Creating And Evaluating A Virtual Reality Environment For Studies Into Social Anxiety Disorders

Daan van der Veen, 6591965, d.vanderVeen@students.uu.nl

Utrecht University

September 5, 2024

Abstract

Social Anxiety Disorder(SAD) is a mental disorder rapidly growing in prevalence. At the time of writing this paper, slightly more than 1 in 20 people is afflicted with SAD. SADs most often start to occur in a person's childhood, at which it is almost most effective to be treated. To make sure that it can be treated and identified a lot of research is done into SADs. However, performing research in the real world is difficult as it needs a lot of actors and risks the participants experiencing negative emotions. To solve these problems Virtual Reality (VR) is used to experiment in. Most VR environments are created by independent companies for a high price and can rarely be adapted to a researcher's own needs. This study has created and evaluated the usability and effectiveness of a Virtual Reality (VR) tool designed to assist researchers with no need for IT skills in setting up experiments for individuals with SAD. The research involved a thematic analysis of user feedback collected during the setup and execution of VR-based experiments. Key findings highlight that while the tool's interface is generally intuitive, issues related to user experience, such as unclear UI elements and limited control over experimental settings, were identified. Participants appreciated the ease of configuration but expressed a desire for more detailed options, particularly concerning the customization of Non-Playable Characters (NPCs) and the overall experimental environment. The immersive qualities of the VR scenarios were praised, though certain aspects, such as the lack of environmental details and inconsistencies in NPC behaviour, detracted from the overall experience. Additionally, the data analysis component was found to be accessible but could benefit from more comprehensive outputs and clearer documentation. The study concludes that while the VR tool has significant potential for SAD research, further refinement is needed to address the identified usability and functionality issues. Future work should focus on enhancing the tool's features, improving the clarity of its interface, and ensuring the reliability of the VR environment to better support researchers in conducting meaningful experiments.

1 Introduction

People often have to deal with mental disorders in their lives, either their own or those of family or friends. Around 15% of the people in the Netherlands have a mental disorder (ten Have et al. (2023)), given that this is such a large group, it is important to understand how to both research and treat this disorder. One of these disorders rapidly growing in prevalence is social anxiety disorder (SAD), which makes up slightly over a third of the mental disorder cases at 5.2% of the populace of The Netherlands(ten Have et al. (2023)). SADs are described as follows by the National Institute of Mental Health: "Social anxiety disorder is an intense, persistent fear of being watched and judged by others" (U.S. Department of Health and Human Services (2022)). It affects things like speaking in public and dating but also more mundane things like eating or going to a restroom in public spaces (U.S. Department of Health and Human Services (2022)). Often SADs start to occur during adolescence (Kessler et al. (2005)). They found that 90% of SADs have occurred by the age of 23. It is quite uncommon for SADs to show up in very young children, but it gets progressively more common the older children become (Ranta et al. (2012)), having the mean being at age 13 (WITTCHEN et al. (1999)). Common signs of SADs when being around or interacting in public environments are rapid heart rate, sweating, trembling, stomach ache, stiff, soft voice, inability to hold eye contact and many more similar things. SAD is often treated through Cognitive Behaviour Theory (CBT) and medicine. even though medicine can be a more immediate help

than CBT, in the long term CBT vastly outperforms most medicine regarding SAD relapse (Zaider and Heimberg (2003)). CBT has multiple styles, these are exposure, cognitive restructuring, social skills training and applied relaxation. Of these styles exposure is the most commonly researched and often applied to help individuals with SAD (Rodebaugh et al. (2004)). A newer technology has been researched and applied in exposure therapy: Virtual reality.

Virtual Reality (VR) has emerged as a transformative technology with the potential to revolutionize various aspects of human experience, ranging from entertainment and gaming to education, healthcare, and beyond. Defined as a computer-generated simulation of an interactive 3D environment, VR enables users to immerse themselves in synthetic worlds, experiencing sights, sounds, and sensations that mimic real-life scenarios. Over the past few decades, VR has evolved from a niche concept into a mainstream phenomenon, thanks to advancements in computing power, graphics rendering, and motion tracking technologies.

