by asking them questions about the interface elements and first aid procedures. This would not have been possible using a within-groups test design, as the participants would have received information about first aid throughout the use of both versions. Furthermore, as we were primarily looking for user engagement, the novelty of the application would have been drastically reduced in the second version they tried. While it took approximately the same amount of time as a within-group test would have taken, it did require significantly more test participants and made sampling more difficult as the two groups should match in demographics.

The test setup was mobile as the application was installed on tablets. This made it easy to set it up and change location when needed. The setup consisted of a tablet with the application installed and a camera to record the participant. The tablet was inclined by 25 degrees in order to make more of their face visible to the camera. It was placed approximately 140 centimeters in front of the participants and approximately 35 centimeters below the participant's eye level at an angle of 10 degrees to film a bottom-up view of their face as their heads were facing down towards the tablet.

The requirements for the target group was set to the following:

- Has prior experience with first aid
- Between the age of 20 and 28 years old
- Lives in Denmark
- Speaks and understands Danish

The first requirement for the target group was that they had prior experience with or training in first aid. The reason for this was that we wanted to help people to keep their knowledge about first aid up to date, as the CPR standards are reviewed twice a year.

The age requirement was derived from several factors. In order to acquire a driver's license in Denmark, you have to have a valid first aid certificate, however these expire after two years. This means that as a lot of Danes acquire their driver's license at the age of 18, most of their certificates expires when they reach the age of 20. Furthermore, as the requirement was not introduced until 2006, people over the age of 28 did not receive first aid training in conjunction with acquire their driver's license. Alongside the age requirement, it was required that the participants lived in Denmark and spoke Danish.

The sampling method used was quota sampling (Bjørner, 2015*b*, p. 62). We also used snowball sampling to maximize our sample size, this was done by asking the participants if they knew

anyone who matched our target group. In order to avoid getting participants who did not fit the target group, the participants were told the predefined characteristics of the target group after their test session. A visualization of our sample size can be seen in figure 3.1.

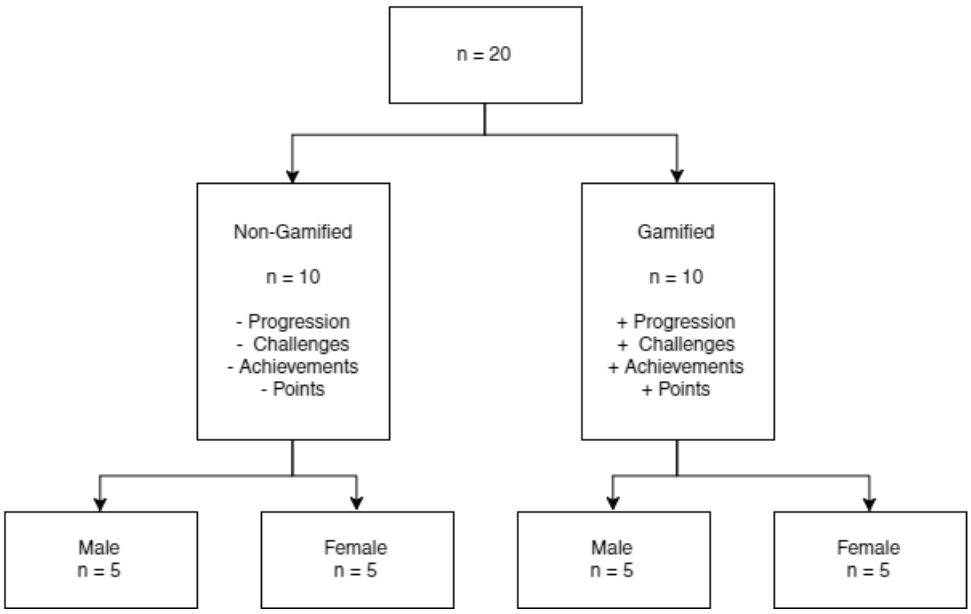


Figure 3.1: Visualisation of the sampling and the differences in the two applications used for the between-groups test design.

The only differences between the two groups are the ones that can be seen in figure 3.1. The two groups receive that same feedback from the buttons and the aesthetics are the same. Nothing in the interfaces has changed, and they have the same theme and colors.

3.2 Data Collection

3.2.1 Data Logging

The prototype logged total time, number of attempts for each scenario, how many times they answered correct and incorrect in each attempt, which pages they had visited, their progression and stars earned in the scenarios. The non-gamified version did not display a progression bar or stars, but the application still tracked this, even though the participants could not see it. Total time was compared with the participants' perceived time in order to detect if they were time distorted. The attempts used for each scenario and number of correct and wrong answers was used to see how well they performed. The progression, stars and number of pages visited showed how much of the application they explored. Figure 3.2 shows an example of a data log and what it contained.

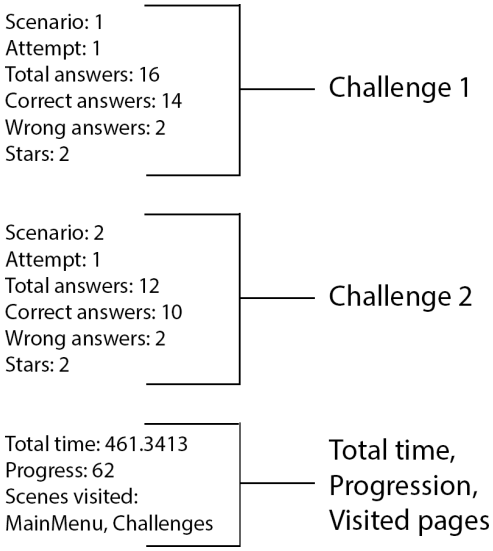


Figure 3.2: Example of how the data was logged.

3.2.2 Interviews

As we wanted to receive in-depth answers from the participants we decided to use interviews rather than questionnaires. The advantage of using an interview rather than a questionnaire is that it allowed the participants to verbally elaborate on their answers right after they have given their score. Additionally, the participants might have been more likely to sufficiently elaborate their answers as the interviewer used follow-up questions. The interviews were recorded in order to ensure that the participants own words were analysed rather than a transcription or notes from the interviewer.

The type of interviews used were in-depth interviews, as we wanted the participants to elaborate on how they experienced the prototype. In-depth interviews made it easier to go into detail about the questions they were asked. Focus groups could have been used, but they are intended to keep a discussion going between a few participants, which can become very broad, and it can be hard to get thorough answers to questions. Conformity and polarisation could also have influenced the interview with a focus group, as the participants could influence each other to feel more engaged than they really were (Bjørner, 2015*b*, p.76). In addition, as the participants were told they could use the application for however long they would like, it would require all the members of the focus group to finish at approximately the same time. However, by not using focus groups we could have missed out on valuable results due to the group dynamic that we would not otherwise find with a 1-on-1 interview.

We did not want to gather data in a too structured manner since we wanted to be able to ask related questions that come to mind during the interview. The interviews were therefore semi-structured, which meant that we could ask open and closed questions (Rogers, Sharp and Preece, 2011, pp. 228-230). This allowed us to follow a certain pattern for each interview (Bjørner,

2015*b*), while having the opportunity to ask related questions if necessary. A manuscript was used to ensure reliability in our interviews, which made sure that every participant was asked the same question in the same manner, furthermore it was the same interviewer throughout the interviews.

3.2.3 Verbal Numeric Rating Scale

During the interview we used a verbal numerical rating scale (VNRS), which is a projective technique used to get people to express their feelings or thoughts towards different subjects (Bjørner, 2015*b*). We chose to use a VNRS that went from 0 - 10, where 0 was the lowest score and 10 the highest. This means that the participants were asked to verbally rate elements from the application. As they were asked in an interview it was possible to ask the participants to elaborate their answer and why they chose e.g. 5 instead of 6 or 7 (Bjørner, 2015*b*). This means that we received numerical values that could be compared between the two groups, along with detailed information that the participants shared. However, self-assessment might yield the wrong results, as the participants might misjudge themselves and give a lower or higher value than the other participants.

3.2.4 Engagement

It seemed appealing to use a questionnaire for measuring engagement, since different questionnaires have already been created for the purpose. Examples can be made of the Game Engagement Questionnaire (GEQ), which was developed for measuring engagement in games (Brockmyer, Fox, Curtiss, McBroom, Burkhart and Pidruzny, 2009), and the User Engagement Scale (UES), which was developed for measuring engagement in online shopping websites (O'Brien and Toms, 2009). We used the GEQ and UES as inspiration for constructing questions for the VNRS since our evaluation consisted of an interview. Table 3.2 shows the questions asked during the interview.

Table 3.2: Translated version of the questions that were asked during the interview regarding engagement. Danish version can be found in Appendix A.

Attributes of engagement	**Questions**
Perceived Time	1 On a scale from 0 - 10, where 0 is not at all and 10 is very much, did you forget about your surroundings while using the application?
Aesthetics	2 On a scale from 0 - 10, where 0 is bad and 10 is good, what did you think of the visuals of the user interface? 3 On a scale from 0 - 10, where 0 is bad and 10 is good, how do you think the visuals presented what you expected? E.g. did you think the buttons looked like button?
Feedback	4 On a scale from 0 - 10, where 0 is bad and 10 is good, how do you think the sounds represented your choices in the application? 5 On a scale from 0 - 10, where 0 is bad and 10 is good, what did you think of the visual feedback received when pressing a button?
Challenge	6 On a scale from 0 - 10, where 0 is hard and 10 is easy, how difficult did you think the scenarios were? 7 On a scale from 0 - 10, where 0 is not at all and 10 is very much, how challenged did you feel while using the application?
Enjoyment	8 On a scale from 0 - 10, where 0 is easy and 10 is hard, what did you think of the navigation in the application? 9 On a scale from 0 - 10, where 0 is not fun and 10 is very fun, how did you feel while using the application?
Interest	10 On a scale from 0 - 10, where 0 is not interesting and 10 is very interesting, how interesting was it to learn about first aid through the application? 11 On a scale from 0 - 10, where 0 is not interested and 10 is very interested, how interested were you in using the application?
Goal	12 On a scale from 0 - 10, where 0 is not clear and 10 is very clear, how clear do you think the goal was with the scenarios?

The questions in table 3.2 are related to certain attributes of engagement from figure 2.2; aesthetics, feedback, challenge, enjoyment, interest, perceived time, and goal. The questions provided detailed information about the participants' experience, since they were asked to elaborate their rating for each question. We decided to limit the amounts of questions to avoid the participants getting fatigued during the interview. The order of which question was asked first was also taken into consideration. The first question was about perceived time and in order to get the most accurate estimate this was asked in the beginning.

3.2.5 Remember

The remember category (category 1) from the Taxonomy of Educational Objectives (Anderson et al., 2001) was used to further evaluate whether participants were engaged. The test participants received questions regarding certain parts of the application during the interview. The questions were related to the content and the user interface of the application, as we wanted to know if any of the two groups remembered elements better than the other. The questions increased in difficulty to see how much the participants remember from the applications. Table 3.3 shows the questions and their difficulty, where 1 is the easiest question.

Table 3.3: Translated version of the question about the application and its content. Danish version can be found in Appendix B.

Difficulty	**Question**
1	"What background color did the buttons have?"
2	"How many subcategories did the recovery position page have?"
3	"How many joints are there in the Chain of Survival?"
4	"How many unexpected out-of-hospital cardiac arrests happen in Denmark each year?"

3.2.6 Apply

The apply category (category 3) was also used. During the interview, the test participants received a piece of paper with a scenario, much like the ones in our application, where we then asked them three questions about the scenario. The "Learn CPR" option in the application contained instructions on what to do in different emergency situations. These questions were asked to test whether they knew what to do, which they should if they read and remembered the instructions in the "Learn CPR" option in the application, because we wanted to know whether or not they have learned something from the application.

Figure 3.3: The image of scenario 3 given to the participants.

The paper scenario was based on a drowning accident as none of the scenarios in the application included a drowning accident. Figure 3.3 the image the participants were shown. The questions asked for the paper scenario can be seen in table 3.4.

Table 3.4: Translated version of the question about the application and its content. Danish version can be found in Appendix C.

Question	**Answering Options**
You are alone and finds an unconscious person in the edge of the water, you pull the person out of the water. What do you do next?	1. Checking if the person is breathing 2. Run for help 3. Call 1-1-2
You check his breathing and finds that he is not. What do you do next?	1. Start giving CPR 2. Place the person in recovery position 3. Call 1-1-2
You call 1-1-2. What do you do next?	1. Start giving CPR 2. Wait for help to arrive 3. You leave him and look for help
You are about to start CPR, how many chest compressions and rescue breaths are correct?	1. 5 rescue breaths followed by regular CPR (30:2) 2. 3 rescue breaths followed by regular CPR (30:2) 3. 2 rescue breaths followed by regular CPR (30:2)

3.2.7 Observations

The participants were video recorded while they used the application, this allowed us to go back and rewatch the sessions, analyse them and interpret the data. The participants were placed in a semi-artificial environment, since we asked the participants to come and try the application at a predetermined spot (Bjørner, 2015*b*, p. 66). The data collected from the observations was the participant's facial expressions, looking at "how" engaged they were in the application. We were measuring enjoyment and challenge since these are factors for engagement (O'Brien and Toms, 2008; Hoffman and Nadelson, 2010; Csikszentmihalyi, 1993). Enjoyment can be detected by looking for signs of happiness, and that can be detected by looking for lip corner pulls, raising the cheeks and narrowing the eyes by raising the cheeks (Rozin and Cohen, 2003, p. 70). Concentration can be detected by looking for narrowing of the eyes, lowering and bringing together the eyebrows (Rozin and Cohen, 2003, p. 73).

