MASTER THESIS

# Usability testing of a virtual reality intervention and possibilities of wearables for treatment in forensic mental healthcare


**AUTHOR**

Thien Vuong

**SUPERVISORS**

Dr. H. Kip

Dr. S. M. Kelders

MSc. M. Kouijzer

HEALTH SCIENCES, UNIVERSITY OF TWENTE


2021

UNIVERSITY OF TWENTE.

# Abstract

**Background**: Treatment of offenders in forensic mental healthcare is complex. Forensic psychiatric patients have often low treatment motivation, suffer from multiple disorders, and have low education skills. Virtual reality (VR) and wearables may be able to improve treatment because of its potential to increase motivation and engagement. Therefore, a VR intervention "Triggers&Helpers" application is being developed in a participatory development process with input from different stakeholders. However, little is still known about suitable methods for the development of VR, such as usability testing. In addition, the use of wearables has been studied with the "Triggers&Helpers" application.

**Objective**: The three research goals are to identify (1) points of improvements in the existing "Triggers&Helpers" application according to therapists, (2) positive points and points of improvements of the usability of VR according to outpatients and (3) possible ways of using wearables with the "Triggers&Helpers" application.

**Methods**: The first usability test and interview were conducted online with five therapists to gain positive, negative and points of improvements about the dashboard of the "Triggers&Helpers" application. The usability test consisted of multiple tasks, which the therapist had to fulfil in the dashboard. During the usability test, the think aloud method was used. The second usability test and interview were conducted with five outpatients to gain first impressions about the "Triggers&Helpers" application and the usability of VR. The VR set with the "Triggers&Helpers" application, and the wearable Empatica E4 were used. The Empatica E4 was used to monitor physiological data, such as the heart rate and tension level of outpatients.

**Results**: Usability tests with therapists resulted in fifteen usability issues that could be divided into five categories, namely login, new template, settings virtual character, settings location, and saving templates issues. Usability tests with outpatients resulted in eight usability issues that could divided in three categories, namely VR device, observation in VR, and moving in VR issues. With the data of the Empatica E4, a table was provided, which displayed the heart rate and tension level that was recorded per scenario. Identified peaks of tension level were related to an occurrence, which means something happened in the VR simulation.

**Conclusion**: The results of both qualitative studies provided insights into the points of improvements of the "Triggers & Helpers" application and the use of VR according to therapists and outpatients. The use of a wearable could have an added value in treatment as an interoceptive awareness tool, together with the "Triggers&Helpers" application. Wearables are a promising tool to use to increase the interoceptive awareness of patients, but more research is needed before integrating wearables in treatments in the forensic mental healthcare.

# Table of contents

# 1. Introduction

Forensic psychiatric patients have increased in the past few years while treatment of these patients is regarded as complex in forensic mental healthcare [1]. Forensic mental healthcare focuses on patients who show aggressive or sexual delinquent behaviour that result in offending and suffer from at least one psychiatric disorder, which may refer to alcohol abuse, antisocial personality, or post-traumatic stress disorder [2]. The complexity of treating forensic psychiatric patients can be partly explained by characteristics of the patients' population. The first barrier is that most forensic psychiatric patients lack motivation to participate their treatment. This barrier could result in a low effectiveness or possible renounce of treatments [3]. The second barrier is that most forensic psychiatric patients have difficulty in reflecting on their behaviour and emotions [4]. This barrier is common among forensic psychiatric patients and mostly due to low education, low cognitive skills, lack of social skills, or lack of reflecting skills [5, 6]. This means that it is difficult for them to effectively participate in their treatment, which mostly contains a lot of reflecting exercises. Third, treatment of forensic psychiatric patients is regarded as complex, because forensic psychiatric patient population is diverse, which refers to different types of offence, mental disorders, personal backgrounds, and personality types [7]. Forensic ment1al healthcare focuses on preventing criminal recidivism via approaching offence related risk factors, such as antisocial behaviour or coping skills [2, 5]. These risk factors can be approached by therapies and interventions that are based on evidence-based approaches. Examples of these approaches are cognitive behaviour therapy (CBT) and risk-needs-responsivity (RNR) principles [8, 9]. However, developing and implementing such approaches in forensic mental healthcare has proven to be challenging. Meta analyses found low effectiveness on clinical measures of interventions that targeting abusers, juvenile delinquency, and relapse prevention of offenders [10]. To improve forensic mental healthcare, a solution could be the use of eHealth in treatment, such as virtual reality (VR) or wearables. eHealth can be defined as technologies to improve and support health, well-being, and quality of care [11].

There are multiple eHealth technologies that seem to be suitable for treatment in forensic mental healthcare. A review pointed out that eHealth technology was able to deal with the complex nature of the forensic psychiatric population [12]. VR technology can take low education of forensic psychiatric patients into account by creating real-life, interactive simulations in which skills can be trained [13]. Furthermore, data on reactions can be gathered via physiological measures which can be integrated in treatment, for example by using wearables [14]. Several studies have shown that VR and wearables are interesting technologies for psychological treatment, since mental health problems such as phobia disorders, post-traumatic stress disorder (PTSD) or addictions are closely intertwined with the perceived environment [15-17]. However, most of these types of technologies are not thoroughly studied in the context of forensic mental healthcare.

Thus, more studies on technologies such as VR or wearables are required to determine the added value for forensic psychiatric patients.

VR is an intervention that has been used to simulate, coach, and educate patients in mental healthcare. In VR, patients can enter computer-generated environments [18]. A great advantage of VR is that VR simulations are not real, but the patient experiences the situation as real. Patients are willing to face more difficult situations in VR, because of the immersive qualities of VR [19]. A recent review pointed out that forensic psychiatric patients might benefit from the immersive qualities of VR [12]. A second advantage, VR can increase motivation of patients because they enjoy the use of VR, which could be a solution for patients that are lacking motivation [20]. The third advantage is that treatment can be tailored to the needs of an individual patient with VR [21]. Behavioural skills and coping strategies can be trained in controlled environments that are tailored to the individual patient's risk factors, without endangering others. Tailoring will create a better fit between VR and the individual patient, which can increase the engagement and effectiveness of the treatment [22]. All these advantages together make VR a promising technology for forensic mental healthcare.

Wearables are also one of the technologies that seem to be suitable for treatment in forensic mental healthcare. Wearables have the ability to monitor physiological signals, such as heart rate and skin conductance by using biofeedback [23]. Sensors are used to measure various bodily signals and make them visible. With biofeedback, patients can gain more insight into their physical reactions and learn to control them, which is called interoceptive awareness [23]. The possibilities of wearables enable the patient to gain new insights and acquire skills. Because wearables measure objective data and can provide immediate feedback, the patient can reflect on what happened and why at that specific moment. For example, the patient can indicate how he or she is feeling, or what caused the anger. The possibilities that wearables offer to monitor this type of data also create new opportunities for coaching, for example "just-in-time" feedback. For example, therapists can indicate when the patient's heart rate increases and start with relaxation exercises or providing tips [23]. In this way, wearables might be used with VR to increase the effectiveness of the treatment.

Several studies have shown that VR and wearables can offer many benefits for forensic mental healthcare, but there is still too little known about how they should be developed, implemented, and evaluated in practice [12, 24-26]. Therefore, a good development process is needed to guarantee a good fit between technology, people, and the context [20]. A way to do this is via participatory development, which supports cooperation with end-users and other stakeholders. In participatory development, a stakeholder can have different roles such as being an informant, who provides input into a product or design. A stakeholder can also be a co-creator, who is actively involved in creating ideas and products [27, 28]. Furthermore, stakeholders can also provide input with identifying problems where technology can be of added value, improving the design,

or identifying issues for implementation. Participatory development can be done by using methods from human-centred design (HCD), such as usability testing, prototyping and interviews [29].

Although participatory development is important, little is known about suitable methods for the development of VR and wearables, such as usability testing [29]. Usability testing is a method to test the functionality of a prototype by observing participants while they attempt to complete tasks on it. With this in mind, the VR application "Triggers&Helpers" is currently being developed in a participatory development process with different stakeholders. The “Triggers&Helpers” can simulate scenarios that are similar to real-life situations. In this study, usability tests are conducted with therapists to find out what they think of the “Triggers&Helpers”. By means of these usability tests, insights can be gained into what therapists and patients think of VR and where are points for improvements. Furthermore, previous research has pointed out the potential of combining biofeedback and VR, but not much is known about the possibilities [23]. This project will therefore provide insight into the participatory usability testing of VR by means of innovative research methods. The main goal of this thesis is to plan and execute usability tests of the VR application “Triggers&Helpers” with therapists and patients and to perform a pilot study about the possibilities of the use of wearables with VR. The three research goals are to identify (1) points of improvements in the existing “Triggers & Helpers” application according to therapists, (2) positive points and points of improvements of the usability of VR according to patients and (3) possible ways of using wearables with the “Triggers&Helpers” application that can improve current treatment.