Currently, VR is already widely used in psychology as a form of treatment for phobias, slowly getting people used to dealing with their phobia without having to interact with it. Two examples are Virtual reality exposure theory (VRET) and the previously mentioned CBT.

There is already research in the place of VR to do research into social anxieties. for example, the study about the impact of ostracism on social anxiety of Rubo and Munsch (2024). There are also studies about how to increase the realism of the virtual reality environment. The paper Bergsnev and Sánchez Laws (2022) is looking into the value of Personalizing a virtual environment for a user. Or the difference between 360-degree videos and a 3D environment (Nason et al. (2019)).

However, alongside its promise, VR also presents challenges and considerations, ranging from technical limitations and ethical concerns to issues of accessibility and inclusivity. As VR continues to evolve and increase, it is imperative to critically examine its implications for society, culture, and individual well-being.

In this Thesis, a virtual reality environment in which research can be done into the behaviour of people with SADs has been created and tested. The purpose of the tool is to provide an accessible alternative to expensive company products and won't require experience with programming or tools like Unity. The tool should be both easy to create experiments in and automatically collect interesting data. The following research question was formed from the mentioned goals "Does the created tool provide an easy environment for researchers, without programming experiences, to create an experiment and obtain understandable and useful data?". This research question was divided into 2 smaller research questions. Research Question 1 (RQ1) is "The VR Tool will be able to be used by the researcher to set up an experiment without any additional help" and Research Question 2 (RQ2) is "The data acquired by the tool can be navigated and worked with, using only the help of the textbook snippets".

Through a comprehensive review of existing literature, an empirical study to define the requirement for the tool and a final experiment in which the functionality of the tool is put to the test, this thesis aims to lay the foundation for creating a tool that can be used for research into SADs. There is also a list of future improvements, either needed or highly recommended, for the improvement of the tool.

2 Literature Review

2.1 Social Anxiety

2.1.1 Impact Of Social Anxiety On Social Relationships

Social Anxiety Disorder (SAD) significantly impacts social relationships, including marital status and friendships. Research by Hart et al. (1999) highlights that socially anxious individuals are less likely to be married and have a higher breakup rate compared to those with lower levels of social anxiety. This phenomenon often stems from the strain placed on relationships due to the individual's anxiety, leading to perceived imbalance by their partner. However, Hart et al. (1999) also found that married individuals exhibited lower levels of social anxiety than their single counterparts, suggesting a potential protective factor in the context of marriage.

Furthermore, Porter and Chambless (2017) found that social anxiety contributes to difficulties in maintaining relationships, with socially anxious individuals having fewer friends and lower-quality friendships. This decline in friendship quality is attributed to the avoidance of social gatherings, limiting opportunities for social interaction and intimacy (Biggs et al. (2011)). The cumulative impact of social anxiety on various aspects of social relationships underscores the importance of understanding and addressing this condition.

2.1.2 Social Anxiety Scales

In the field of social anxiety research, significant efforts have been directed towards the development of questionnaires designed to measure levels of social anxiety in individuals. These instruments, commonly referred to as Social Anxiety Scales (SAS), serve as valuable tools for assessing the severity and impact of social anxiety symptoms. Among the various variants of SAS, prominent examples include the Liebowitz Social Anxiety Scale (LSAS), the Social Avoidance and Distress Scale (SADS), and the Social Interaction Anxiety Scale (SIAS), among others.

The LSAS is a 24-item scale which is split into 13 performance-related questions and 11 social situation questions (Center (2023)). The scale is most often used in clinical trials and as an effectiveness rating of a cognitive behaviour therapy (CBT). It is also acknowledged as a self-report scale.

The SADS focuses on measuring distress, discomfort, fear and avoidance in social scenarios (Watson and Friend (1969)). The questionnaire is 28 items long and is built up of true and false questions. scoring high on the scales often is connected to individuals with low self-confidence, a low need for social affiliation and a higher need for social deference.

The SIAS is made up of 20 5-point Likert scale questions. Its main function is as a self-report scale used for tracking treatment progress (Herbert et al. (2014)). The score of this scale is correlated to the social anxiety score of the person and shows highly similar outcomes to the Social Phobia and Anxiety Inventory (SPAI).