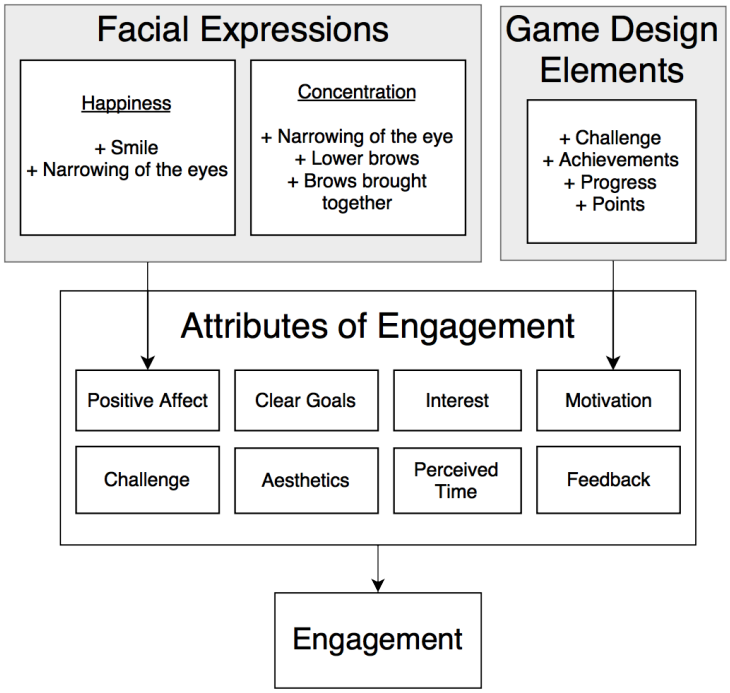


Figure 3.4: Visualization of enjoyment and concentration leading to engagement.

Rozin and Cohen (2003, p. 73) explain that concentration can easily be misinterpreted as confusion. Two rating scales were developed for observing enjoyment and concentration during the test. The rating scales were based on the Facial Action Coding System (FACS) (Cohn, Ambadar and Ekman, 2007). The two rating scales are shown in table 3.5 and 3.6. The relevant Action Units (AUs) from FACS can be found in Appendix D. The two scales have three different levels, where certain AUs have been used to determines the intensity of the facial expression. The letter after the number in each FACS code represents the intensity of that AU, where A is minimum intensity and E is maximum intensity.

Table 3.5: A representation of the table used by the observers to determine their "level" of enjoyment.

	1. Little happy	**2. Moderate happy**	**3. Very happy**
Exrpession definition	Small lip corner pull	Medium lip corner pull with mouth closed. Raising cheeks	High lip corner pull with mouth open. Narrowing the eyes by raising the cheeks
FACS Code	12A	6C, 12C	6E, 12E
Their expression			

Table 3.6: A representation of the table used by the observers to determine their "level" of concentration.

	1. Little concentrated	**2. Moderate concentrated**	**3. Very concentrated**
Exrpession definition	Little narrowing of eyes	Medium narrowing of eyes, head might be tilted forward	Maximum narrowing of the eyes, lowered eyebrows and eye brows brought together, head tilted forward, maybe hand near face
FACS Code	4A, 4B	4B, 6A, 57B	4C, 6C, 42D, 44D, 57C
Their expression			

Some ethical issues could have emerged when it was decided to record the participants, however they were asked to sign an informed consent form that included consent to video and audio record them during the test and interview (see Appendix E). This stated that their privacy and data will be kept confidential and will only be shared with the supervisor of the study and the examinator at the exam. Furthermore, as the participants were placed in front of the camera, they were well aware of them being video recorded.

3.2.8 Time

As the participants were asked to use the application for as long they would like to, the total amount of time was logged. The logged time was used based on the theory that the more you like an application, the more likely you are spending time using it. However, as the test

participants are placed in an artificial environment, they might feel inclined to spend more time on the application after they are disengaged and have lost interest in the application.

Additionally, the total amount of play time was compared to the participants' perceived time. Several studies that have researched user engagement have used perceived time as a measurement of how engaged the users are (O'Brien and Toms, 2009; Brockmyer et al., 2009). Even though none of the mentioned studies have any sources or data to defend their claims, we agree that if you are engaged in an activity your accuracy of perception of time might be reduced. Furthermore, (Csikszentmihalyi, 1993, p. 178) described that time distortion is a dimension of flow, where track of time is lost and seems to go by faster.

It was important that the test participants did not have access to any devices that could show the time both during and after the test as this could potentially invalidate the metric. To avoid this the participants were asked to leave watches, phones, etc. with the observer at the far end of the table.

3.2.9 Limitations

A reliability problem we had was that the location changed during some of the test sessions, otherwise every measurement method was exactly the same for all participants. Four out of the 20 interviews were conducted at a test conductors home in a noisier but more intimate test environment and at a later hour than the other 16.

Most of the participants were tested in a populated environment, which means that they might have gotten distracted by loud noises or people passing.

3.3 Data Handling and Analysis

Content analysis was used to find the frequency of terms and facial expressions in order to gain an overview of the data. Traditional coding (Bjørner, 2015*a*, pp. 98-99) was used to organize, categorise and analyse the data from the interviews and the recordings. While the interviews did not have predefined categories, the video recordings did, these can be found in table 3.5 and 3.6. Three coders were used to perform the content analysis on both the interview and the recordings.

The coders started with recognising the themes from the interviews individually, thereafter the data was coded and organised into several categories and subcategories. These were then collected, and a final decision was made on which were used. Table 3.7 presents the categories and subcategories that were used.

Table 3.7: Categories derived from the content analysis.

Main Categories	**Subcategories**
Forgot surroundings	
Simplistic	
Too textual	
Clear goal	
Fun	
Feedback	Positive Negative
Bias	Distractions Test environment
User interface	Positive Negative
Challenge	Appropriate Easy
Humorous	Positive Negative
Game Design Elements	Positive Negative
Aesthetics	Positive Negative
Interest	Positive Negative
Learning	Informative Good alternative
Lacking	Physical interaction Content Sound Interactivity

Percentage agreement was used to give an indication of the overall agreement between the coders. The percentage agreement was calculated using algorithm (3.1).

$$\frac{(Coder\ Reliablity_{C_1C_2}) + (Coder\ Reliability_{C_1C_2}) + (Coder\ Reliability_{C_1C_2})}{Numbers\ of\ Coder} \tag{3.1}$$

Krippendorff's Alpha (KALPHA), for both the interviews and the observations, was calculated

in order to check for intercoder reliability (Hayes and Krippendorff, 2007). The KALPHA should be between 0 and 1, where 0 means no reliability and 1 means perfect reliability (Hayes and Krippendorff, 2007). Data can be quite reliable if the KALPHA is above 0,8, however it should be above 0,67 for drawing uncertain conclusions (Hayes and Krippendorff, 2007). The average KALPHA for both groups can be seen in table 3.8, where all are below 0,67. Graphs over the KALPHA for both groups can be found in Appendix F and percentage agreement in Appendix G.

Table 3.8: Average KALPHA and average agreement percentage results for interviews and observations for both groups.

	Gamified	**Non-gamified**
Interviews KALPHA	0,537	0,546
Interviews Agreement %	82,4%	82,9%
Observations KALPHA	0,320	0,364

The numbers collected from the VNRS were compared to the participants elaboration, to see if we could get a better idea of how engaged they were. The elaborations were also used to confirm that the number they chose was the correct one. In addition, the ratings from the VNRS were compared to the observations, to see if their self-assessed level of enjoyment matched their facial expressions.

For analysing the video recordings, it was possible to either analyse them manually or use automatic software recognition. While using automatic software recognition would have been faster than manually analysing it, we decided against it due to factors such as camera sensor sensitivity against light could make the software unreliable, furthermore the angle that the camera recorded at was not optimal for achieving reliable results. However, the manual approach does require a considerable amount of training. Despite this we decided to manually code the observations, as we believe the angle would make the software too unreliable and that the human eye might be able to detect expressions the software could miss. Table 3.5 and 3.6 show the facial expressions that the coders were to observe, afterwards the KALPHA was calculated to find the intercoder reliability of the codes.

3.3.1 Depth of Visit

We have defined the term "depth of visit" to see how much the participants used the application. We believed that the participants who used the gamified version would spend more time and become more engaged compared to those who used the non-gamified version. To see if this was the case, the data logs were compared to the questions asked after the semi-structured interview.

The logged data contained information about how much of the application they had explored. Furthermore, it was cross-checked with the observations from the recordings, to make sure that they spent the time using the application.

To see if the group that used the gamified version was more likely to retry challenges in which they had not completed without making a mistake, the logged data also contained information about how many tries they had in each challenge and their result.

Chapter 4

Design & Implementation

It was decided that the application should consist of several and different scenarios, since two-thirds of the cardiac arrest occurs at home. Hence we have two scenarios with different scenery. In one of the scenarios there is a bystander which can help you find things, such as an Automated External Defibrillator (AED). In the second scenario the player is alone and can not receive help from any bystanders.

Our questions and knowledge is based on the European Resuscitation Council Guidelines for 2015 (Monsieurs et al., 2015), since the application was designed to be used by Danes. The visual theme of the application was determined to match the topic, therefore red and white colors were used for most of the text and buttons in the game, since these are generally colors associated with first aid. The background color was decided to be a neutral blue color, in order to avoid a conflict with the red buttons and make interactable objects more noticeable.

Humor was decided to be a useful factor because of our targeted audience being in the age group of young individuals (20-28 years of age). Agencies such as 'The Danish Road Safety Council' has previously made use of humor in their infomercial about how kids should wear protective headgear while riding a bike (Rådet for Sikker Trafik, 2016). Furthermore, Weinberger and Gulas (1992) support this and state that humor attracts attention and can be used to make a dull subject more interesting. Since first aid is a subject that can be quite tiresome to read about, we designed some of the answering options in the challenges to have a humours formulation. This should also help the users obtain a certain amount of enjoyment, which is a part of the positive effect (O'Brien & Toms, 2008; Sweetser & Wyeth, 2005).

Unity3D was used to develop the application. This was beneficial as Unity featured a full UI-system. Using this over hardcoding would allow us to avoid having to code an action listener for each button and use a "what you see is what you get" editor (WYSIWYG). All assets, including scripts, can be found in the "/assets/" folder in appendix H.

Acknowledgement to third party owners of intellectual property can be found in appendix I.

4.1 Game Design Elements

4.1.1 Rewards

Rewards were implemented in the application, since they give the user clear visual feedback based on performance, which can motivate the user to try the scenarios more than once (Marczewski, 2015; O'Brien and Toms, 2008; Sweetser and Wyeth, 2005). Furthermore the users were unable to continue to scenario 2, if they got less than two stars. The icon chosen to indicate points was decided to be a star, since many games or gamified systems use this as their reward icon (see figure 4.1).

Figure 4.1: Example of when a user has gotten two stars in a scenario. The earned stars are yellow, and unearned are grey.

It was discussed whether the point icon should be a heart or a first aid icon to follow the theme of the application, however it was decided against, since it could have confused the users.

Code 4.1 shows how the EndScenario.cs script determines how many stars are rewarded based on the amount of total answers the user had. Line 1 checks if the scenario is the scenario 1, if it is then it checks the amount of total answers. The condition in line 2 determines that the user can have no wrong answers in order to receive three stars, because it says that the user should have less than or equal to 14 total answers. If this number was set to 12 instead of 14, then this would mean that the user could get 3 stars even if he answered two questions incorrectly. Line 4 and 6 do the same for one and two stars, the only difference is the amount of answers that change.

```
1 if(player.scenarioNumber == 1)
2     if (scenario.totalAnswers <= 14)
3         scenario.Stars = 3;
4     else if (scenario.totalAnswers <= 18)
5         scenario.Stars = 2;
6     else if (scenario.totalAnswers <= 24)
7         scenario.Stars = 1;
8     else
9         scenario.Stars = 0;
```

Code 4.1: Snippet from EndScenario.cs - Code snippet of the reward showing system.

4.1.2 Achievements

Achievements can be a motivational factor, since it provides clear feedback when showing the user their progress, it also gives the users a reason to continue using the application, especially if they want to unlock as many as possible. The way we visualised the achievements can be seen in figure 4.3, where the user has unlocked two of the nine available. The unlocked achievements are highlighted with a bright yellow color to give a clear indication of the user's accomplishments. The user is shown a popup that enters the screen from the top, to give them instant feedback that the task they just did awarded them an achievement, figure 4.2 shows an example of the popup box that the user is shown.