# 2. Methods

## 2.1 Setting

Forensic mental healthcare focuses on identifying and treating risk factors of an individual patient [8]. The advantages of VR have led to a development of a new intervention by an interdisciplinary "VooRuit met VR" project team, consisting of key stakeholders such as patients, therapists, researchers, VR developers, and a health sciences student. In this study, usability tests were conducted with therapists and patients from two Dutch forensic mental clinics: Transfore and De Waag. Transfore and De Waag are organizations that offer forensic mental healthcare. Transfore treats patients in the east of the Netherlands who suffer from delinquent behaviour, such as problems with aggression or sexuality. Transfore treats more than 1500 people every year who exhibited transgressive behaviour. As a foundation, Transfore is part of the Dimence Group and has approximately 200 employees. Different in- and outpatient clinics are located in the Netherlands where patients go to be treated. De Waag was also participating the interdisciplinary project team and is a co-financier in this project.

## 2.2 Triggers and Helpers

Regardless of a forensic psychiatric patient's offence, diagnosis, or type of treatment, forensic mental healthcare always focuses on what 'triggers' a patient. During treatment, factors that increase the chances of someone committing an offense again, will be studied. Those risk factors, such as dynamic risk factors, are important in treatment of offenders [30]. Once identified, these specific risk factors can be targeted in the treatment to decrease the chance of committing an offense. Acute dynamic risk factors are dependent on the context and state of the offender. Since these factors are only relevant during short periods of time and in specific situations, they are harder to target in treatment. Therefore, the application for VR was developed in collaboration with the company CleVR, called the "Triggers&Helpers" application. The "Triggers&Helpers" application is currently a functional prototype version, a working prototype built to validate the design. The goal is to identify triggers of patients and practice their helpers in personalized virtual environments. It is important to identify these triggers in order to deal with these challenging situations with helpers. Examples of these helpers are relaxation exercises, listening to music, or distracting oneself by playing games on a smartphone. The "Triggers&Helpers" can offer the therapist and patient many tools to identify personally relevant triggers in realistic context and to practice with helpers in a controlled setting. In order to use the "Triggers&Helpers" application, hardware is needed to launch the "Triggers&Helpers", which consisted of multiple components. The hardware was provided by CleVR. The required components are shown in Figure 1. The tablet, which is wired with the VR glasses and laptop, launched the dashboard of the "Triggers&Helpers"

application. The voice-morphing microphone distorts the therapist's voice, which can be used in role-play with patients.

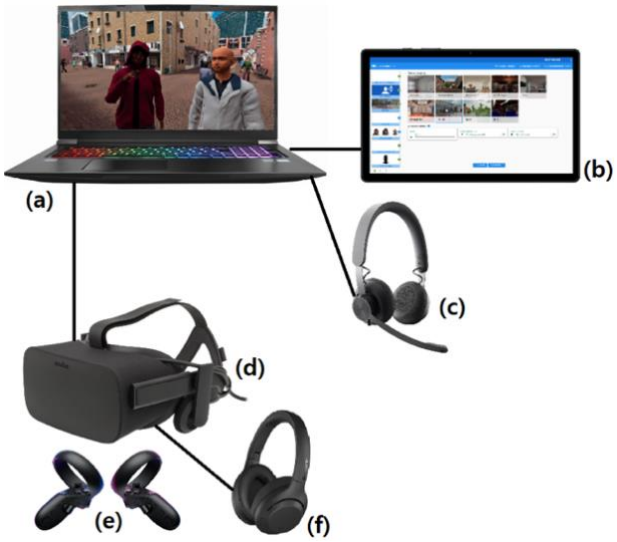


***Figure 1**. Setup of VR consisted of laptop (a), tablet with dashboard of "Triggers&Helpers" application (b), voice-morphing microphone (c), VR glasses (d), VR controllers (e), noise cancelling headphone (f).*

Therapists could use the application by navigating on the dashboard, that is launched on the tablet. The dashboard is made up of different building blocks to create a virtual session. The first block allowed the therapist to create a new session or to load a template that is saved. The second block is to select a type of exercise, such as walking around or role-play exercises. These are the only relevant exercises in this study, because only those are included in the usability test. With walking around exercises, patients could walk through virtual environments such as a shopping street. Virtual role-playing could also be used, in which the therapist "play" another virtual character via a voice-morphing microphone. When an exercise is selected, a specific environment can be chosen in the third block, for example on the corner of the shopping street. In the fourth building block, the settings are specific to the type of exercise that is chosen. For walk around exercise, types of virtual characters can be set, which will appear in the selected environment, and a specific location can be set for role-play exercise such as the corner of the street. In the fifth building block, social options, such as crowd and staring behaviour, can be set for walk around exercises. Two options can be set for role-play, which are the position of the virtual characters and the types of virtual characters. Triggers and helpers can be added in the last block. A list of triggers is provided, which is divided in different categories, and there is also an option to change the trigger to a helper.

## 2.3 Study 1 – Usability test and interview with therapists

### 2.3.1 Participants

The target group for the first study were therapists that are specialized in forensic mental healthcare. These therapists are involved in any type of treatments that are focused on

outpatients, which are patients who attends a hospital or a clinic for treatment. Requirements for participating in this study were therapists who are potential end-users of the "Triggers&Helpers" application, and who are trained in the use of VR. A total of five therapists were included in the first study because studies have shown that about five participants are required in usability tests to identify most flaws and points of improvement [31]. Four therapists were recruited from De Waag and one therapist from Transfore by using convenience sampling via therapists and researchers who are affiliated with the "VooRuit met VR" project team. These therapists were not part of the "VooRuit met VR" project team and did not have experience in the use of the "Triggers&Helpers" application. Therapists were informed beforehand about the goal and nature of the study via verbal informed consent. They participated voluntary and were able to withdraw at any time during the study, which is communicated.

### 2.3.2 Materials and procedure

The five usability tests were conducted in March 2021 online via Zoom. The usability test consisted of six scenarios and an interview. An overview of the scenarios is provided in Table 1. In one scenario, multiple smaller tasks were provided, which the therapist had to fulfil in the application. The usability tests were conducted online by one researcher, one master student, and one engineer of CleVR and took approximately 45 minutes. The researcher (MK) was leading the usability tests, provided the therapists with tasks and conducted the interview. The master student (TV) observed the usability test, took notes, wrote down the number of mistakes that the therapist made, and the time the therapist needed to complete one task. The engineer is the software developer of company CleVR, which took care of the presentation of the "Triggers&Helpers" application via Zoom. The application was launched on the stakeholder's computer and the screen was shared with the therapist. The controls were handed over, which allowed the therapist to use the application on his or her own computer via Zoom. During the usability test, the think aloud method was used, which is a stable method for understanding the thought processes and problems of test participants [32, 33]. The advantages of this method are that data during the usability test was continuously collected without many specific questions. The therapist was asked to think aloud to enable the researchers to gain insight into the first impressions and thinking process. The therapist was asked to say whatever comes into his or her mind.

***Table 1.*** *Scenarios within the usability test, which therapists had to fulfil.*

| Scenario | Description |
|---|---|
| **1** | Login the "Triggers&Helpers" application. |
| **2** | Open new template and create a walking around situation in a bus half occupied with only women. |

| | |
|---|---|
| **3** | Save the template that was created in scenario 2, including accessibility for all therapists. |
| **4** | Create a roleplay situation in grocery store with a security guard and grocery store employee. Add a trigger of your choice and start the roleplay. |
| **5** | Create a new roleplay situation in a park with a police officer and man wearing a hoodie. |
| **6** | Change the roleplay situation that is created in task 5 to a roleplay situation in a shopping street with one old man that is sitting on a bench. |

Afterwards, an interview with nine questions was conducted. The interview was based on four categories such as positive elements, difficulties that therapists have experienced, suggestions of improvements for existing components, and suggestions of points of improvements for new components. The interview was created by the researcher (MK). The overview of the interview is provided in Table 2. These questions supported the findings of the usability test to gain more insight into the results.