2.1.3 Exposure Theory

Exposure therapy (ET), a key treatment for social anxiety disorder (SAD), systematically exposes individuals to feared social situations to reduce anxiety responses. Rooted in cognitive-behavioural therapy (CBT), exposure therapy aims to habituate individuals to feared stimuli and challenge maladaptive beliefs. Research explores various modalities, including in vivo and virtual reality exposure, both effective in fear reduction and social functioning improvement. Cognitive restructuring techniques complement exposure, targeting negative thoughts and fostering realistic perspectives. A newer alternative way to ET might be game streaming as this simulates social interaction, exposing individuals to potential audience scrutiny (Frommel et al. (2021)). As it provides a controlled environment for tailored exposure, facilitating gradual desensitization. Streaming gameplay introduces performance pressure akin to social evaluation concerns, intensifying anxiety. Individuals fear judgment for their gameplay, mistakes, or social awkwardness, amplifying anxiety. Moreover, game streaming offers social support and reinforcement from viewers, enhancing confidence and coping skills (Frommel et al. (2021)).

2.2 Social Anxiety In VR

2.2.1 Virtual Reality Exposure Therapy

To research SADs it is needed to understand how to create an environment in which participants experience this. This is exactly the purpose of Virtual Reality Exposure Therapy (VRET), where the participants get put into scenarios to experience different types of SAD of varying degrees of intensity. For example, Coelho et al. (2009) examined the possibilities of Virtual reality exposure theory (VRET) for acrophobia (fear of heights). They found advantages to be lower cost, time investment, and social pressure removal. In a study about VRET for arachnophobia, a fear of spiders, 89% of the participants mentioned that they would rather face their fear in a virtual environment than in a real-life setting (Garcia-Palacios et al. (2001)). This shows that performing research in VR might lessen the emotional impact that confronting user with their fears normally has, making it more morally acceptable. The paper by Arnfred et al. (2023) combines multiple studies using VR in combination with phobias, of which 21 studies were for SAD. 17 of these were implemented in a rendered Virtual environment of differing types of social activities (presentations, riding the bus, job interviews). In these studies, the three most effective ways to influence the "difficulty" of the environment were the amount of NPCs, the gender of the NPCs and the attitude (Bored, Neutral, Excited). Similar results were obtained by Emmelkamp et al. (2020), finding the same ways of controlling the intensity of the environment. They additionally found the importance of verbal interaction with NPCs and concluded that the interactivity of the environment is positively connected to the perceived level of immersion. These findings provide important variables to be controlled in the tool for the researcher to be able to adapt to the intensity of the created environment.

2.2.2 VRET For Adolescents

SADs most often occur at a relatively young age (Kessler et al. (2005)). Wong Sarver et al. (2013) researched a tool for treating children with SADs with the question if there is a place for such a

tool. Even though the tool has the potential to be used for adolescents, it needed altercations that were not needed for adults. One of these problems is boredom, adolescents have a lower interest in performing the tasks showing the importance of external features to keep the content engaging. Another study looking at the feasibility of VR exposure as an assessment and treatment modality for youth with SAD found that virtual environments simulating parties and public speaking created significantly more distress than neutral environments, showing that VR anxiety also occurs in VR (Parrish et al. (2016)). They also found that even though the level of immersion was acceptable, the environments did not fully connect to the experiences of the children as the environment was often made from the viewpoint of an adult dealing with SAD. These studies show that there is value and room to research the topic further as the requirements for children do differ from those of adults.