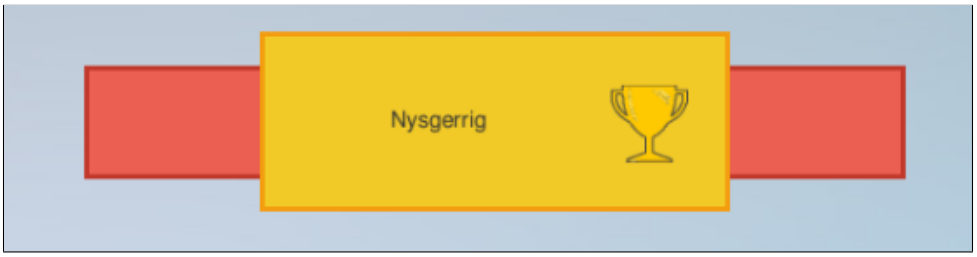


Figure 4.2: Popup the user is shown when entering the Facts page.

A trophy icon was used on the achievements page, since this is a well recognised icon symbolising an achievement. Examples of its usage can be seen on the Xbox One's (2013) achievement pop-up, Playstation's trophy system (2013) and in the game Peggle (2007). Furthermore, we added four stars to the icon when the user unlocked an achievement. This should hopefully further invite the users to feel that they have accomplished something. The stars also indicate a reference between the scenarios and achievements, which was designed to invite the users to try out the challenges.

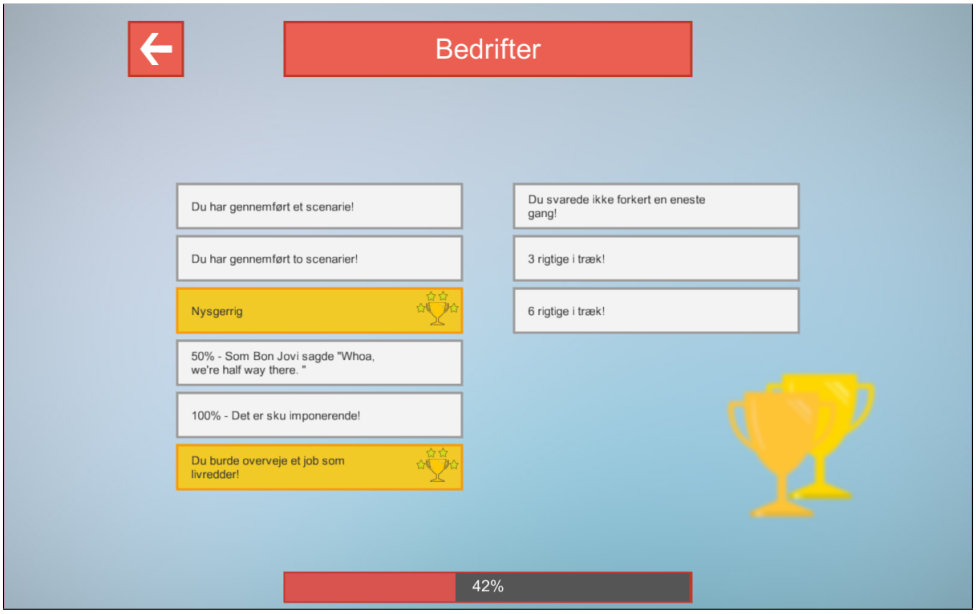


Figure 4.3: Image from the Achievements page, where the accomplished achievements are yellow and marked with a trophy.

The name of some of the achievements was designed to be humorous. Achievement 4 in 4.2 is an example of this choice. An example can be made of achievement 4 on line 4 in code 4.2. When the user reached 50% progress they would be awarded an achievement called "50% - Whoa, we're half way there.", which is a reference to Bon Jovi's song "Livin' On A Prayer". Achievement 6 in line 6 was unlocked when participants had opened all the subpages in the "Learn CPR" page, and it was called "You should consider a job as lifesaver!" (translated to English), which also was intented to be humorous.

```
1 ach1 = new Achievement {Name="Du har gennemført et scenarie!", isUnlocked =
      false};
2 ach2 = new Achievement {Name="Du har gennemført to scenarier!", isUnlocked =
      false};
3 ach3 = new Achievement {Name="Nysgerrig", isUnlocked = false};
4 ach4 = new Achievement {Name="50% - Som Bon Jovi sagde "Whoa, we're half way
      there. " ", isUnlocked = false};
5 ach5 = new Achievement {Name="100% - Det er sku imponerende!", isUnlocked =
      false};
6 ach6 = new Achievement {Name="Du burde overveje et job som livredder!",
      isUnlocked = false};
7 ach7 = new Achievement {Name="Du svarede ikke forkert en eneste gang!",
      isUnlocked = false};
8 ach8 = new Achievement {Name="3 rigtige i træk!", isUnlocked = false};
9 ach9 = new Achievement {Name="6 rigtige i træk!", isUnlocked = false};
```

Code 4.2: Snippet from Data.cs - Achievements are initialized.

4.1.3 Progress Bar

It was decided that the progress bar should be a horizontal bar that would fill up based on the user's progress, similar to how hit point bars and energy bars are usually displayed. Furthermore, the color of the progress bar was set to be red, since it would match the rest of the theme in the application. We chose red as this was more noticeable than white because most of the backgrounds are a neutral color close to white.

The progress bar itself is not a motivational factor, however as progress and clear visual feedback both are motivational factors (Marczewski, 2015; O'Brien and Toms, 2008; Sweetser and Wyeth, 2005), we decided to visualise it to the users to motivate them to use the entire application, rather than stop early. Figure 4.4 shows an example of the progress bar from the application filled 42%.

The shape of the progress bar is designed to be an icon, which should resemble something that can be filled up. Furthermore, the progression itself is an index, since an index is defined by something sensory. It also indicates a connection between the start of the progress bar and the end, since it flows towards the end of the progress bar. The percentage in the progress bar is a symbol, which indirectly indicates how much the user has progressed, and how much more he is able to progress (Port, 2000; Atkin, 2013).

Figure 4.4: Image of the progress bar from the gamified version.

Code 4.3 increments the progression when the user enters a new unique scene. The if-statement on line 4 checks if the current scene has been visited before, by checking an array of pages names that has been visited before. If the page is unique then line 5 increments the progression counter by the float in line 1 and line 6 adds the name of the page to the array, so it will not increment if the user visits this page again. If the value was set to 15f instead of 7f, then the progression would increase too rapidly with 15% for each page visited and the total progression would exceed 100%.

```
1 incrementValue = 7f;
2 sceneName = SceneManager.GetActiveScene().name;
3
4 if (!Data.Scenes.Any(str => str.Contains(sceneName))) {
5     Data.Progression += incrementValue;
6     Data.Scenes.Add(sceneName);
7 }
```

Code 4.3: Snippet from IncrementProgression.cs - Code that increments the progress bar when a new unique scene has been visited.

Similar to code 4.3, code 4.4 also increments the progression counter. However, this was based on how many stars they received in both of the two scenarios. The increment value is set to approximately 8.46f because the increment value for the pages is set to 7f. These two values have to be matched in order to make sure that the max progression is at 100%. Code snippet 4.4 is for the first star in the first scenario. Line 1 checks if the user has tried the scenario before, if they have then the next if-statement will check whether there is a one or more stars and whether progress for this star in this scenario has not been incremented before. If the if-statement returns true then it will increment the progress and change the boolean on line 4 to true, so it will not increment again, if they retry the scenario and receive 1 star again.

```
1 if (Data.ScenarioWithMostStars(1).Stars > 0) {
2     if (Data.ScenarioWithMostStars(1).Stars >= 1 && !Data.scenario1ProgStars[0])
          {
3         Data.Progression += 8.466667f;
4         Data.scenario1ProgStars[0] = true;
5     }
```

Code 4.4: Snippet from Progres.cs - Increments the progress if they have received 1 star in scenario 1.

4.2 Scenarios

It was decided to include challenges to the design as interactivity is a part of engagement (O'Brien and Toms, 2008). Questions with a single answering option should not be used, as this would be too easy, which was discussed in the Analysis chapter. Additionally, the users should not feel that they were under time pressure, as we wanted them to take their time to analyse the scene and apply the correct procedures. The application featured two scenarios during the testing phase; one in public and one at home. In the gamified version the second scenario was locked until the users have completed the first one and gotten at least two stars. This was to give the users clear feedback on how they performed in the scenarios and to motivate them to play through the scenario again, if they had not gotten three stars in it. Figure 4.5 and figure 4.6 show the second scenario being locked and unlocked respectively.

Figure 4.5: Image from the "Challenges" page, where no stars have been earned in the first scenario.

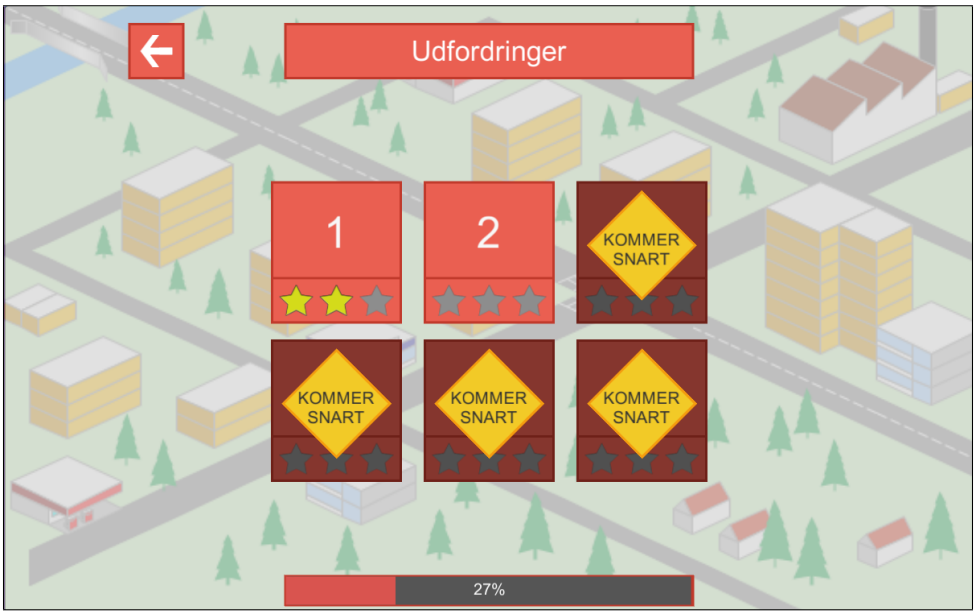


Figure 4.6: Image from the "Challenges" page, where two stars have been earned in the first scenario, and the second scenario gets unlocked.

Both scenarios were controlled by a single script called playerController.cs. This controlled the general mechanics for the scenarios such as disabling and enabling answering options, moving game objects, moving and zooming the camera and deciding how many wrong answers the users were allowed to have for reaching 3 stars. Furthermore, it was also the script that would show a dialog box if the users answered wrong and if it was the last question in the scenario. Initially, it

was meant to work as a framework, which could allow us to easily setup more scenarios, however it was quickly based on temporary work that ultimately ended up working, but did not adhere to basic software engineering rules such as low coupling, high cohesion, and polymorphism. Figure 4.7 shows the playerController.cs script in the Unity inspector.

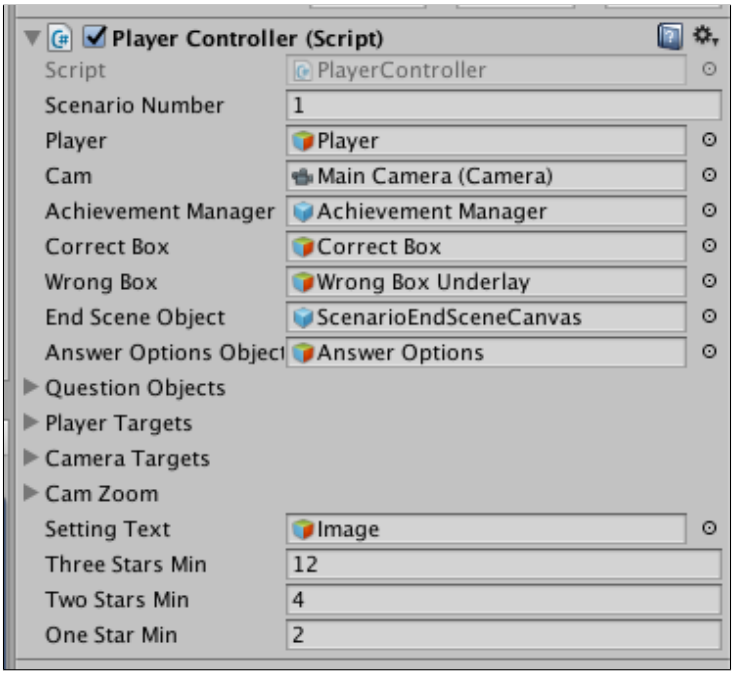


Figure 4.7: playerController.cs in the Unity inspector panel.