***Table 2.*** *Overview of questions of the interview with therapists.*

| # | Questions |
|---|---|
| **1** | What is your first impression of using the dashboard? |
| **2** | What did you find difficult about using the dashboard? |
| **3** | What did you find easy about using the dashboard? |
| **4** | Did you need more information on the dashboard to build a scenario? If so, what information? |
| **5** | What else would you like to change about the dashboard? |
| **6** | What would you like to add to this dashboard to improve it? |
| **7** | Are there any environments, characters or triggers and helpers that you are missing so far? |
| **8** | Do you find the overview of triggers and helpers clear? How would you improve this? |
| **9** | How would you use this VR application in treatment? Within which treatments? What could that look like? |

### 2.3.3 Data analysis

After the usability tests were conducted, the recordings of the five participants were analysed by the researcher (MK) and the master student (TV). The data of the observation during the usability test were structured per therapist in a table with an issue identification, where the issue happened, task description, and a concise description of the issue. The table is provided in the result section in the Tables 6. Each usability issue had a grade of severity that is influenced by task criticality, issue frequency and issue impact. To prioritize the usability issue, the first thing was to set the criticality score of each task [34]. The criticality score rated the impact on the therapist if the task is not accomplished by a numeric value. The description of each numeric value is displayed in Table 3. The second thing to set, was the impact score, which rated how much the task has impacted the therapist that was trying to accomplish the task. This was done the same way as how the criticality score was set by using the numeric value. The last score that

had to set, was the issue frequency, which was determined by the total of participants that occurred the issue.

***Table 3.*** *Description of the numeric values to determine the criticality score and impact score [34].*

| Numeric Value | Description |
|---|---|
| **5** | The issue prevents the participant from accomplishing the task. |
| **3** | It causes frustration and/or delay. |
| **2** | It has a minor effect on task performance. |
| **1** | It is a suggestion from the participant. |

The severity of each issue is calculated by multiplying the three variables above, which is shown in Equation 1. When the severity value is known, the usability issues can be sorted with higher severity values having a higher priority than usability issues with a smaller severity value. This value can be calculated with the formula below when task criticality, frequency, and impact are known. Lastly, a summary of the main usability issues is given for each individual scenario.

$$Severity = \ Criticality\ score * Impact\ score * \frac{Frequency}{total\ participants}$$

***Equation 1****: Formula to calculate the severity score.*

The audio recordings of the interviews were transcribed verbatim and coded by the master student (TV). A coding scheme was deductively created with four main codes, which were based on the type of questions that are provided in Table 2. The answers of each interview were marked and subcodes were inductively drawn up based on those marked answers. The provided subcodes were sorted in types and counted how many subcodes of the same type occurred. Finally, the definition of the codes was added, and the number of times a subcode occurred in all interviews. After feedback from other researchers, the coding schemes were adapted and used to code all interviews.

## 2.4 Study 2 – Usability test and interview with outpatients

### 2.4.1 Participants

The target group for the second study were outpatients who attended a forensic mental healthcare clinic for treatment. Three inclusion criteria have been drawn up for this study to recruit outpatients in order to gain insight into positive and negative aspects of the use of VR. The first inclusion criterium was that outpatients were willing to participate voluntary. The second criterium was including outpatients without experiences with VR and the "Triggers&Helpers" application to elicit their first impressions and experiences. The last inclusion criterion was that participation was only allowed when a therapist indicated that the usability test would not be uncomfortable or damaging for the outpatient. The usability tests were conducted with five male outpatients of an aggression

regulation group treatment from Transfore. With therapists were discussed, which outpatients could participate the study. Furthermore, an ethical approval was given by the Ethics Committee of the University of Twente (Behavioural, Management and Social Sciences; Request number 210177 and 210645).

### 2.4.2 Materials and procedure

The usability tests were conducted in June 2021 and took around 40 minutes per outpatient. For the usability test, the VR set with the "Triggers&Helpers" application, and the wearable Empatica E4 were used. The outpatients were put on the VR glasses, which they could enter a neutral virtual world that was launched by the application. By using a controller, the outpatients could walk around and explore the virtual world. The outpatients were also given an Empatica E4 wearable. It was mainly to get a first insight into whether the Empatica E4 could be an added value to use with the "Triggers&Helpers" application. The Empatica E4 was used to monitor physical data, such as the heart rate and tension level of outpatients. The Empatica E4 was worn on the wrist by the outpatient and was connected via Bluetooth to the "E4 Realtime" app that was installed on a smartphone. With this app, the measured physical data could be registered and stored.

During this study, the outpatients' therapist was present, next to one researcher (MK) and one master student (TV). The researcher (MK) conducted the usability test and evaluation interview. The master student (TV) kept track of time, started the recordings of VR and Empatica E4, and observed the level of tension of the outpatient by viewing real-time data of the Empatica E4. All outpatients were informed beforehand about the goal and nature of this study and were also informed about the Empatica E4. All outpatients signed an informed consent form. The usability test consisted of four tasks, which illustrated how the "Triggers&Helpers application" might be used in treatment for outpatients. An overview of the task is provided in Table 4.

***Table 4****. Task within the usability test with outpatients.*

| Task | |
|---|---|
| **1** | Getting the outpatient used to VR glasses and noise-cancelling headphone. |
| **2** | Walk around exercise in shopping street with a few people walking around quietly and little noise in the background. There is a neutral atmosphere. |
| **3** | Walk around exercise in shopping street with few more people walking around, more noise in the background, and sounds of car brakes and car driving away. There is a neutral atmosphere. |
| **4** | Role-play exercise with virtual male character about the weather and neutral hand gestures, such as thumps up and goodbye wave. |

Between every task, outpatients were taking a break and asked how they experienced the task. Furthermore, the tension level of outpatients was checked by asking them about these levels on a scale of one to ten and by observing the data of the Empatica E4. When the tension level exceeded the value of twenty micro-Siemens (μS), the test should be

stopped for a while before it could escalate [35]. At the end, outpatients were asked seven evaluation questions, provided in Table 5. The interview was based on three categories, which was positive elements, difficulties that outpatients have experienced and points of improvements. The interview was created by the researcher (MK). To conclude the session, questions were asked about the whole experience with VR, and each outpatient was thanked for their collaboration with a VVV-voucher with value of ten euros.

***Table 5.*** *Overview of questions of the interview with outpatients.*

| # | Questions |
|---|---|
| **1** | What is your first impression of VR? |
| **2** | What did you like/positive/good about VR? |
| **3** | Are there things that you liked less/less good/negative about VR? |
| **4** | How realistic/real did it feel? |
| **5** | What did you notice about yourself? |
| **6** | What struck you when you were in VR? |
| **7** | Would you like to use VR in a treatment? What should we take into account? |

### 2.4.3 Data analysis

After performing the usability tests, three types of data were obtained, namely the screen recordings from the VR, audio from the entire usability test and interview, and recordings from the Empatica E4. These three raw datasets were merged into one video by using a video editor. The part of the usability test was analysed and again structured per outpatient in a table as in the first study. Each usability issue had again a grade of severity that is influenced by task criticality, issue frequency and issue impact. The most important usability issue had the highest severity value. A summary of the main usability issues is given for each individual scenario. The audio recordings of the interviews were transcribed verbatim and coded by the master student (TV). A coding scheme was deductively created with three main codes, which were based on the type of questions that are provided in Table 5. For each interview, the answers were marked for each main code. Subcodes were drawn up based on the marked answers. Finally, the definition of the codes was added, and the number of times a subcode occurred in all interviews. After feedback from other researchers, the coding schemes were adapted and used to code all interviews. The data from the Empatica E4 was analysed for heart rate and tension level of each outpatient. For each scenario, it was checked whether there were peaks visible. When a peak was registered, the peak was compared to what was occurred in the screen recordings. A table was created with a description of the occurred event and how high the peak level was.

# 3. Results

## 3.1 Study 1 – Usability test with therapists

All five therapists went through six scenarios of the usability test, which were divided in smaller tasks. In total, there were fifteen usability issues that could be divided into five

categories. In Figure 2 is shown that issues that were related to settings for characters occurred the most, followed by settings for location and creating new templates. For example, therapists could not find or recognize the correct type of virtual characters, such as a security guard. The severity of these usability issues was higher compared to other usability issues. Saving templates and login problems were the least common.