If the user pressed the correct answer a method in playerController.cs named startNextQuestion() would run. This method checked which scenario the user was in and which question the user was on. Based on these it decided what to show the user next. Code 4.5 shows the first part of the condition statement that disabled the back button and the setting text in the top of the video (line 2-3). Then it starts a coroutine called queAmbulance, however this will be explained later. If the question was 1 over the amount of questions in the scenario, it would end the scenario and show the user a dialog of how well they performed, and if they played the gamified version it would also show them their earned stars.

```
1  if (question == noOfQuestions) {
2      GameObject.Find("BackButton").gameObject.SetActive(false);
3      GameObject.Find("Scenario Text").gameObject.SetActive(false);
4      StartCoroutine("queAmbulance");
5      question++;
6  } else if (question == noOfQuestions + 1) {
7      endOfScenario = true;
8      EndSceneObject.gameObject.SetActive(true);
9
10     ...
```

Code 4.5: Snippet from playerController.cs - Conditional statements that check which number of the question the user is on.

Figure 4.8 shows an example of the end scene that the gamified users were shown. It shows them their stars earned, the total amount of tries, how many correct answers and how many wrong answers they had. The non-gamified would be shown the exact same, except their version would not show the stars. Furthermore, this scene also gave them to option to retry the scenario or go back to the menu.

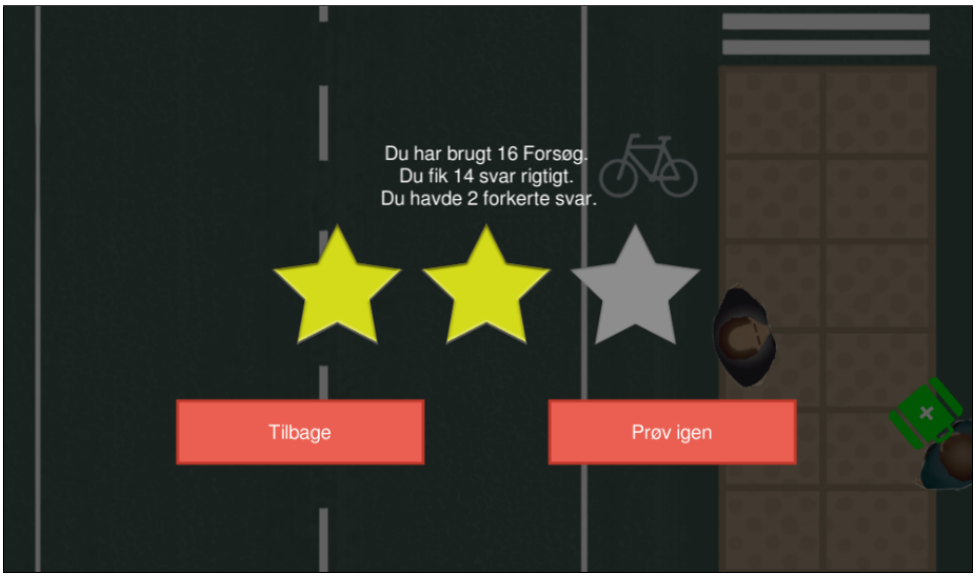


Figure 4.8: End scene shown to the users using the gamified version.

If neither of the conditional statements in code 4.5 were true, then the code would check if the question was equal to or smaller than the amount of questions. This could only be true if there were still questions the user yet had to receive. If this was the case then the code would start a coroutine at line 5 that handled all the player movements and how the camera was positioned and zoomed. After the next question had been initiated, either of the else if-statements on line 7, 11 or 15 would run depending on which question the user was on. These conditional statements all controlled the movement of other objects through the coroutine moveObject(GameObject, Target). The three conditional statements together control how the bystander in scenario 1 runs off to find an AED.

```
...


} else if(question <= noOfQuestions) {
    StartCoroutine("nextQuestion");
    if (question == 0) {
    } else if (question == 2) {
        s1RotatePlayer = true;
        GameObject AEDTarget1 = GameObject.Find("AEDPerson Target 1");
        StartCoroutine(moveObject(GameObject.Find("AEDPerson"), AEDTarget1, 4.0f
            ));
    } else if(question == 8) {
        GameObject AEDTarget2 = GameObject.Find("AEDPerson Target 2");
        StartCoroutine(moveObject(GameObject.Find("AEDPerson"), AEDTarget2, 1.0f
            ));
        getAEDRotation = true;
    } else if(question == 10) {
        GameObject AEDTarget3 = GameObject.Find("AEDPerson Target 3");
        getAEDRotation = false;
        StartCoroutine(moveObject(GameObject.Find("AEDPerson"), AEDTarget3, 1.0f
            ));
    }
}
```

Code 4.6: Snippet from playerController.cs - Executed code if the number of the question is smaller than or equals to number of questions

4.2.1 Scenario 1 - Public

Scenario 1 was decided to be in a public space, since it is good to know how to act when there are other people around, who can use to make the process more sufficient. Furthermore, we wanted a scenario where we could implement an AED, since it is good to know what to do when you have one at your disposal. As they are placed around public spaces, it was decided that it should be in a public space. Therefore, nearly half the questions for scenario 1 were about how to use an AED and what to do if you have people who can help you.

4.2.2 Scenario 2 - Home

It was decided to create scenario 2 in a different environment as that would add variety for the users. The scenario was designed as an home environment, since 72.3% of out-of-hospital cardiac arrests occur in private homes (Wissenberg et al., 2013). The users had to be challenged by getting different questions under different circumstances compared to those in the previous scenario. Questions that were not considered as basic CPR were therefore implemented, examples of such questions were regarding alternative rescue breaths (mouth-to-nose) and psychological first aid.

4.3 Audible Feedback

Sounds were implemented to give the users feedback on their actions and improve usability. When the participants clicked a button and got an answer correct or wrong a sound was played. Additionally, the ambulance in scenario 1 had both engine and break sounds added. Neither of the sounds were synthesized by us (see Appendix I for references). The button played a "click" sounds as feedback. It seemed intuitive to use a simple click sound since this is often used by both mobile and desktop applications. The feedback for when participants answered correctly was a sound that was supposed to revoke positive feelings in the participants (see appendix J) and a sound with lower pitch was used for when they answered incorrect (see appendix K). Two sounds were used for the ambulance, one was just an engine sound and the other was a sound of wheels screeching. The purpose was to add some humour to when the ambulance stopped, by making it sound like it was driving very fast and then braking suddenly.

4.4 Data

The data regarding the challenges, progression, rewards, achievements and time were all stored in the Data.cs script. Line 3 and 4 in code 4.7 are variables declared for time and progression. Line 6 and 7 create two lists of type Scenario, which are used to store data for each scenario the participant tries.

```
1  public static string ParticipantsName { get; set; }
2
3  public static float TotalTime { get; set; }
4  public static float Progression { get; set; }
5  ...
6  public static List<Scenario> scenario_1 = new List<Scenario>();
7  public static List<Scenario> scenario_2 = new List<Scenario>();
8  public static List<Achievement> achievements = new List<Achievement>();
9  ...
10 ...
11 ...
12 public static bool TimeAndProgressSaved { get; set; }
13
14 public static Achievement ach1, ach2, ach3, ach4, ach5, ach6, ach7, ach8, ach9;
```

Code 4.7: Snippet from Data.cs - The Data.cs script contained all the data that was relevant to what the user had achieved.

The reason for using a list instead of an array is because arrays have to be a fixed size, while lists can grow dynamically, and it had to be a collection of the Scenario class. We did not know how many times the participants would try each scenario, so it would be impossible to create an array with a fixed size. The class for a Scenario is seen in code 4.8, which contained properties for correct and wrong answers, along with total answers and stars.

```
1 public class Scenario {
2     public int wrongAnswers { get; set; }
3     public int correctAnswers { get; set; }
4     public int totalAnswers { get; set; }
5     public int Stars { get; set; }
6 }
```

Code 4.8: Snippet from Scenario.cs - Scenario class with its properties.

Line 8 in code 4.7 creates a list of type Achievement, which was used to store all achievements. This had to be a list, since it had to be a collection of the Achievement class, just like the two scenario lists. The Achievement class is seen in code 4.9, which only had two properties; a string for the name and a boolean to check if the achievement has been unlocked.

```
1 public class Achievement {
2         public string Name { get; set; }
3         public bool isUnlocked { get; set; }
4 }
```

Code 4.9: Snippet from Achievement.cs - Achievement class with its properties.

Logging the data

Data had to be logged while participants were using the application. A file was automatically created for each participant, where all the data for the particular participants test session was saved. The file was saved locally on the tablet.

Scenario data such as total answers, correct answers etc. was saved when participants either finished a scenario or pressed the "Back" button in the top left to exit a scenario. The SaveCurrentData() method seen in code 4.10 is responsible for saving the scenario data to a file. Line 8 checks if the file for the current participant has been created already, if not, then it will create a string array with the current scenario data and on line 11 create the file with the data. If the file already had been created, then the else-statement on line 12 would run, which creates a string with the current scenario data and appends it to the already existing file. Appending the text made sure that data was never overwritten, but just added after the last line in the file. The three dots that keep appearing in line 9 are for simplicity and they represent "Environment.NewLine", which just makes a line break in the file. Using the NewLine command ensured that a new line was added no matter what operating system the application were executed on.

```
1  public void SaveCurrentData(){
2     string path = Application.persistentDataPath + Path.DirectorySeparatorChar +
          Data.ParticipantsName + ".txt";
3
4     if (scenarioNumber == 1)
```

```
5       {
6           lastScenario = Data.scenario_1.Count - 1;
7
8           if (!File.Exists(path)){
9               string[] createText = { "Scenario:" + scenarioNumber.ToString() +
                    ... + "Attempt:" + (lastScenario + 1).ToString() + ... + "Total
                    answers:" + Data.scenario_1[lastScenario].totalAnswers.ToString
                    () + ... + "Correct answers:" + Data.scenario_1[lastScenario].
                    correctAnswers.ToString() + ... + "Wrong answers:" + Data.
                    scenario_1[lastScenario].wrongAnswers.ToString() + ... + "Stars:
                    " + Data.scenario_1[lastScenario].Stars.ToString() + ... };
10
11              File.WriteAllLines(path, createText);
12          } else {
13              ...
14              File.AppendAllText(path, appendText);
15          }
16      }
```

Code 4.10: Snippet from PlayerController.cs - SaveCurrentData method for saving data from the current scenario.

Chapter 5

Findings

All gathered data from the test can be found in the appendices. The interviews for both groups can be found in Appendix L, recordings from the test in Appendix M and data logs in Appendix N. Data on demographics for the gamified group can be found in Appendix O and non-gamified group in Appendix P. Full VNRS results can be found in Appendix Q for the gamified group and Appendix R for the non-gamified group. The coder ratings for the interviews can be found in Appendix S, and the observations in Appendix T.

5.1 Reliability

During the interviews the participants were asked how much of their surroundings they forgot (question 2, see table 3.2). Sixteen of the test participants were tested in a noisy environment at Aalborg University Copenhagen and the last four in a smaller but equally noisy environment at a test conductor's home. Ten participants, five from each group, mentioned this as a disturbance that made them less likely to forget their surroundings. Most of them reported general background noise as people talking and walking by. However, NG6 mentioned that a loud whistle was playing during the test (Appendix L, NG6, 00:36). The whistle was also present while G9 was testing (Appendix M, G9, 02:49), who also reported noise as a disturbance. The ten participants who reported disturbances had an average rating of 6,1 for the first question, whereas the other ten participants who did not mention disturbances rated it 7,8 on average. Out of the ten participants who did not report any disturbances, five from the gamified group had an average of 7,2 compared to the five from the non-gamified group who had an average of 8,4.

Most of the participants were both students at Aalborg University Copenhagen and acquaintances of at least one of the test conductors. Participant G5 expressed that he knew it was not a professionally made application and that he did not have a lot of expectations for it (Appendix L, NG5, 06:28). NG2 expressed that he thought he knew what was being evaluated, and

stated that he was looking for distractions throughout the test, since he expected that the test conductors would test if he noticed any distractions caused by them.

5.2 Engagement

The participants' ratings for each of the questions asked in the interviews were averaged and can be seen in figure 5.1. The questions used for the VNRS and their numbers can be found in table 3.2 on page 14.

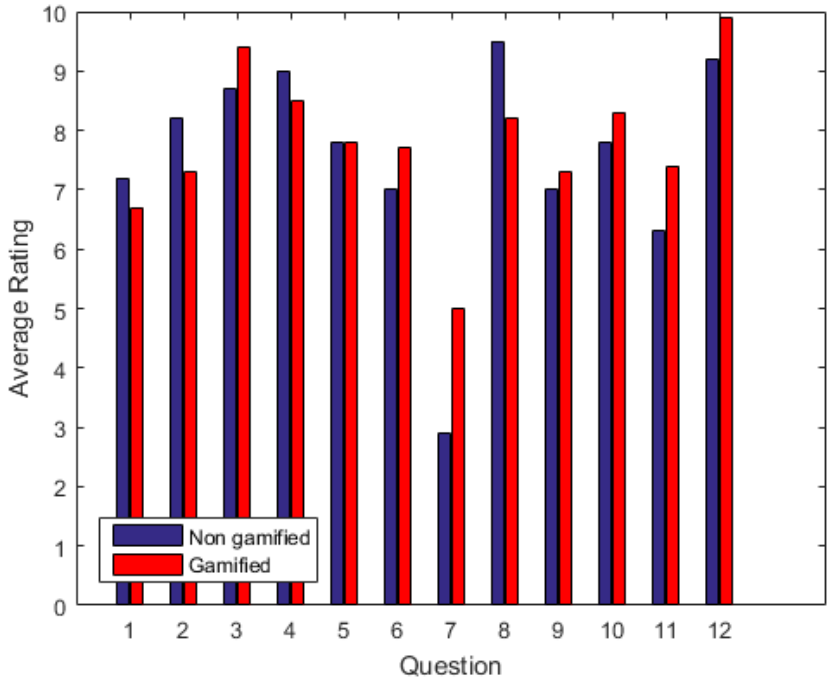


Figure 5.1: Average VNRS ratings for each question in the two groups (Non-gamified and gamified).

Figure 5.1 shows that the average ratings are similar across the two groups, however question 7 did have a relative high difference. Question 7 asked the users to rate how challenging the application was. The question was rated on average to 5 for the gamified group and 2.9 for the non-gamified group. This means that the gamified group felt more challenged than the non-gamified. The gap can be explained by the fact that the application lacked a challenging aspect, which means that the challenge attribute might not have been the reason for them being engaged (Sweetser and Wyeth, 2005; O'Brien and Toms, 2008). Some participants in the gamified group stated that the achievements and points were a challenge for them to obtain, which may be the reason they felt more challenged.

5.2.1 Interest

On average question 10 was rated 8,3 by the gamified group and 7,8 by the non-gamified, however a few thought that there could have been videos instead of the text and illustrations, some were also missing physical activation (Appendix L, G2, 05:30; NG4, 08:00; NG8, 08:10). The logged data shows that the gamified group played scenario 1 more frequently compared to the non-gamified group. The results can be seen in table 5.1.

Table 5.1: Average number of tries in each scenario for both groups.

Groups	**Scenario 1**	**Scenario 2**
Gamified	1,4	0,7
Non-gamified	1	0.8

The reason the gamified group tried scenario 1 more often than the non-gamified group might be because of interest, since their average rating on question 10 was higher. Users can become drawn to an activity by interest (O'Brien and Toms, 2008; Csikszentmihalyi, 1993, p. 188), and interest can keep them engaged (O'Brien and Toms, 2008). However, scenario 2 was seldom retried by either group, which can be because of the similarity of the scenarios, since novelty affects interest (O'Brien and Toms, 2008), and scenario 2 might not have contained enough new content.

5.2.2 Challenge

In question 6, the averages were 7,7 for the gamified and 7 for the non-gamified, which means that the difficulty level of the scenarios was too low. Several of the participants stated that due to the diversity of the answering options and their prior knowledge it was easy to eliminate the wrong answers, likewise they also stated that the questions were leading (Appendix L, G2, 03:55; G4, 02:40; G8, 03:55; NG3, 04:30; NG9, 03:30).

The average amount of concentration in the three defined categories can be seen in table 5.2 where the amount of concentration was rated by the coders. The gamified group was reported to be significantly more concentrated, since the difference between the two groups was 1,37 in "2. Moderately concentrated" and 0,74 in "3. Very concentrated".

Table 5.2: Average level of concentration rated by the coders.

Level of concentration	**Gamified**	**Non-gamified**	**Difference**
1. Not or little concentrated	2,93	2,48	0,45
2. Moderately concentrated	2,30	0,93	1,37
3. Very concentrated	1	0,26	0,74

The difference in the average level of concentration between the two groups could be due to motivation and the achievements, since the non-gamified group had no external rewards, where the gamified did. If a user is not enjoyed or there are no external rewards, then the user will stop the activity (Csikszentmihalyi, 1993, 188). G2 stated *" ... the progression bar kept me going ... "* (Appendix L, G2, 06:15) and G1 stated *" ... I liked and felt motivated by the achievements, and I wanted to unlock all of them ... I wanted to know how I could get all the achievements."* (Appendix L, G1, 02:30).

5.2.3 Perceived Time

We assumed that the participants who used the gamified version would spend more time on the application compared to those who used the non-gamified version. However, the result was completely opposite. The group that used the gamified version had an average play time of 564 seconds and the non-gamified had 619,5 seconds, which can be seen in table 5.3.

The level of measurement was ratio as the data was time. An Anderson-Darling test was used to check if the data followed a normal distribution. The gamified group returned an h-value of 1, which meant we accepted the alternative-hypothesis that the data did not follow a normal distribution. The data was not normally distributed and non-parametric, therefore a Wilcoxon rank sum test was computed on the percentage difference in played time and perceived time within the two groups to see if there was a significant difference between the two groups' time distortion. See Appendix U for the MATLAB computations.

Table 5.3: Play time and perceived time for each group.

Group	**Play Time (seconds)**	**Perceived Time (seconds)**
Gamified	564	654
Non-gamified	619,5	648

The Wilcoxon rank sum test returned a p-value of 0,2123. The null-hypothesis could not be rejected as the p-value was well above the significance level of 0,05. This meant that there was

no significant difference between the percentage difference in played time and perceived time between the two groups.

G1 showed signs of engagement, since the coders gave her several signs of enjoyment and concentration, likewise she also rated 8 to question 1. Furthermore, she stated *" ... I liked and felt motivated by the achievements, and I wanted to unlock all of them ... I wanted to know how I could get all the achievements"* (Appendix L, G1, 02:30), which shows that she was motivated to keep using the application even though her perception of time was not distorted. The difference between her time and perceived time was only 27 seconds, which is a minor difference, and indicates that she did not lose track of time. This could mean that the perceived time attribute mentioned in figure 2.1 is not always a requirement for engagement, since G1 was deemed to be engaged based on all the other attributes.

The small amount of difference between the two groups' perceived time and play time could be due to the participants not being in flow, motivated by the motivational factors, or be completely engaged (Csikszentmihalyi, 1993; Marczewski, 2015; O'Brien and Toms, 2008). G7 stated several things affecting her experience, this being disturbing background noises, deceiving pictures and text placement issues (Appendix L, G7, 00:40, 01:00, 05:00). Her facial expression remained almost the same throughout the test, which also could be an indication of her level of enjoyment being small. See figure 5.2 for a frame of G7's facial expression.

Figure 5.2: Picture of participant G7's general facial expression throughout the test

These factors could have had a negative effect on her perceived time, not making her forget the surroundings and therefore not lose track of time, which could be the reason for her perceived time only being 5 minutes, not giving the correct presentation of the two groups perceived time shown in table 5.3.

5.2.4 Positive Affect

The gamified group's averaged score was 7,3 while the non-gamified group averaged score was 6,8 to question 11. Several of the participants also stated that it was a fun experience (Appendix L, G1, 05:40; G3, 04:55; G6, 04:00; NG4, 06:50; NG5, 03:45; NG8, 07:30; NG10, 04:25). Furthermore, only two participants did not smile through the usage of the application. Table 5.4 shows the coders' ratings of the participants' enjoyment. Even though the observations show that there is a difference in how many signs the participants showed of happiness, the two groups, it very small. Based on the observations its hard to say if either of the two groups experienced a positive affect from the application.

Table 5.4: Average number of signs of happiness rated by the coders.

	Gamified	**Non-gamified**	**Difference (N-NG)**
1. Little happy	2,36	2,17	0,21
2. Moderate happy	2,2	2,2	0
3. Very happy	0,43	0,93	0,50

The participants that reported the highest amount of perceived time were G3 and NG3. They thought they had used 30 minutes on the application and said that the application was fun but serious (Appendix L, G3, 04:55), had good basics (Appendix L, NG3, 05:50), was funnier than traditional learning of CPR, encaptivating (Appendix L, G3, 05:25, 06:00), and that it was game-like (Appendix L, NG3, 07:15).

One of the participants who was coded to not have smiled throughout the test was G2. Figure 5.3 shows his general facial expression throughout the entire session and after the test he only rated 5 in question 9. In his elaboration he stated that it was not a fun application as it was used to learn. However, he went on to state that the learning aspect *" ... was done in a humorous way."* (Appendix L, G2, 04:52). Which means that despite his lower rating of question 9 compared to the groups average (7.3), he might still have enjoyed the application. However, this could also support what was stated in the Methods chapter about self-assessment, since his rating was different from many of the other participants' ratings.

Figure 5.3: Snippet from G2 recordings at 01:28.

NG5 also showed no signs of enjoyment, figure 5.4 shows her general expression throughout her usage of the application. Based on her rating from question 9 she found the application very fun as she rated it 9. However, her answer was very hesitant (Appendix L, NG5, 03:34). Her elaboration was also hesitant and argued that the illustrations could have been more entertaining had they been animations (Appendix L, NG5, 03:40).

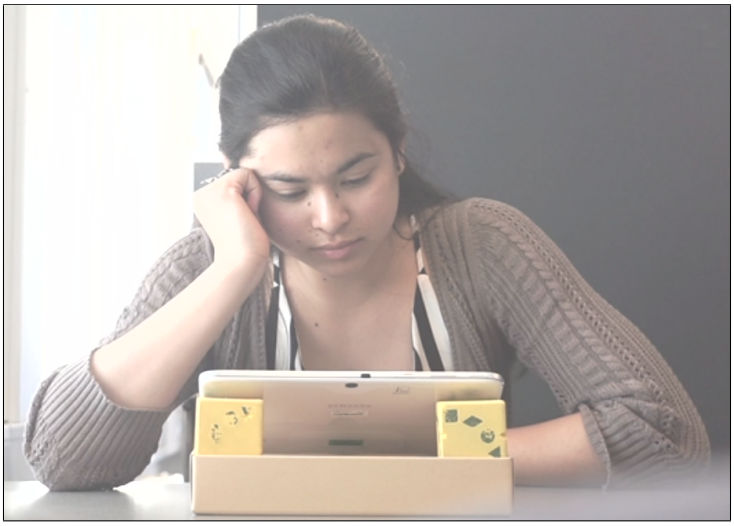


Figure 5.4: Snippet from NG5 recordings at 02:05.

5.2.5 Motivation

NG3 and G3 tried the first scenario twice while G5 and G6 tried it three times. Three of them said the scenarios were easy (Appendix L, G3, 04:00; G5, 08:45; G6, 07:45). G1, G2 and G3 mentioned that they wanted to unlock all the achievements (Appendix L, G1, 02:30; G2, 06:15; G3, 00:30), and G3 was the one in the gamified group that used the application longest, where his perceived time also was the highest. He said *" ... since there were achievements, I was very interested in it "* (Appendix L, G3, 00:30). This means that the achievements could have been extrinsically motivating for him (Marczewski, 2015). His actual play time was 13 minutes and 5 seconds and he thought that he used the application for 30 minutes. The big difference between the play time and perceived time means his sense of time was altered, and therefore could be in the state of flow (Csikszentmihalyi, 1993). But Csikszentmihalyi (1993) states that people should be involved in an intrinsically motivated task in order to be in flow, which could mean that G3's motivation went from being extrinsic (from the achievements) to intrinsic (from being challenged) (Marczewski, 2015).

The non-gamified group had an average of 78,9% progression and gamified had 77,4%. The gamified group's progression was lower than expected, however all of the gamified group's participants had visited all the pages in the application, where two participants of the non-gamified group missed a few. That means that the non-gamified group performed better in the scenarios, since their average progression was higher.

5.2.6 Feedback, Aesthetics & Clear Goal

Question 4 regarding sound feedback had an average rating of 9 in the non-gamified group and 8,5 in the gamified group. Six participants did not notice the sound feedback from the buttons, but they rated them anyway (NG2, NG4, NG6, G1, G2, G5). G2 stated that *"It's like you've achieved something when you answer correctly ... but it wasn't something I considered."* (Appendix L, G2, 02:29). The visual feedback from the buttons was rated 7,8 (question 5) in average for both groups even when six participants did not remember any visual feedback from the buttons (NG3, NG5, NG7, NG9, G3, G7). Question 5 was specifically regarding the visual feedback from the button, but G6, G7 and G10 took some visual illustrations into account when rating the feedback, because they thought some of them were confusing. Three mentioned that the visual illustrations should have been animated (Appendix L, NG5, 05:10; G4, 05:34 ;G10, 0:55).

In question 12, nearly every participant stated the goal with the scenarios was very clear, with the exception of a few. NG7 stated she was confused by the humor which suddenly occurred (Appendix L, NG7, 07:45). She was uncertain whether the goal was to compare your score with your friends or if it actually was an application to test your level of knowledge. However, the non-gamified group had an average rating of 9,2 while the gamified group's average was 9,9.

5.2.7 Depth of Visit

As stated earlier we thought that the more engaged participants were, the more pages they would visit and the more they would remember from the application. Eighty percent of the non-gamified group visited all possible pages, where the participants in the gamified group visited all the pages. Two of the participants (G9 and NG5) did not try any scenarios, despite NG5 having access to both scenarios as they were unlocked for her. In addition, G9 and NG5 used the application for 353 seconds and 264 seconds respectively compared to the gamified average of 619,5 seconds and the non-gamified average of 564 seconds. Furthermore, G9 showed very little enjoyment where NG5 showed no enjoyment at all. Even though they did not spend a lot of time with the application and their visits were not as deep as hoped, both participants still managed to answer correctly on half of the questions about the application after the interview, which is on par with the average for both groups. Furthermore, NG5 did not visit all of the possible pages, this combined with her short play time, the data suggests that she lacked interest for the application and thus was never engaged in it. G9 did visit all the scenes, but never tried a scenario, the reason for this could be poor usability. However, as only two out of 20 participants, we assume that this is not the case.