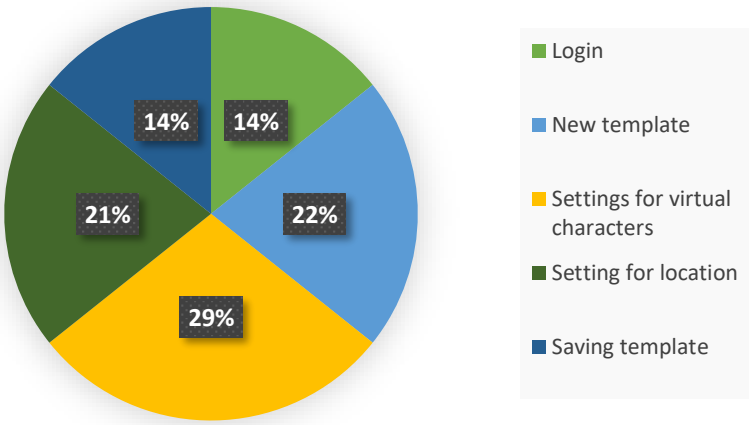


***Figure 2**: Overview of the usability issues from therapists, divided into five categories.*

An overview of the obtained usability issues per scenario is shown in Table 6. The table is divided in different columns, which are displayed the gradings of task criticality, impact, frequency, and severity for each usability issue. Each usability issue in Table 6 is also provided with a description of the issue and where in the dashboard the issue occurred.

***Table 6:*** *Overview of smaller tasks within six scenarios that occurred issues for therapists.*

| ID | Task | Category | Task criticality | Description | Impact | Frequency | Severity[1] |
|---|---|---|---|---|---|---|---|
| **Scenario 1: Login in application** | | | | | | | |
| **1** | Fill in name therapist | Login | 1 | Therapist had to delete "login name" to fill own name | 1 | 1 | 0,2 |
| **2** | Fill in name outpatient | Login | 1 | Therapist had to delete "login name" to fill outpatient's name | 1 | 1 | 0,2 |
| **Scenario 2: Walking around session in bus half full occupied** | | | | | | | |
| **3** | Set profile group occupation | Settings for characters | 5 | The function of this option was unclear, and therapists did not know what the numbers were standing for | 5 | 5 | 25 |
| **4** | Set occupation in bus half full | Location settings | 2 | Therapist set occupation to maximum; Therapist wanted to set occupation with people icons | 3 | 2 | 2,4 |
| **5** | Select profile group with only female characters | Settings for characters | 2 | Therapist did not use "select all" button to select female characters; Some female characters did not look like woman | 1 | 2 | 0,8 |
| **Scenario 3: Saving created template** | | | | | | | |
| **6** | Select access to patient/therapist/everyone | Saving | 3 | For therapist unclear what difference was between "save for therapists" and "save for patient" | 2 | 2 | 2,4 |
| **7** | Fill in description of template | Saving | 1 | Therapist did not see "fill in description" bar | 5 | 1 | 1 |
| **Scenario 4: Role-play in grocery store** | | | | | | | |
| **8** | Selecting characters (security & grocery employee) | Settings for characters | 5 | For therapist was not clear which is the right character because list of characters not clearly displayed with name/function; Position of a character is not clear | 5 | 5 | 25 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **9** | Select location in grocery store with right number of characters | Location settings | 3 | For therapist not clear that the number of characters was displayed in pictures; Therapist though the outpatient also counts as character | 5 | 3 | 9 |
| **10** | Create new template | New template | 1 | Not clear how to create a new template after completing previous one | 3 | 2 | 1,2 |
| **Scenario 5: Role-play in park** | | | | | | | |
| **11** | Create new template | New template | 1 | Not clear how to create a new template after completing one | 3 | 1 | 0,6 |
| **Scenario 6: Adjust previous role-play** | | | | | | | |
| **12** | Select location in shopping street with right number of characters | Location settings | 3 | Filter option for selecting location is not clearly visible | 3 | 1 | 1,8 |
| **13** | Selecting characters (police & old man) | Settings for characters | 3 | Old characters do not look old enough | 1 | 1 | 0,6 |
| **14** | Adjust current template | New template | 1 | Therapist saved current template and started new one | 2 | 1 | 0,4 |

[1] The severity is calculated by the formula: $Severity = \ Criticality\ score * Impact\ score * \frac{Frequency}{total\ participants}$

### 3.1.1 Scenario 1: Login in application

The first scenario, where therapists had to use the login function, was overall successfully completed. Only one therapist had trouble with filling in the name (ID 1 & ID 2). The therapist had to remove the “type your name” indicator that was filled in automatically.

### 3.1.2 Scenario 2: Walking around in bus half full occupied

Therapists were experiencing more usability issues in the second scenario, especially with settings for virtual characters and location. Selecting only female characters was especially difficult and unclear for everyone, which had the highest severity of 25 (ID 3). All therapists stated that the description and function were unclear, which led to ignoring or randomly selecting a value for this option. As a result, not only female characters were appearing, but also male characters. Additionally, issues occurred with settings for locations. Therapists found it difficult to set the occupation of the bus to a maximum of a half (ID 4). Two therapists were not able to use a slide bar to set the occupation of the bus.

### 3.1.3 Scenario 3: Saving created template

Saving created templates in the third task was overall successfully completed. Only one issue occurred during this scenario. One therapist overlooked the option to fill in a description before saving the template (ID 7). As a result, the therapist could not save the template.

### 3.1.4 Scenario 4: Role-play in grocery store

In the fourth session, therapists had to create a role-play session in a grocery store. Therapists were experiencing more usability issues that were mostly related to settings of virtual characters and location. Firstly, three therapists were experiencing difficulties with selecting the right location for role-playing (ID 9). For example, when a virtual environment has been chosen in the third building block, a specific location in the environment can be set in the fourth building block, such as at the entrance of the grocery store. Different options of locations were displayed on the location tab with pictures, which shows the location and the number of virtual characters in VR. However, for three therapists was this unclear. It was unclear that the number of characters were displayed on the pictures, and they assumed that the user wearing the VR glasses also counted as a character.

Second, none of the therapists were able to select the right character, such as the security guard and grocery employee (ID 8). Therapists were looking for a filter to display only characters with professions, but that option did not exist. As a result, none of them were able to pick the security guard and grocery personnel without help of the researcher. It also happened that the characters were standing in the wrong place in the created scenario. In the simulation, the grocery employee was standing next to the counter and the security guard behind the counter, which was not correct if the situation was in real-life. None of the therapists were realizing that without help of the researcher.

### 3.1.5 Scenario 5: Role-play in park

The fifth scenario was especially focused on creating a role-play exercise in a park with a police officer and man wearing a hoodie, which was overall successfully completed. Only one issue, that was more related to new templates, occurred during this scenario. One therapist could still not find the "create new template" button to start a new template after completing the previous one (ID 11).

### 3.1.6 Scenario 6: Adjust previous role-play

The last scenario was overall successfully completed. The number of characters, that is displayed on the location tab with pictures, was still unclear for one therapist (ID 12). The therapist did not see the filter option to display the number of characters on a location. One therapist could not easily find an old man character because old characters do not look old enough (ID 13). One therapist started a new template instead of adjusting the template that was created in the fifth scenario (ID 14).

## 3.2 Study 1 - Interview with therapists

After the usability test, five therapists evaluated what they have experienced while using the dashboard. The main and subcodes that resulted from the answers are provided in Table 7.

***Table 7**: Results of interviews with therapists in code scheme.*

| Main codes | Subcodes | Definition | Total |
|---|---|---|---|
| **Positive points of the dashboard** | First impression | Positive opinion about the usability of the VR dashboard after first use. | 5 |
| | Clear design | Dashboard is easy to use because of logical visualization of options and settings. | 3 |
| | Overview of the settings | A summary of the composite scenario, displayed on the left side of the dashboard, is a good feature which gives the therapists a good overview of the chosen settings immediately. | 2 |
| | Clear overview of triggers | The overview of the list of triggers is clearly visible. | 2 |
| | Easy to learn | The dashboard is overall easy to learn after using it a few times. | 1 |
| **Negative points of the dashboard** | Selecting characters | The function for selecting virtual characters is difficult to understand and not clearly visible. | 3 |
| | Long list triggers | The list of triggers with character sentences is too long. | 3 |
| | Profile group | The function for setting a profile group for a specific scenario is difficult to understand. | 2 |

| | | | |
|---|---|---|---|
| | Number of characters | Unclear to see how many characters are present on a specific location. | 1 |
| | Unclear words | Some words are difficult to understand. | 1 |
| **Suggestions for improvements: existing components** | More character types | Therapists wanted to see more different types of characters to choose from. | 2 |
| | Switch positions of triggers | The long lists of triggers should be placed on the bottom of the dashboard. | 2 |
| | Fold-out option for triggers | A feature to fold-out a specific type of triggers. | 2 |
| | Position of character | Better visualization settings for position of a character on a location. | 1 |
| | Filter for location | Better visualization for filter to select a location with the right number of characters. | 1 |
| | Realistic environment | Virtual environments may look more like the real life. | 1 |
| **Suggestions for improvements: new components** | New filter options for characters | Adding filter options to search for a virtual character with a specific job or virtual characters with a specific length. | 3 |
| | Add behavioural triggers | Adding new triggers that are related to the category behavioural triggers. | 2 |
| | New environments | Adding more new virtual environments. | 1 |
| | Overview created session | A visual overview of the selected options to see how it looks before starting a session. | 1 |
| | Add subgroup triggers | Adding subgroups for each category of triggers for a clearer overview | 1 |
| | Add forensic triggers | Adding more forensic related triggers | 1 |