After the interview they were given a scenario on paper regarding a drowning accident. As the questions in this scenario were similar to those in the application, they were also too easy. This was proved by the fact that out of 80 asked questions (4 per participant) only 14 questions were answered incorrect. The amount of correct answers for each group can be found in table 5.5 and 5.6.

Table 5.5: Amount of correct answers answered in the questions about the scenario in each group and the difference between the two.

	Gamified	**Non-gamified**	**Difference (N-NG)**
Question 1	9	10	-1
Question 2	8	8	0
Question 3	10	9	1
Question 4	6	7	-1

Table 5.6: Amount of correct answers answered in the questions about the interface in each group and the difference between the two.

	Gamified	**Non-gamified**	**Difference (N-NG)**
Question 1	4	7	-3
Question 2	6	6	0
Question 3	3	7	-4
Question 4	0	2	-2

Chapter 6

Discussion

The topic, first aid, could have had an effect on the participants' interest, as it was stated by some of them that they knew it was an important topic, but felt that they did not need it, since they believed that they would never be in a situation where it was needed. Furthermore, interest can be disrupted if the user does not find the topic or activity appealing (O'Brien and Toms, 2008), which might have happened since the average of how many times the two groups tried scenario 2 was 0,8 times in the non-gamified and 0,7 times in the gamified. The lack of interest in scenario 2 could be explained by the amount of challenge they reported, where both of the groups indicated that there was a lack of challenge, even if they only tried scenario 1.

The scenarios should have been more challenging, nearly everyone who tried the scenarios scored a high amount of correct answers in their first try. The lack of challenge in the scenarios might have affected motivation, interest and perceived time. If the users do not feel challenged there is less of a chance that the users would be engaged in the activity (O'Brien and Toms, 2008; Hoffman and Nadelson, 2010; Pausch et al., 1994). Therefore, the interest could also be affected by the lack of challenge in the scenarios as they might become bored and lose interest in the application (O'Brien and Toms, 2008). Furthermore, the lack of challenge could also have had an affect on their perceived time, if you are having fun and enjoy an activity your time perception might become distorted (O'Brien and Toms, 2008; Csikszentmihalyi, 1993). However, there seem to have been a small difference on the amount of concentration between the two groups, which could mean that the gamified group was more concentrated because they felt that they had to unlock the achievements. This could lead to them being more motivated, not by interest, but by the game design elements that were implemented in the gamified version (Marczewski, 2015) as several participants stated that the achievements and points were motivational factors for them.

During member checking (Bjørner, 2015*a*, p. 109) some participants were asked when they were most concentrated, they stated that it was while they were reading about CPR and not, as we assumed, in the scenarios. To properly see if they were concentrated during the scenarios, we

should have logged time spent on each page, so this could be compared to the observations. This would allow the coders to distinguish between concentration while reading and concentration while being challenged.

To increase the challenge in the scenarios, the answering options should have been more alike. Some participants stated that having two or more options look similar would make it harder. For example, two answering options about how to perform rescue breaths; one through the nose and one through the mouth. While both could be correct, only one of them would be the right answer. This could have caused the participants to feel more challenged as their skill level might have matched the difficulty level better and in return avoid them being bored and affect their perceived time.

The progress bar provided the gamified group with feedback on their progression, which means they knew exactly when to stop using the application, since the it would display 100%. Participant G3 looked at the test conductor while proudly proclaiming "100% easy!" (Appendix M, G3, 12:54) when he finished. As the non-gamified group did not receive any feedback on their progression we believe that they kept playing even after they had completed all scenarios and explored the entire application.

We also assumed that if a participant in the gamified group did not receive three stars in a scenario, it would be more likely that they would retry the scenario, since each star affected the progression by 8,46%. However, as the difficulty of the scenarios was too easy, many of the participants got three stars on their first try. If the scenarios were harder they might have been motivated to retry them until they reached 100% since a matched player skill and challenge affects engagement (Csikszentmihalyi, 1993; Sweetser and Wyeth, 2005; O'Brien and Toms, 2008) and therefore spend more time on the application.

Furthermore, the application should have featured more content, as most of the participants progression were equal to each other. Fourteen out of the 20 participants' progress was 100%, 91% and 81%. More content would mean that the progress percentages would be more diverse, which would make it easier to see how far each individual participant got.

The observations showed little to no difference between the two groups' enjoyment, however several of the participants stated that they thought it was fun, and the gamified group rated 7,3 and the non-gamified group 7 in average in question 9. G2 stated that he had fun when using the application, however the coders did not assign any signs of enjoyment to him. These opposite findings could suggest that there were flaws within the pre-defined coding scheme for the facial expressions, since people do not express feelings the same way, he might as well have been more enjoyed than some of the other participants. A few days after the test, member checking was used to validate the data (Bjørner, 2015*a*, p. 109). He was asked if he enjoyed using the application, he answered yes and was surprised when told he did not show any signs of enjoyment during the test.

The differences in how much time the groups spent on the application could have been affected

by their rated enjoyment factor, since if you are enjoying a task, your perceived time could be affected (O'Brien and Toms, 2008; Csikszentmihalyi, 1993).

Achievements can be a motivational factor (Marczewski, 2015) and this became apparent in the interviews with the participants. The level of motivation provided by the achievements varied between them, since their time invested in the application was very different compared to each other. This might also have affected their perceived time, since the participant who used the application for the longest duration also had the most extreme time distortion, which could indicate that he was more engaged than the others (Csikszentmihalyi, 1993).

There was almost no difference between the average progression of the two groups, even though progression was used as a motivational factor (Marczewski, 2015) in the gamified version. However, the gamified version increased a participant's progression visually when a new page was visited. All the participants in the gamified group visited all the pages, which might indicate that they were motivated by progression. They did however not always get 100% progression since they still needed to get 3 stars in each scenario, but their perception of time might have had an affect on why they did not keep trying to get three stars as not having sufficient time might disengage them (O'Brien and Toms, 2008).

The application proved to provide good aesthetics while also having a clear goal, which was believed to help them get engaged, as pleasing aesthetics and specific goals affect point of engagement (O'Brien and Toms, 2008). The visual and audio feedback was not always noticed, and some of the ratings might have skewed the results as the participants rated them despite not having noticing any feedback, but the overall feedback was good and the navigation was very easy, which prevented participants from getting disengaged by poor usability (O'Brien and Toms, 2008).

We did not know how much knowledge the participants had within first aid since some of them might have forgotten most of what they learned from their last CPR course while others could still have remembered most of it.This meant that we could not know if their answers to the scenario 3 were influenced by the information given in the application or if they knew the answers beforehand. Instead we should have given them the paper scenario before they tried the application and have them answer questions regarding the third scenario. Then when they had tried the application, we should have presented the same paper scenario and asked them whether or not the wanted to change any of their answers. If they would change some of their answers we would be able to conclude that they have learned something from the application or that it had refreshed their knowledge.

A major reliability issue that was discovered in the findings was the low KALPHA value of both the interviews and the observations. However, this did not necessarily mean that the data was unreliable since each coder only went through data once, without having any previous experience in coding content. Additionally, Donato, Bartlett, Hager, Ekman and Sejnowski (1999) state that it requires over 100 hours of training before a coder is considered minimally competent

to code facial expressions using FACS. We believed that the KALPHA values for both groups would have been significantly higher if the coders had analysed the data multiple times and had extensive training in both coding and using FACS.

Ten participants reported that the test environment had disturbed them throughout the test. If these ten were excluded, then the remainder of the participants in the non-gamified group scored higher on question 2 than the gamified. Based on this we believe that the test environment has had an effect on the results. In addition, they also show that the gamified version failed to make the participants more likely to forget their surroundings and increase engagement (O'Brien and Toms, 2008). However, factors such as the small sample size of the remaining participants (n=10) and that the numbers were based on self-assessment made it difficult to come to a reliable conclusion.

The sampling of the participants could also have influenced the test. As 17 of the participants were students at Aalborg University Copenhagen, they all had semester projects to hand-in within four weeks of the test. Eleven of them were at 6th semester, which means that they were working on their Bachelor's project at that time. This could have caused them to disengage themselves early and stop the test, because they did not feel that they had time for it (O'Brien and Toms, 2008). Furthermore, all of the test participants were acquaintances of the test conductors, this means that they might have had some knowledge about the test before it started. This could have caused them to inflate their ratings in order to please the test conductors. Additionally, one participant's (G5) low expectations could have caused him to inflate his ratings, as he might have been surprised by the quality of the prototype. It could also have caused him to deflate his ratings as he knew it was not going to be a product ready to be released. Another participant (NG2) expressed that he had an idea of what the test was evaluating, which caused him to spend time looking for distractions caused by the conductors. Fortunately, he was wrong in his assumption, however it is impossible to say if it had any influence on his level of engagement, as he spent time elsewhere rather than on the application.

Chapter 7

Conclusion

This study has investigated how game design elements can be used for creating an engaging and informative application, where first aid was chosen as a theme with focus on CPR. First aid certificates expire after two years, however not all people are motivated to renew them, but game design elements can be motivational factors for users, and we thought these could be used to the advantage of getting people engaged in refreshing their knowledge in a mobile application. The following attributes of engagement have been used for measuring engagement; positive affect, interest, clear goal, motivation, aesthetics, feedback, perceived time and challenge. These attributes were based on prior studies by O'Brien and Toms (2008), Sweetser and Wyeth (2005) and Csikszentmihalyi (1993). To evaluate whether this was true or not, two applications were developed and tested by two different groups using a between-groups test design.

The game design elements we implemented did not affect the participants engagement as we had hoped. The lack of match between the player skill level and difficulty level of the scenarios made it problematic to state whether or not the game design elements had the desired effect. However, some of them were still stated to have a positive effect on the participants' level of engagement, this being achievements and points. We discovered that some game design elements negatively affected engagement due to the way that they were implemented. This means that adding multiple elements does not automatically mean that engagement will increase, but each of them should be wisely considered before implemented.

To gauge the participants engagement we used a VNRS within semi-structured interviews. The questions were based on questionnaires from O'Brien and Toms (2009) and Brockmyer et al. (2009). This allowed us to get a self-assessed idea of their level of engagement. However, many participants asked to have the questions repeated, which could mean that the questions was not formulated properly. In addition, an inexperience interviewer was used, which means that the participants could have understood the questions differently.

Observations were used to analyse enjoyment and concentration in the participants' facial expressions using FACS, where three coders analysed the data. It required a large amount of time

to analyse facial expressions, and KALPHA was used to check for intercoder reliability which showed that most of the observations were not reliable. This implied that the facial expressions should have been analysed several times, especially since none of the coders had any training in analysing facial expressions. However, it did give an indication of where their attention was directed at, which showed if they had an interest in the application or not.

Even though we were unable to measure whether or not participant learned anything from the application, they stated several times that it was informative, refreshed their knowledge and even taught elements of first aid they did not know.

This study showed a minor increase in engagement when the system featured game design elements. It was speculated that this was an indirect cause. The test participants that were rated to be the most engaged all mentioned the game design elements as a motivational factor to keep playing. Most often getting all achievements was the reason for them to keep playing. However, there was no evidence for this claim other than participants' own assessment and their desired play time.

In order to reach more decisive findings, several factors should be improved; that is experienced coders, a larger sample size, a better sampling of the participants and better implemented game design elements.

Bibliography

Anderson, L. W., Krathwohl, D. R. and Bloom, B. S. (2001), *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*, Allyn & Bacon.

Atkin, A. (2013), 'Peirce's theory of signs'. `http://plato.stanford.edu/entries/peirce-semiotics/` [Accessed: May 24, 2016].

Bartle, R. (1996), 'Hearts, clubs, diamonds, spades: Players who suit muds', *Journal of MUD research* **1**(1), 19.

Bjørner, T. (2015*a*), Data analysis and findings, *in* T. Bjørner, ed., 'Qualitative methods for Consumer Research', Hans Reitzels Forlag, Denmark, chapter 3, pp. 97–106.

Bjørner, T. (2015*b*), Data collection, *in* T. Bjørner, ed., 'Qualitative methods for Consumer Research', Hans Reitzels Forlag, Denmark, chapter 2, pp. 57–96.

Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroom, E., Burkhart, K. M. and Pidruzny, J. N. (2009), 'The development of the game engagement questionnaire: A measure of engagement in video game-playing', *Journal of Experimental Social Psychology* **45**(4), 624–634.

Charlier, N. and De Fraine, B. (2013), 'Game-based learning as a vehicle to teach first aid content: a randomized experiment', *Journal of school health* **83**(7), 493–499.

Cohn, J. F., Ambadar, Z. and Ekman, P. (2007), 'Observer-based measurement of facial expression with the facial action coding system', *The handbook of emotion elicitation and assessment* pp. 203–221.

Csikszentmihalyi, M. (1993), *The Evolving Self: A Psychology for the Third Millennium*, Perennial.

DeRose, D. J. and Laurel, B. (1993), 'Computers as theatre'.

Deterding, S. (2011), Situated motivational affordances of game elements: A conceptual model, *in* 'Gamification: Using Game Design Elements in Non-Gaming Contexts, a workshop at CHI'.