### 3.2.1 Positive points of the dashboard

This main code referred to all the positive characteristics of the dashboard after using the dashboard according to the therapists. According to three therapists, the application was easy to use because of clear visualization of options and settings [**Clear design**]. Therapist 1 clarified this with:

> *"It is just easy because of the icons that clearly indicate what it stands for."*

Furthermore, two therapists stated that the overview on the left was an incredibly useful feature [**Overview of the settings**]. The overview was designed so that therapists could see which settings have been made. Therapist 5 explained why it was useful:

> *"I also like to see the left overview with the green checkmarks, which indicate if it is set correctly."*

### 3.2.2 Negative points of the dashboard

Therapists have experienced some difficulties while using the dashboard. Three therapists stated again that searching for the right character was difficult [**Selecting characters**]. It was not clearly visual which character was, for example, a security guard or a grocery store employee. Therapist 2 stated:

> *"Searching for those characters is time-consuming. Now I know, for example, where the police officer is, but I didn't find it quickly.*

Two therapists stated once again that settings for profile group were unclear and difficult [**Profile group**]. With profile group, types of virtual characters could be set that will appear in the scenario, such as female characters. Therapist 2 explained that with:

> *"What that crowd will look like is unclear. You don't know in advance how the profile group will look like."*

One therapist was confused about the options that are displayed on the location tab with pictures, which were displaying the number of possible characters that the therapist could set on that specific location [**Number of characters**]. Therapist 3 clarified that with:

> *"I was confused by the security guard that was displayed [on the example screen in the location tab]. But I know now that I have to pay attention to the number of characters."*

### 3.2.3 Suggestions for improvements: existing components

Therapists provided points that could be improved. Firstly, two therapists indicated that more types of virtual characters could be added that are more related to the forensic mental healthcare [**More character types**]. Therapist 5 explained further with:

> *"You can add characters that look more intimidating. I have a client who has a problem with people that look too casual. Maybe there could also be characters in it that look like civil servants."*

The second point was a better visualization of the settings for changing the position of a virtual character in VR [**Position of character**]. For example, the grocery employee was standing next to the counter and the security behind the counter, which was incorrect. These settings to change the position was hidden on the right side of the screen, but therapists stated that it was easy to overlook that. Therapist 1 explained that with:

> *"Perhaps the sidebar where you can switch those characters from place could be clearer. It is still unclear what position a character is in."*

Third, three therapists specifically stated that the list with trigger sentences is too long [**Long list triggers**]. However, therapists have provided points to improve the list of triggers. Two therapists have suggested switching the position of the triggers. The longest list with triggers should be at the bottom, while short lists of triggers should be shown first. Furthermore, two therapists suggested using a fold-out feature for each category of triggers, which if you click on a category, it expands with all associated triggers. Another therapist stated to add subgroups for each category of triggers. Therapist 3 explained that with:

*"I think it's fine without fold-out function, but in subgroups, such as neutral, questions, reactions. You can also make a subgroup with negative comments, for example."*

### 3.2.4 Suggestions for improvements: new components

Therapists stated with different options to improve the dashboard. Two therapists wanted to see a filter option to easily find characters with certain professions [**New filter options for characters**]. The therapist also stated that it could be useful to have a filter for characters with different lengths. Therapist 4 explained that with:

*"Because occasionally you have clients who are tall, and they already look intimidating. Then it sometimes helps to make a role-play with a character of the same height."*

Therapists also indicated that there could be added more virtual environments, such as a construction site, or certain existing environments were too neat [**New environments**]. Therapist 5 stated that with:

*"In terms of environment, I find workplaces a bit limited, such as missing a construction site. Slightly more physical workplaces."*

Lastly, one therapist stated that it could be useful to have an overview of how the created environment will look like before the simulation starts [**Overview created session**]. This way, it is easier to adjust a setting without closing the whole simulation. Therapist 1 explained that with:

*"It would be nice, as a therapist, to see the created concept by yourself, perhaps by clicking a special button in the last step."*

## 3.3 Study 2 – Usability test with outpatients

All five outpatients went through four tasks of the usability test, which were divided in smaller tasks. In total, there were eight usability issues that could be divided into three categories. In Figure 3 is shown that issues that were related to observation in VR occurred the most, followed by issues with the VR device and moving in VR. In general, outpatients were experiencing most difficulties with observations in VR. For example, they indicated that the VR simulation did not look realistic enough. The controls to walk in VR had the highest severity value.

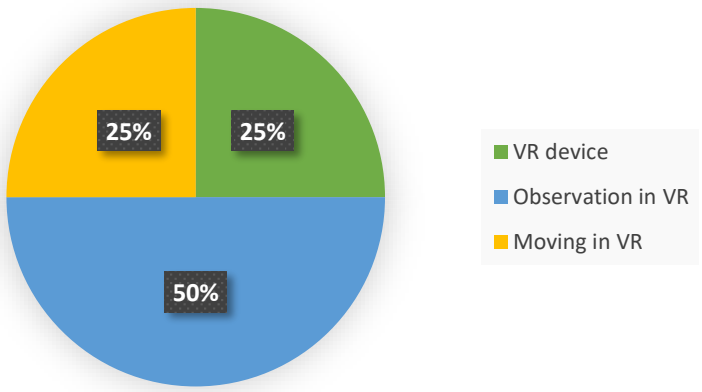


***Figure 3**: Overview of the usability issues from outpatients, divided into three categories.*

An overview of the obtained usability issues is shown in Table 8. The table is divided in different columns, which are displayed the gradings of task criticality, impact, frequency, and severity for each usability issue. Each usability issue in Table 8 is also provided with a description of the issue and where the outpatient is located when the VR-glasses is put on.

***Table 8:*** *Overview of smaller tasks within four scenarios that occurred issues for outpatients.*

| ID | Task | Category | Task criticality | Description | Impact | Frequency | Severity[1] |
|---|---|---|---|---|---|---|---|
| **Scenario 1: Getting used to VR** | | | | | | | |
| **1** | Put VR glasses on first time | VR device | 2 | Outpatient had to get used to VR | 2 | 2 | 1,6 |
| **2** | Looking around in VR first time | Observation in VR | 3 | Surrounding seemed not realistic without details; little dizzy after using VR | 2 | 1 | 1,2 |
| **3** | Walking around in VR using joystick first time | Moving in VR | 0 | No issues occurred | 0 | 0 | 0 |
| **Scenario 2: Walking in shopping street with few virtual characters** | | | | | | | |
| **4** | Put VR glasses and headphone on second time | VR device | 0 | No issues occurred | 0 | 0 | 0 |
| **5** | Looking around in VR second time | Observation in VR | 3 | Virtual characters were walking far-fetched; little dizzy after using VR | 2 | 2 | 2,4 |
| **6** | Walking around in VR using joystick second time | Moving in VR | 5 | The controls were difficult to use; walking speed too slow | 3 | 2 | 6 |
| **Scenario 3: Walking in shopping street with more virtual characters and car sounds** | | | | | | | |
| **7** | Put VR glasses and headphone on third time | VR device | 0 | No issues occurred | 0 | 0 | 0 |
| **8** | Looking around in VR third time | Observation in VR | 3 | Surrounding seemed not realistic enough | 2 | 1 | 1,2 |
| **9** | Walking around in VR using joystick third time | Moving in VR | 5 | The controls were difficult to use; walking speed too slow | 3 | 2 | 6 |
| **10** | Car sounds | Interaction in VR | 0 | No issues occurred | 0 | 0 | 0 |
| **Scenario 4: Role-play with conversation about weather** | | | | | | | |
| **11** | Put VR glasses and headphone on fourth time | VR device | 2 | Looking through VR was blurry without own glasses | 2 | 1 | 0,8 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **12** | Virtual character is standing in front of outpatient | Observation in VR | 3 | Virtual character seemed not realistic enough (2x) | 2 | 2 | 2,4 |
| **13** | Conversation about weather with hand gestures | Interaction in VR | 0 | No issues occurred | 0 | 0 | 0 |

[1] The severity is calculated by the formula: $Severity = \ Criticality\ score * Impact\ score * \frac{Frequency}{total\ participants}$

### 3.3.1 Scenario 1: Getting used to VR

In the first task, the aim was to get the outpatients used to the VR glasses by putting on the glasses on outpatients and letting them walk around in a virtual waiting room. There were no major issues. Outpatients 4 and 5 indicated that it took some time to get used to VR (ID 1) and outpatient 3 reported feeling a little dizzy after completing the first scenario (ID 2).