Deterding, S., Sicart, M., Nacke, L., O'Hara, K. and Dixon, D. (2011), Gamification. using game-design elements in non-gaming contexts, *in* 'CHI'11 Extended Abstracts on Human Factors in Computing Systems', ACM, pp. 2425–2428.

Dickey, M. D. (2005), 'Engaging by design: How engagement strategies in popular computer and video games can inform instructional design', *Educational Technology Research and Development* **53**(2), 67–83.

Donato, G., Bartlett, M. S., Hager, J. C., Ekman, P. and Sejnowski, T. J. (1999), 'Classifying facial actions', *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **21**(10), 974–989.

Entertainment, S. (2013), 'Playstation'. `http://us.playstation.com/community/mytrophies/` [Accessed: May 23, 2016].

Faghihi, U., Brautigam, A., Jorgenson, K., Martin, D., Brown, A., Measures, E. and Maldonado-Bouchard, S. (2014), 'How gamification applies for educational purpose specially with college algebra', *Procedia Computer Science* **41**, 182–187.

Flatla, D. R., Gutwin, C., Nacke, L. E., Bateman, S. and Mandryk, R. L. (2011), Calibration games: making calibration tasks enjoyable by adding motivating game elements, *in* 'Proceedings of the 24th annual ACM symposium on User interface software and technology', ACM, pp. 403–412.

Ford, J. K. and Weissbein, D. A. (1997), 'Transfer of training: An updated review and analysis', *Performance improvement quarterly* **10**(2), 22–41.

Hayes, A. F. and Krippendorff, K. (2007), 'Answering the call for a standard reliability measure for coding data', *Communication methods and measures* **1**(1), 77–89.

Hoffman, B. and Nadelson, L. (2010), 'Motivational engagement and video gaming: A mixed methods study', *Educational Technology Research and Development* **58**(3), 245–270.

Hwa, S. P. (2009), Create an edutainment learning environment for mathematics learning in primary schools, *in* 'Proceeding of 2nd International Conference on Teaching and Learning (ICTL 2009): Achieving Educational Excellence through Student Centred ApproachISBN', pp. 978–983.

Khaddage, F., Lattemann, C. and Acosta-Díaz, R. (2014), Mobile gamification in education engage, educate and entertain via gamified mobile apps, *in* 'Proc Society for Information Technology & Teacher Education International Conference', pp. 1654–1660.

Krathwohl, D. R. (2002), 'A revision of bloom's taxonomy: An overview', *Theory into practice* **41**(4), 212–218.

Lazzaro, N. (2004), Why we play games: Four keys to more emotion without story, *in* 'Game Developers Conference'.

Malone, T. W. (1982), Heuristics for designing enjoyable user interfaces: Lessons from computer games, *in* 'Proceedings of the 1982 conference on Human factors in computing systems', ACM, pp. 63–68.

Marczewski, A. (2015), 'Even ninja monkeys like to play: Gamification, game thinking and motivational design'.
`http://www.gamified.uk/user-types/` [Accessed: May 22, 2016].

Microsoft (2013), 'Xbox one'. `http://www.xbox.com/en-US/xbox-one/experience?xr=shellnav` [Accessed: May 23, 2016].

Monsieurs, K. G., Nolan, J. P., Bossaert, L. L., Greif, R., Maconochie, I. K., Nikolaou, N. I., Perkins, G. D., Soar, J., Truhlář, A., Wyllie, J., Zideman, D. A., Khalifa, G. E. A., Alfonzo, A., Arntz, H.-R., Askitopoulou, H., Bellou, A., Beygui, F., Biarent, D., Bingham, R., Bierens, J. J., Böttiger, B. W., Bossaert, L. L., Brattebø, G., Brugger, H., Bruinenberg, J., Cariou, A., Carli, P., Cassan, P., Castrén, M., Chalkias, A. F., Conaghan, P., Deakin, C. D., Buck, E. D. D., Dunning, J., Vries, W. D., Evans, T. R., Eich, C., Gräsner, J.-T., Greif, R., Hafner, C. M., Handley, A. J., Haywood, K. L., Hunyadi-Antičević, S., Koster, R. W., Lippert, A., Lockey, D. J., Lockey, A. S., López-Herce, J., Lott, C., Maconochie, I. K., Mentzelopoulos, S. D., Meyran, D., Monsieurs, K. G., Nikolaou, N. I., Nolan, J. P., Olasveengen, T., Paal, P., Pellis, T., Perkins, G. D., Rajka, T., Raffay, V. I., Ristagno, G., Rodríguez-Núñez, A., Roehr, C. C., Rüdiger, M., Sandroni, C., Schunder-Tatzber, S., Singletary, E. M., Skrifvars, M. B., Smith, G. B., Smyth, M. A., Soar, J., Thies, K.-C., Trevisanuto, D., Truhlář, A., Vandekerckhove, P. G., de Voorde, P. V., Sunde, K., Urlesberger, B., Wenzel, V., Wyllie, J., Xanthos, T. T. and Zideman, D. A. (2015), 'European resuscitation council guidelines for resuscitation 2015: Section 1. executive summary', *Resuscitation* **95**, 1 – 80. European Resuscitation Council Guidelines for Resuscitation 2015European Resuscitation Council Congress - Prague 2015.
URL: *http://www.sciencedirect.com/science/article/pii/S0300957215003500*

Nike (2016), 'Nike+'.
`https://secure-nikeplus.nike.com/plus/` [Accessed: May 22, 2016].

O'Brien, H. L. and Toms, E. G. (2008), 'What is user engagement? a conceptual framework for defining user engagement with technology', *Journal of the American Society for Information Science and Technology* **59**(6), 938–955.

O'Brien, H. L. and Toms, E. G. (2009), 'The development and evaluation of a survey to measure user engagement', *Journal of the American Society for Information Science and Technology* **61**(1), 50–69.

Pausch, R., Gold, R., Skelly, T. and Thiel, D. (1994), What hci designers can learn from video game designers, *in* 'Conference companion on human factors in computing systems', ACM, pp. 177–178.

Pivec, M., Dziabenko, O. and Schinnerl, I. (2004), 'Game-based learning in universities and lifelong learning:"unigame: social skills and knowledge training" game concept', *Journal of Universal Computer Science* **10**(1), 14–26.

Popcap (2007), 'Peggle'. http://www.popcap.com/peggle-1 [Accessed: May 23, 2016].

Port, R. (2000), 'Icon, index and symbol: Types of signs'. http://www.cs.indiana.edu/~port/teach/103/sign.symbol.short.html [Accessed: May 24, 2016].

Quesenbery, W. (2003), *The five dimensions of usability*, Vol. 20, Lawrence Erlbaum Associates Mahwah, NJ.

Robson, K., Plangger, K., Kietzmann, J. H., McCarthy, I. and Pitt, L. (2016), 'Game on: Engaging customers and employees through gamification', *Business Horizons* **59**(1), 29–36.

Rogers, Y., Sharp, H. and Preece, J. (2011), *Interaction Design: Beyond Human - Computer Interaction*, Interaction Design: Beyond Human-computer Interaction, Wiley.

Rozin, P. and Cohen, A. B. (2003), 'High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of americans.', *Emotion* **3**(1), 68.

Rådet for Sikker Trafik (2016), 'Nederen forældre - nederen er kærlighed'. "https://www.youtube.com/watch?v=s7A-eLLn_Ts"[Accessed: May 23, 2016].

Shernoff, D. J., Csikszentmihalyi, M., Shneider, B. and Shernoff, E. S. (2003), 'Student engagement in high school classrooms from the perspective of flow theory.', *School Psychology Quarterly* **18**(2), 158.

Sweetser, P. and Wyeth, P. (2005), 'Gameflow: a model for evaluating player enjoyment in games', *Computers in Entertainment (CIE)* **3**(3), 3–3.

Unit9 (2013), 'Lifesaver mobile - android apps on google play'. https://play.google.com/store/apps/details?id=air.com.unit9.lifesaver.android.mobile&hl=en [Accessed: May 22, 2016.

Von Ahn, L. and Dabbish, L. (2008), 'Designing games with a purpose', *Communications of the ACM* **51**(8), 58–67.

Webster, J. (1988), Making computer tasks at work more playful: Implications for systems analysts and designers, *in* 'Proceedings of the ACM SIGCPR conference on Management of information systems personnel', ACM, pp. 78–87.

Weinberger, M. G. and Gulas, C. S. (1992), 'The impact of humor in advertising: A review', *Journal of advertising* **21**(4), 35–59.

Wissenberg, M., Hansen, C. M., Mortensen, R. N., Folke, F., Lippert, F., Christensen, E. F., Hansen, P. A., Lang-Jensen, T., Henriksen, O. M. and Torp-Pedersen, C. (2013), 'Hjertestop uden for hospital i danmark'. http://genoplivning.dk/wp-content/uploads/2013/10/Dansk-Hjertestopregister-2001-2011.pdf [Accessed: May 22, 2016].

Zichermann, G. and Cunningham, C. (2011), *Gamification by design: Implementing game mechanics in web and mobile apps*, " O'Reilly Media, Inc.".

Appendix A

VNRS Questions - Danish

Attributes of engagement	**Spørgsmål**
Opfattet Time	1 På en skala fra 0-10, hvor 0 er overhovedet ikke og 10 er glemte alt, hvor meget glemte du omgivelserne omkring dig?
Æstetik	2 På en skala fra 0-10, hvor 0 er dårligt og 10 er godt, hvad synes du om det visuelle i brugergrænsefladen ? 3 På en skala fra 0-10, hvor 0 er dårligt og 10 er godt, hvordan synes du det visuelle var præsenteret i forhold til det du troede det var? F.eks. synes du knapperne præsenterede knapper?
Feedback	4 På en skala fra 0-10, hvor 0 er dårlig og 10 er god, hvordan synes du lydene repræsenterede dine valg i applikationen? 5 På en skala fra 0-10, hvor 0 er dårlig og 10 er god, hvordan var den visuelle feedback du fik da du trykkede på knapperne i applikationen?
Udfordring	6 På en skala fra 0-10, hvor 0 er svært og 10 er nemt, hvor ville du placere sværhedsgraden på scenarierne? 7 På en skala fra 0-10, hvor 0 er lidt og 10 er meget, hvor udfordrede følte du dig prøvede applikationen?
Glæde	8 På en skala fra 0-10, hvor 0 er svær og 10 er nem, hvordan synes du applikationen var at navigere i? 9 På en skala fra 0-10, hvor 0 er ikke sjovt og 10 er meget sjovt, hvordan havde du det da du brugte applikationen?
Interesse	10 På en skala fra 0-10, hvor 0 er ikke interessant og 10 er meget interessant, hvor interessant var det at lære omkring førstehjælp gennem applikationen? 11 På en skala 0-10, hvor 0 er ikke interesserede og 10 er meget interesserede, hvor interesseret var du i at bruge applikationen?
Mål	12 På en skala fra 0-10, hvor 0 er ikke klart og 10 er klart, hvor klart var målet med scenarierne så?

Appendix B

Questions about the application - Danish

Sværhedsgrad	**Spørgsmål**
1	“Hvilken baggrundsfarve havde knapperne”
2	“Hvor mange menuer/underkategorier var der i aflåst sideleje siden?”
3	”Hvor mange led var der i overlevelseskæden?”
4	“Hvor mange uventede hjertestop sker der uden for hospitalet om året i Dänemark?”

Appendix C

Questions about first aid - Danish

Spørgsmål	**Svar mulighed**
Du er alene og ser en person ligge i vandkanten, du hiver ham væk fra vandet, personen er bevidstløs, hvad gør du?	1. Tjekker om han trækker vejret 2. Løber efter hjælp 3. Ring 1-1-2
Efter du har tjekket om han trækker vejret, hvilket han ikke gør, hvad gør du så?	1. Start HLR 2. Læg personen i aflåst sideleje 3. Ringer 1-1-2
Efter du har ringet 1-1-2, hvad gør du så?	1. Starter HLR 2. Afventer hjælp 3. Du forlader ham for at finde hjælp
Du skal til at starte HLR, hvor mange brystkompressioner og pust er det rigtige antal at starte med?	1.5 pust efterfulgt af almindelig HLR (30:2) 2. 3 pust efterfulgt af almindelig HLR (30:2) 3. 2 pust efterfulgt af almindelig HLR (30:2)

Appendix D

Relevant FACS Action Units

Illustrations borrowed from Cohn, Ambadar and Ekman (2007)

4 Brow Lowerer — *Corrugator supercilii, Depressor supercilii*

6 Cheek Raiser — *Orbicularis oculi, pars orbitalis*

12 Lip Corner Puller — *Zygomaticus major*

42 Slit — *Orbicularis oculi*

44 Squint — *Orbicularis oculi, pars palpebralis*

57 Head forward ---

Appendix E

Informed Consent

Samtykke Erklæring

Dette er et samtykke for et gruppe projekt lavet af gruppe 623 (MED6-2016).

Samtykke for deltagelse i evaluering af en prototype

Jeg deltager hermed til at teste et projekt lavet af gruppe 623 på Aalborgs Universitet København. Jeg acceptere at test prototypen og bliver interviewet bagefter.