### 3.3.2 Scenario 2: Shopping Street with few virtual characters

In the second scenario, outpatients had to walk through a shopping street which was quiet with some virtual characters that were walking around. Outpatient 1 indicated that virtual characters did not look realistic (ID 5). Additionally, the biggest issue were the controls to walk in VR (ID 6). Outpatients 2 and 5 indicated that it was difficult to walk around in VR. The outpatients had some difficulties with walking forward with the joystick and steering with the head at the same time. Furthermore, outpatient 2 also indicated that the speed of walking was too slow.

### 3.3.3 Scenario 3: Shopping Street with more virtual characters and car sounds

In the third scenario, outpatients had to walk through the same shopping street again, but there were more virtual characters presented. There were no major issues occurred during this scenario. Outpatients 2 and 5 indicated again that the controls were difficult to use (ID 9). Outpatient 3 indicated that the surroundings did not look realistic enough.

> *"I look at the details and notice that I don't see my limbs. Small details are still missing. You can clearly see that it is a computer image."*

### 3.3.4 Scenario 4: Role-play with conversation

The last scenario was a role-play with a virtual character. This scenario was also divided into different tasks. Outpatient 2 indicated that the image was blurry without his glasses when he put on the VR glasses (ID 11). For outpatients 3 and 5, the virtual character did not look realistic enough (ID 12).

## 3.4 Study 2 - Interview with outpatients

After the usability test, outpatients evaluated what they have experienced while using VR. The codes that resulted from the answers are provided in Table 9. Furthermore, description of the codes has been given and the total of outpatients that stated the provided code in the interview.

***Table 9:*** *Results of interviews with outpatients in code scheme.*

| Main code | Sub code | Description | Total |
|---|---|---|---|
| **Positive points of VR** | Positive first impression | First opinion of the use of VR was positive. | 4 |
| | Identifying issues | VR is a nice intervention to use in treatment to detect forensic related issues. | 1 |
| | | | |
| **Negative points of VR** | Not realistic | Virtual characters and environments did not look like the real world. | 4 |
| | Controls are difficult | Outpatients were not used to these controls. They are normally used to the controls in game consoles. | 3 |
| | VR image blurry | Outpatient did not see clearly through the VR glasses. | 2 |
| | | | |
| **Suggestions for improvements** | Characters more realistic | Virtual characters should look more real to have more effects on outpatients. | 3 |
| | Controls like games | Controls in VR of how to walk should be the same as current game devices, such as PlayStation. | 2 |
| | Walking speed | An option to walk faster in VR. | 2 |

### 3.4.1 Positive points of VR

The first main code refers to all the positive characteristics of using VR according to the outpatients. Four outpatients indicated that they enjoyed using VR [**Positive first impression**]. Moreover, outpatient 5 indicated that this is a good idea to use in treatment. Using VR will make it easier to make things clearer by showing them in VR simulations instead of just a conversation with therapists. Outpatient 5 stated that with:

> *"It's great that you can make things even clearer with VR. Telling is more difficult than showing."*

### 3.4.2 Negative points of VR

Outpatients were also given points that they disliked when using VR. Four outpatients stated that virtual characters and environments did not look realistic enough [**Not realistic**]. Outpatient 2 stated that with:

> *"The movements and appearance of the characters are still unnatural."*

Moreover, the controls were difficult to use [**Controls are difficult**]. For the outpatients, it was unnatural to walk forward with the joystick while steering with your head. As a result, they unknowingly steered in the wrong direction, leading them somewhere they preferred not to go. Outpatient 2 explained that with:

> *"Only the movement is still weird and does not work smoothly because you have to move the joystick forward while you have to move your head to steer. Actually, it should be like in games."*

Two outpatients also indicated that the image of VR sometimes became blurry [**VR image blurry**]. One of them uses glasses in daily life but took them off during the usability test.

### 3.4.3 Suggestions for improvements

Outpatients provided different points to improve the "Triggers&Helpers" application. Two outpatients suggested adding an option to walk faster in VR because the speed of walking around in VR was too slow [**Walking speed**]. Moreover, three outpatients suggested to make virtual characters more realistic [**Characters more realistic**]. Outpatient 1 explained that with:

> *"It does help a bit more if you see a real character in front of you instead of an animated one. I think it will have more impact."*

## 3.5 Study 2 - Physiological arousal of outpatients

The Empatica E4 wearable was used by all five outpatients during the usability test. The recorded physiological data of each outpatient were analysed, which is shown in Table 10. No data were included from outpatient 3, because the Empatica E4 did not connect properly with the "E4 Realtime" app. The table displayed the heart rate and tension level for each outpatient and showed that heart rates of all outpatients remained constant during the whole usability test. Several peaks of the tension level have been detected from the data. For each scenario, a description of "peak moments" has been given, which stated what happened when heart rate or tension levels rose.

What stood out was that the tension level of outpatient 2 constantly rose during the whole usability test. His tension level did not drop since the start of the usability test. However, outpatient 2 stated that his tension level was low. The outpatient particularly indicated that he was irritated by the controls, which was the same moment when his tension level started to rise. Furthermore, it was noticed that the tension level of each outpatient rose when a conversation started with a virtual character.

***Table 10:*** *Recorded physical data of each outpatient, with a description of "peak moments".*

| Scenario | HR (bpm) | Tension level (µS) | Descriptions of peaks |
|---|---|---|---|
| **Outpatient 1** | | | |
| 1 | No data | No data | No peaks detected. |
| 2 | 90 – 100 | 2.4 – 3.5 | No peaks detected. |
| 3 | No data | 2.5 – 4 | Outpatient bumped into virtual character. Tension level briefly rose to 4. |
| 4 | No data | 2.5 – 4 | Tension level rose to 4 when outpatient started conversation with the virtual character and dropped after conversation. |
| **Outpatient 2** | | | |
| 1 | 90 – 100 | 3.0 – 7.8 | Tension level rose while observing the waiting room. The outpatient indicated that it took some time to get used to. |

| | | | |
|---|---|---|---|
| 2 | 90 – 100 | 10 – 17.5 | As soon as the outpatient put on the VR glasses, the tension level rose constantly. |
| 3 | 90 – 100 | 11.5 – 17.5 | Tension level rose constantly during this scenario. The outpatient indicated that he was irritated by the controls. |
| 4 | 90 – 100 | 12.5 – 18.5 | Tension level rose when outpatient started conversation with the virtual character and dropped after conversation. |
| **Outpatient 4** | | | |
| 1 | 70 – 80 | 0.75 – 1.5 | No peaks detected. |
| 2 | 70 – 80 | 1.0 – 2.5 | No peaks detected. |
| 3 | 70 – 80 | 2.5 – 3.5 | No peaks detected. |
| 4 | 70 – 80 | 2.75 – 4.5 | Tension level rose when outpatient started conversation with the virtual character and dropped after conversation. |
| **Outpatient 5** | | | |
| 1 | 80 – 90 | 1.5 – 2.5 | No peaks detected. |
| 2 | 80 – 90 | 2.0 – 4.0 | No peaks detected. |
| 3 | 90 – 100 | 4.5 – 9.5 | Tension level rose constantly during this scenario. The outpatient indicated that he was irritated by the controls. |
| 4 | 80 – 90 | 9.5 – 13.5 | Tension level rose when outpatient started conversation with the virtual character and dropped after conversation. |

# 4. Discussion

The aim of this study was to identify points of improvements of the existing "Triggers&Helpers" application according to therapists, positive and points of improvements of the usability of VR according to patients and possible ways of using wearables with the "Triggers&Helpers" application that can improve current treatment. Several points arose from the two usability tests and interviews. Therapists found it especially important that the "Triggers&Helpers" application should have a clear overview. Therapists indicated that the overview on the left side of the dashboard was useful to see what has already been set and which settings still needed to be set. Despite the overview on the left side of the dashboard, therapists had to use many settings to build a scenario. It sometimes happened that therapists lost track of what they were trying to set and did not exactly know what the scenario would look like. Therapists indicated that this can be solved by adding several filters to the settings, such as selecting professions or locations to speed up the building process. In the second usability test, outpatients were also positive about the use of VR device and the "Triggers&Helpers" application. Most outpatients have seen or used VR before for gaming but indicated that VR could be used in treatment to detect forensic related issues. However, outpatients also indicated in interviews that the controls were difficult to use, because the controls were different than they were used to with gaming. According to outpatients, this can be improved by making the controls the same as in games. A second point that could be improved was the appearance of the virtual characters and the environment. Outpatients indicated that it was not very realistic, but they still wanted to use the "Triggers&Helpers" application because they do feel a presence of people when they walked in the virtual world. The Empatica E4 wearable was used during usability tests with outpatients to mainly monitor heart rates and tension levels. There were no abnormalities in the measured heart rates. The heart rate of all outpatients was constant and there were no outliers. What especially stood out were the findings of tension levels. The Empatica E4 showed changes in physiological measures, which may indicate that an outpatient was irritated, startled, or nervous. An example was that the tension level started to rise when an outpatient bumped into a virtual character.