1. Jeg deltager i dette projekt frivilligt og jeg forstår at jeg kan stoppe test sessionen når jeg har lyst.

2. Testen tager omkring 15-20 minutter og jeg er ikke tvunget til at svare på spørgsmålene under interviewet.

3. Jeg forstår at personlige informationer vil blive holdt fortroligt og sikkert.

4. Jeg giver mit samtykke til at blive videooptaget under testen og blive lydoptaget under interviewet.

5. Jeg giver mit samtykke til at data samlet og optagelser fra test sessionen vil blive brugt til en eksamen for Medialogi på Aalborg Universitet København.

6. Jeg forstår at videooptagelser fra test sessionen og lydoptagelser fra interviewet vil blive delt med gruppens supervisor og censor til eksaminationen.

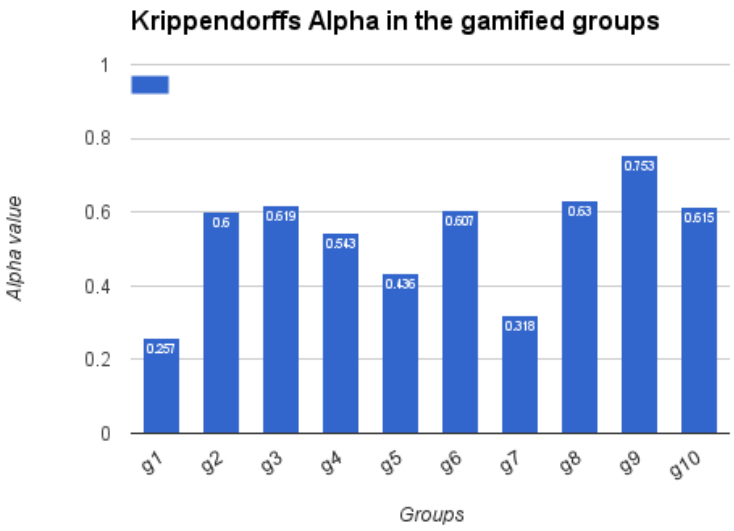
7. Jeg har læst og forstået alt hvad der er blevet forklaret på dette stykke papir og jeg vil gerne være med i eksperimentet.

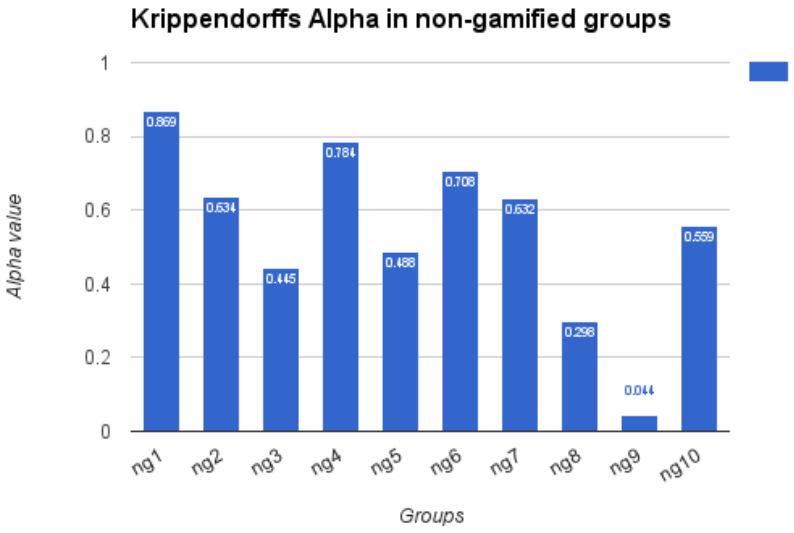
_____ _____

Underskrift Dato

Appendix F

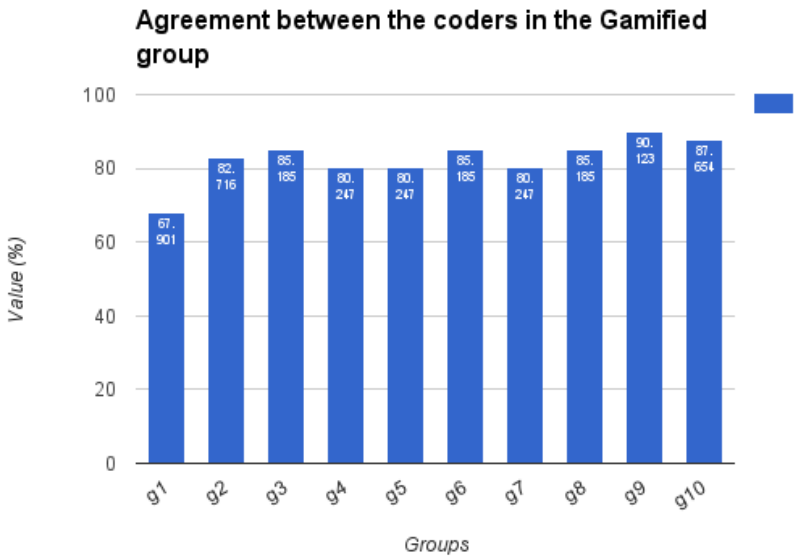
Krippendorff's Alpha Graphs

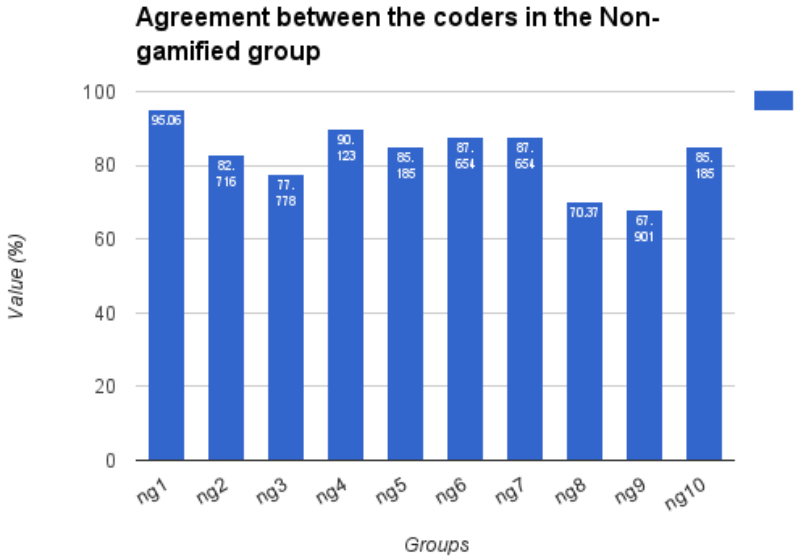




Appendix G

Coder Agreement Percentage Graphs





Appendix H

Project Assets

Please see folder *"Appendix H - Unity Project"* on the attached CD.

Scripts can be found in "Appendix H - Unity Project/Assets/Scripts/"

Images can be found in "Appendix H - Unity Project/Assets/Images/"

Sounds can be found in "Appendix H - Unity Project/Assets/Sounds/"

Appendix I

Credit to Third Parties

Main Scene
First Aid suitcase

http://www.graphicsfuel.com/2012/05/first-aid-medical-kit-icon-psd/

Learn CPR (Lær HLR)
Icons on buttons

https://design.google.com/icons/

Basic HLR (Basis HLR)
Icons on buttons

https://design.google.com/icons/

Illustrations:

http://www.cprguidelines.eu/assets/downloads/guidelines/S0300-9572(15)00327-5_main.pdf

Recovery Position (Sideleje)
Illustrations:

Can't find original

AED (Hjertestarter)
Illustrations:

http://www.cprguidelines.eu/assets/downloads/guidelines/S0300-9572(15)00327-5_main.pdf

Challenges (Udfordringer)
Scenario 1 Ambulance wheels screeching (cars045.wav)

http://www.grsites.com/archive/sounds/category/18/?offset=36

Ambulance engine sound

https://www.freesound.org/people/msorbo/sounds/31015/

Achievements (Bedrifter)

Facts (Fakta)
http://genoplivning.dk/wp-content/uploads/2013/10/Dansk-Hjertestopregister-2001-2011.pdf

Appendix J

Positive Sound

Please see "Appendix J - Positive Sound.wav" on the attached CD.

Appendix K

Negative sound

Please see "Appendix K - Negative Sound.wav" on the attached CD.

Appendix L

Interviews

Please see folder *"Appendix L - Interviews"* on the attached CD. Each file contains interview for each participant. The name of the file states the ID of the participant (E.g. G1.mp4 is participant G1)

Appendix M

Recordings for Observation

Please see folder *"Appendix M - Recordings"* on the attached CD. Each file contains recordings for each participant. The name of the file states the ID of the participant (E.g. G1.m4a is participant G1)

Appendix N

Data Logs

Please see folder *"Appendix N - Data logs"* on the attached CD. Each file contains the data log for each participant. The name of the file states the ID of the participant (E.g. G1.txt is participant G1)

Appendix O

Gamified Demographics

ID	Age	Gender	First aid knowledge (0-5)*	Last course**
G1	24	Female	2	6
G2	24	Male	4	1
G3	24	Male	4	7
G4	23	Male	5	1
G5	23	Male	4	5
G6	23	Male	3	7
G7	24	Female	3	1
G8	20	Female	2	3
G9	23	Female	2	5
G10	27	Female	2	7

*Based on the question: *"On a scale from 1-5, how would you rate your knowledge about first aid?" (0 being poor and 5 being good*

**Years since last first aid course taken.

Appendix P

Non-gamified Demographics

ID	Age	Gender	First aid knowledge (0-5)*	Last course**
NG1	25	Male	3	7
NG2	23	Male	2	5
NG3	23	Female	4	1
NG4	25	Female	2	5
NG5	20	Female	3	2
NG6	28	Male	5	3
NG7	23	Female	2	5
NG8	24	Female	3	3
NG9	23	Male	2	5
NG10	26	Female	3	8

*Based on the question: *"On a scale from 1-5, how would you rate your knowledge about first aid?" (0 being poor and 5 being good*

**Years since last first aid course taken

Appendix Q

Gamified VNRS results

	Gamified										
	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	Average
Q1	8	7	8	8	6	8	5	7	5	5	6.7
Q2	7	8	8	8	7	6	9	6	7	7	7.3
Q3	9	8	10	10	8	10	10	10	9	10	9.4
Q4	7	4	10	10	6	9	10	10	9	10	8.5
Q5	8	8	5	8	5	8	10	9	8	9	7.8
Q6	6	9	9	4	9	9	7	6	10	8	7.7
Q7	7	2	3	3	9	4	8	4	5	5	5
Q8	6	10	10	9	0	9	10	10	10	8	8.2
Q9	9	5	7	9	5	8	10	7	7	6	7.3
Q10	8	7	10	7	7	5	10	10	10	9	8.3
Q11	6	7	8	8	3	7	10	9	8	8	7.4
Q12	10	10	10	10	10	9	10	10	10	10	9.9
Average	9.1	8.5	9.8	9.4	7.5	9.2	10.9	9.8	9.8	9.5	9.35

Appendix R

Non-gamified VNRS results

Non-Gamified

	NG1	NG2	NG3	NG4	NG5	NG6	NG7	NG8	NG9	NG10	Average
Q1	10	6	7	5	10	6	7	7	6	8	7.2
Q2	8	7	10	8	8	8	7	9	7	10	8.2
Q3	10	8	7	10	10	9	10	6	8	9	8.7
Q4	10	9	10	8	8	5	10	10	10	10	9
Q5	8	5	10	10	7	8	5	7	8	10	7.8
Q6	10	2	9	10	5	9	4	7	7	7	7
Q7	0	1	6	0	2	6	2	5	2	5	2.9
Q8	10	9	10	10	10	10	10	8	9	9	9.5
Q9	8	2	8	7	9	8	7	8	4	9	7
Q10	7	6	9	8	10	7	9	6	6	10	7.8
Q11	2	5	9	5	6	8	6	6	6	10	6.3
Q12	10	10	10	9	9	9	7	10	8	10	9.2
Average	9.3	7	10.5	9	9.4	9.3	8.4	8.9	8.1	10.7	9.06

Appendix S

Interview Coding

Please see "Appendix S - Traditional Coding of interviews.pdf" in the attached CD.

Appendix T

Observation Coding

Please see "Appendix T - Traditional Coding of observations.pdf" in the attached CD.

Appendix U

Wilcoxon rank sum test computations

```
%Matlab code
a = [166.6670;114.9430;143.7130;107.1430;121.9510;
    70.3130;108.6960;105.1550;148.7600;60.6530];

b = [104.7120;83.4490;229.2990;102.7400;
    99.3380;109.6890;43.0420;78.6370;118.9800;73.6840];

%% Anderson Darling Test

% An anderson darling test is performed in order to test if the data is
    normally distributed.
% The adtest tests whether the data meets the assumption of being
    normally distributed.
[hDarlingA pDarlingA] = adtest(a)
[hDarlingB pDarlingB] = adtest(b)

% The p-value from the adtest needs to be small to be able to reject the
     null-hypoteses.
% The value from the gamified group was low, so we cannot assume that
    the data is normally distributed.

%% Wilcoxon rank sum test

[p,h,stats] = ranksum(a,b)
```