An important finding that arose, was that therapists found it important that the dashboard of the "Triggers&Helpers" application should be clear and structured. Building scenarios for each forensic psychiatric patient will take a lot of time, especially if the dashboard is not clearly structured or is lacking an overview of how the scenario will look like. Treatment of these forensic psychiatric patients is already regarded as complex due to its diversity and a time-consuming process [6, 7, 36]. Therapists often have several patients, each with an individual problem. Therefore, having a structured dashboard could be more supportive by means of tunneling, which means that therapists should be guided by the dashboard through every building block that is needed to build a

personalized scenario [37]. This way, therapists do not have to work as hard to build every personalized scenario and could focus more on their patients.

Another important point that arose, was that outpatients stated that virtual characters do not look realistic enough. Several outpatients indicated in interviews that the VR simulation had less effect on them, because they noticed that the virtual characters were not real. Although outpatients indicated that they did have the feeling that there was "someone" standing in front of them during role-play, they were reminded again that it was not a real person when they looked closely at the virtual character. However, the Empatica E4 showed that their tension was rising when role-play was started, which could imply that they did feel a presence. A recent study has shown that people can react and behave in a realistic way in VR environments [38]. Experiencing behaviour in VR is correlated with the "sense of presence" [39, 40]. This sense of presence can improve behaviour skills and knowledge transfer, partly because of a situated performance in VR [38]. Because the scenarios in the usability test were neutral, outpatients may have had more freedom to look at details and were therefore not triggered by any forensic content. It could be that if scenarios were more focussed on the individual problems, their attention would be more focused on experiencing and performing behaviour in a forensic related setting. However, more research is required on the added value of the "Triggers&Helpers" application with forensic related scenarios.

A third important finding is related to the possible added value of the Empatica E4 as an interoceptive awareness tool and the possible effectiveness of biofeedback. In this study, for example, the measured tension level was rising when an outpatient virtually bumps into something or someone. One of the assumptions is that these measured physiological data could lead to increases in interoceptive awareness, which will help patients to regulate their physiology better [41-44]. Interoceptive awareness is the ability to sense and interpret physiological signals [45-47]. Several studies have argued that interoceptive awareness is a key component to many interventions and may be the primary mechanism, by which therapists derive benefits from the treatment [48, 49]. One of the barriers of forensic psychiatric patients is that they find it difficult to reflect their behaviour and emotions [4-6]. It becomes even more difficult to reflect when they are focused on the virtual world in which they find themselves. In these cases, the Empatica E4 could support therapists in monitoring patient's physiological data. Outpatients indeed indicated in the second study that virtual characters and environments did not look realistic, but the Empatica E4 showed that they did feel something because their tension increased at certain moments. These moments can be discussed together with the therapist, which could improve their interoceptive awareness. Furthermore, interoceptive awareness has been positively related to decision making, which could be a benefit for both therapists and patients [50]. They could discuss together what they saw in the physiological data when tension was rising, and how the patient could recognize and deal

with it the next time. However, this was just a pilot study with five outpatients and more research needs to be done to confirm these assumptions in a forensic related setting.

## 4.1 Strengths and Limitations

The strength of this study is the use of usability tests and interviews in combination with the think-aloud method in order to gather several new insights. This combination of methods resulted in conclusions and recommendations for the future. Because therapists and outpatients communicated their thoughts orally, an amount of data was collected while they were familiarising themselves with the "Triggers&Helpers" application and VR [51]. The obtained data could be analysed immediately after the usability test and coding schemes were deductively drawn up without further changes. In the interviews, outpatients and therapists were asked about their experiences and ideas in an open, explorative way. Moreover, different stakeholders were often involved as informants in the whole developing process who provided valuable feedback and points that otherwise would have been overlooked by the researchers.

A limitation of this study is the fact that only five outpatients were included. Although sources say that a minimum of five participants is enough to get results from usability tests, this was not enough for this study [31]. All five outpatients came from the same aggression regulation group of Transfore and wanted to participate in the test with VR. Self-selection bias could have influenced the results because outpatients with a possible negative attitude towards VR might not have participated in this study. However, the outpatients were asked in interviews to provide a broader perspective than their own experiences, for example by asking about the opinions and ideas to use VR in treatment for other forensic related issues.

## 4.2 Implications for future research & practice

As a recommendation for further research in the future, the usability tests could be done with multiple outpatients from another group or organization. Despite the self-selection bias limitation, it would be useful to extend the current findings by examining another group with another type of problem to obtain broader insights of the "Triggers&Helpers" application and VR. Furthermore, scenarios for the usability test with outpatients could also be drawn up that is more forensic related, but it must remain ethically responsible. To do this, it must be done in a treatment setting, under the guidance of a trained therapist. The Empatica E4 has been used by five outpatients of the same treatment group, which could be again a self-selection bias. There is only investigated what possibilities the Empatica E4 can offer and whether it has added value for treatments. The findings of the Empatica E4 in this study implied that it could be useful to use wearables with VR, but that was only in a neutral setting. For further research, it is recommended to study more outpatients from another group in order to gain a broader insight of the Empatica E4. This can be done the same as with the "Triggers&Helpers" application. A

usability test and interview combined with think-aloud method could be drawn up that is more focused on the use of the Empatica E4 with both therapists and outpatients. In this way, positive, negative, and points for improvement can be obtained from therapists and outpatients.

# 5. Conclusion

The results of both qualitative studies provide insights into the points of improvements of the "Triggers&Helpers" application and the use of VR according to therapists and outpatients. This study pointed out that therapists found it necessary that the dashboard is structured so that it is faster and easier to use. The VR scenarios could be designed more realistically according to the outpatients. However, the usability test was only conducted with five outpatients from the same treatment group and consisted of only neutral scenarios. For further research, more outpatients from other groups could be included and scenarios could be built that are more related to forensic mental healthcare. The use of a wearable could have an added value in treatment as an interoceptive awareness tool together with the "Triggers&Helpers" application, but more research is needed. Therapists and more outpatients could be involved in the future to elicit their opinion that is more focused on the use of wearables in treatment. All in all, applying VR in combination with wearables has the potential to improve the current forensic mental healthcare by beneficially supporting both the patient and the therapist as a interoceptive awareness tool.

# References

[1] Ronald C. Kessler, et al. 2007. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry 6(3), 168−176. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2174588/

[2] Arboleda-Florez, J. (2006). *Forensic psychiatry: contemporary scope, challenges and controversies.* World Psychiatry, 5(2), 87.

[3] Drieschner, K. H., & Boomsma, A. (2008). Validation of the treatment motivation scales for forensic outpatient treatment (TMS-F). Assessment, 15(2), 242-255.

[4] van Gemert-Pijnen JEWC, Kip H, Kelders SM, Sanderman R, Kelders SM, Kip H. (2020). Introducing eHealth. In: van Gemert-Pijnen JEWC, editor. eHealth Research, Theory and Development: A Multi-Disciplinary Approach. Abingdon, UK: Routledge.

[5] Kip, H. (2021). The added value of eHealth in treatment of offenders: Improving the development, implementation and evaluation of technology in forensic mental healthcare. University of Twente. https://doi.org/10.3990/1.9789036551311

[6] Deenik, J., Tenback, D. E., Tak, E., Blanson Henkemans, O. A., Rosenbaum, S., Hendriksen, I., & van Harten, P. N. (2019). Implementation barriers and facilitators of an integrated multidisciplinary lifestyle enhancing treatment for inpatients with severe mental illness: the MULTI study IV. BMC health services research, 19(1), 740. https://doi.org/10.1186/s12913-019-4608-x

[7] Franke, I., Vogel, T., Eher, R., & Dudeck, M. (2019). Prison mental healthcare: recent developments and future challenges. Current opinion in psychiatry, 32(4), 342-347.

[8] Landenberger NA, Lipsey MW. The positive effects of cognitive-behavioral programs for offenders: a meta-analysis of factors associated with effective treatment. J Exp Criminol (2005) 1(4):451–76. doi:10.1007/s11292-005-3541-7

[9] Andrews DA, Bonta J. Rehabilitating criminal justice policy and practice. Psychol Public Policy Law (2010) 16(1):39. doi:10.1037/a0018362

[10] Babcock JC, Green CE, Robie C. Does batterers' treatment work? A metaanalytic review of domestic violence treatment. Clin Psychol Rev (2004) 23(8):1023–53. doi:10.1016/j.cpr.2002.07.001

[11] van Gemert-Pijnen JEWC, Kip H, Kelders SM, Sanderman R, Kelders SM, Kip H. (2020). Introducing eHealth. In: van Gemert-Pijnen JEWC, editor. eHealth Research, Theory and Development: A Multi-Disciplinary Approach.

[12] Kip H, Bouman YH, Kelders SM, van Gemert-Pijnen LJ. eHealth in treatment of offenders in forensic mental health: a review of the current state. Frontiers in psychiatry. 2018;9:42.

[13] Greenberg E, Dunleavy E, Kutner M. Literacy Behind Bars: Results from the 2003 National Assessment of Adult Literacy Prison Survey. NCES 2007–473. Washington, DC: National Center for Education Statistics (2007).

[14] Levesque DA, Ciavatta MM, Castle PH, Prochaska JM, Prochaska JO. Evaluation of a stage-based, computer-tailored adjunct to usual care for domestic violence offenders. Psychol Violence (2012) 2(4):368–84. doi:10.1037/a0027501

[15] Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: a review of research in virtual reality. Frontiers in psychology, 6, 26.

[16] Bordnick, P. S., Graap, K. M., Copp, H. L., Brooks, J., & Ferrer, M. (2005). *Virtual reality cue reactivity assessment in cigarette smokers.* CyberPsychology & Behavior, 8(5), 487-492.

[17] Pericot-Valverde, I., Germeroth, L. J., & Tiffany, S. T. (2016). *The use of virtual reality in the production of cue-specific craving for cigarettes: a meta-analysis*. Nicotine & Tobacco Research, 18(5), 538-546.

[18] Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., & Slater, M. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychological medicine, 47(14), 2393-2400.

[19] Guillén, V., Baños, R. M., & Botella, C. (2018). Users' opinion about a virtual reality system as an adjunct to psychological treatment for stress-related disorders: A quantitative and qualitative mixed-methods study. Frontiers in psychology, 9, 1038.

[20] Turner WA, Casey LM. Outcomes associated with virtual reality in psychological interventions: where are we now? Clin Psychol Rev. 2014 Dec;34(8):634-44. doi: 10.1016/j.cpr.2014.10.003. Epub 2014 Oct 18. PMID: 25455627.

[21] Fromberger, P., K. Jordan, and J.L. Müller, Anwendung virtueller Realitäten in der forensischen Psychiatrie. Der Nervenarzt, 2014. 85(3): p. 298-303.

[22] Alley, S., Jennings, C., Plotnikoff, R. C., & Vandelanotte, C. (2016). Web-based video-coaching to assist an automated computer-tailored physical activity intervention for inactive adults: a randomized controlled trial. Journal of medical Internet research, 18(8), e223.

[23] Kip, H., Oberschmidt, K., Bierbooms, J., Dijkslag, D., Kelders, S., & Roelofsen, B. (2019). *Technologie in de forensische zorg: Crossing borders*. Kwaliteit Forensische Zorg.

[24] Kip H, Oberschmidt K, Bierbooms J, Dijkslag D, Kelders S, Roelofsen B. Technologie in de forensische zorg–Crossing borders. Kwaliteit Forensische Zorg. 2019;1.

[25] Kip H, Oberschmidt K, Bierbooms JJPA. eHealth Technology in Forensic Mental Healthcare: Recommendations for Achieving Benefits and Overcoming Barriers. International Journal of Forensic Mental Health. 2020:1-17.

[26] Bierbooms J, Bouman Y, Dijkslag D, Kimpen R, Muller J, Wieske R. Do's en don'ts van e-health in de forensische ggz. Kwaliteit Forensische Zorg (KFZ). 2015.

[27] Scaife M, Rogers Y, Aldrich F, Davies M, editors. Designing for or designing with? Informant design for interactive learning environments. Proceedings of the ACM SIGCHI Conference on Human factors in computing systems; 1997.

[28] Beerlage-de Jong N. eHealth vs. infection: Participatory development of persuasive eHealth to support safe care. 2016.

[29] Burns C. Human-centred design. eHealth Research, Theory and Development: Routledge; 2018. p. 207-27.

[30] Kip, H. (2021). The added value of ehealth in treatment of offenders. Gildeprint.

[31] Nngroup | Why You Only Need to Test with 5 Users. (2000). https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/. Retrieved 2021-06-06.

[32] Jaspers, M. W. (2009). A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. *International journal of medical informatics*, *78*(5), 340-353.

[33] Årsand, E., & Demiris, G. (2008). User-centered methods for designing patient-centric self-help tools. *Informatics for health and social care*, *33*(3), 158-169.

[34] Toptal | *Turning Usability Testing Data into Action without Going Insane*. (2021). https://www.toptal.com/designers/usability/turning-usability-testing-data-into-action#:~:text=Typically%2C%20a%20usability%20test%20involves,a%20walk%20in%20the%20park. Retrieved 2021-04-06.

[35] Biopac systems | EDA introductory guide. (2015). https://www.biopac.com/wp-content/uploads/EDA-Guide.pdf. Retrieved 2021-05-21.

[36] Thieme, A., McCarthy, J., Johnson, P., Phillips, S., Wallace, J., Lindley, S., ... & Olivier, P. (2016, May). Challenges for designing new technology for health and wellbeing in a complex mental healthcare context. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 2136-2149).

[37] Purpura, S., Schwanda, V., Williams, K., Stubler, W., & Sengers, P. (2011, May). Fit4life: the design of a persuasive technology promoting healthy behavior and ideal weight. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 423-432).

[38] Martirosov, S., & Kopecek, P. (2017). Virtual reality and its influence on training and education-literature review. *Annals of DAAAM & Proceedings*, *28*.

[39] Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, *3*, 74.

[40] Riva, G. (2011). Presence, actions and emotions: A theoretical framework. *Journal of CyberTherapy and Rehabilitation*, *4*(2), 204-206.

[41] Blascovich, J., & Mendes, W. B. (2010). Social psychophysiology and embodiment.

[42] Cameron, O. G. (2001). Interoception: the inside story—a model for psychosomatic processes. *Psychosomatic medicine*, *63*(5), 697-710.

[43] Chentsova-Dutton, Y. E., & Dzokoto, V. (2014). Listen to your heart: The cultural shaping of interoceptive awareness and accuracy. *Emotion*, *14*(4), 666.

[44] Mehling, W. E., Price, C., Daubenmier, J. J., Acree, M., Bartmess, E., & Stewart, A. (2012). The multidimensional assessment of interoceptive awareness (MAIA). *PloS one*, *7*(11), e48230.

[45] Füstös, J., Gramann, K., Herbert, B. M., & Pollatos, O. (2013). On the embodiment of emotion regulation: interoceptive awareness facilitates reappraisal. *Social cognitive and affective neuroscience*, *8*(8), 911-917.

[46] Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K., & Critchley, H. D. (2015). Knowing your own heart: distinguishing interoceptive accuracy from interoceptive awareness. *Biological psychology*, *104*, 65-74.

[47] Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nature reviews neuroscience*, *3*(8), 655-666.

[48] Mehling, W. E., Price, C., Daubenmier, J. J., Acree, M., Bartmess, E., & Stewart, A. (2012). The multidimensional assessment of interoceptive awareness (MAIA). *PloS one*, *7*(11), e48230.

[49] Fissler, M., Winnebeck, E., Schroeter, T., Gummersbach, M., Huntenburg, J. M., Gaertner, M., & Barnhofer, T. (2016). An investigation of the effects of brief mindfulness training on self-reported interoceptive awareness, the ability to decenter, and their role in the reduction of depressive symptoms. *Mindfulness*, *7*(5), 1170-1181.

[50] Lux, E., Hawlitschek, F., Adam, M. T., & Pfeiffer, J. (2015). Using live biofeedback for decision support: Investigating influences of emotion regulation in financial decision making.

[51] Bolle, S., Romijn, G., Smets, E. M., Loos, E. F., Kunneman, M., & van Weert, J. C. (2016). Older cancer patients' user experiences with web-based health information tools: a think-aloud study. *Journal of medical Internet research*, *18*(7), e5618.