2.2.3 Fear Of Public Speaking VR Treatment

Fear of public speaking, as the name suggests, is the fear of talking/presenting to a group of people. There is quite an old study on using VR to help with this fear. The study of North et al. (1998) compared two groups of which one group had a public speaking VR setting and the other group a trivial VR setting. The participants showed similar symptoms for public speaking in the VR setting as in the real world setting, which being an increase in heart rate, sweating, dry mouth and feelings of discomfort. Just as in real-life treatment, extended experience in VR environments helped in lowering the level of perceived anxiety. The lowered perceived anxiety in the VR setting also translated to real-world public speaking, showing that there is a use for VR in treating fear of public speaking. (Slater et al. (1999), Pertaub et al. (2001), Pertaub et al. (2002)) Three similar studies were performed on the effect of the behaviour of a virtual audience on the perceived anxiety of a speaker in a VR setting. The audience showed positive, neutral or negative behaviour and in all three studies, a non-neutral audience increased the perceived anxiety, with a negative audience increasing it the most. This proves the importance of the NPCs reacting to the User as well as the influence of the emotions the NPCs express.

2.3 Variables For Immersion Control In VR

VR is used in many more fields than only in the treatment or research of SADs and other phobias. These other fields might light up additional features important to increase the perceived immersion. Studies about the importance of immersion have found mixed results for the effects on treatment. However, in these studies, a positive correlation is found for the effect of immersion on perceived social anxiety (Price and Anderson (2007)). Given that the tool is for research and not for treating participants, immersion is an important factor in making sure that the participants experience anxiety as realistically as possible Servotte et al. (2020). At the moment questionnaires to evaluate the experience in an immersive VR world are used to evaluate how immersive a VR environment is, for example, the questionnaire created by Tcha-Tokey et al. (2016). This questionnaire asks about features the user experiences in a VR environment, like the responsiveness of the environment, experience of time and emotions and many more. Selzer and Castro (2023) have created a metric for determining the immersion of a VR environment, they have done this using all manners of variables that are found in a VR environment and adding weights to them. Some variables that can lead to improvement of immersion are the type of locomotion, sharpness of visuals and audio types, as follows from Selzer and Castro (2023). These features were taken into account when designing the tool to identify the types of controls that are beneficial, differences of importance between different aspects of the environment and features that were useful to add to the tool.

2.3.1 NPC Emotions

The emotional state expressed through facial expressions and type/tone of speech seems to affect gaze avoidance for people with a social anxiety disorder following Weeks et al. (2013). They compared the difference between gaze avoidance between neutral, happy and angry clips through covert eye-tracking. For the happy as well as the angry cases participants have shown a significantly lower amount of time until they looked away. This is proved even further by McTeague et al. (2011), this study was on four different emotions on which they measured electrocortical dynamics between a control group consistent of people with low to no social anxiety and an experimental group consisting of people with a high social anxiety value. The outcome of the study has been that high social anxiety people will perceive a higher level of anxi-

ety when emotional facial expressions are shown in pictures. Whereas, people with low social anxiety did not perceive a significant difference between neutral pictures and those showing emotions. In these studies, emotions were primarily indicated through the mouth and eyebrows, for example, scrunching eyebrows and an upside-down smile to indicate a negative emotional state.

2.3.2 Perspective

In a virtual environment, you either view the environment through the avatar or from behind the avatar, this is the difference between the first-person perspective (fpp) and the third-person perspective (tpp). It is often assumed that to be able to feel a connection to a body you need to be in the fpp. However, Liou et al. (2023) put this assumption to the test and compared bodily function using a haptic sensor between fpp and tpp, leading to finding no significant difference between the two perspectives. Either perspective would function equally well to provoke a sense of control over the body.

2.3.3 Avatar Customization

When you are walking around in an environment you often control a character or avatar. This avatar can look like a generic person or you can be given the option to change characteristics like hair style and colour, gender and many other things. Customizing the avatar, especially when made to look like the player, increases how well the player identifies as the character (Rheu et al. (2022),Safran et al. (2024)) The behaviour a player shows can be affected by the acts the avatar performed. Ratan and Sah (2015) found a positive interaction between avatar embodiment and performance by the player after the avatar has performed the same action. Rheu et al. (2022) found that when an avatar embodies the player the way the player controls the avatar changes, prioritizing actions that would look similar to what the player would do in the situation. So, letting a player customize an avatar would most likely lead to the actions in the virtual space being more representative of how the actions would be performed in the real world.

2.4 Gaze Tracking

A lot of studies have been done about gaze avoidance by people with social anxiety (Fernandes et al. (2017), Howell et al. (2015), Weeks et al. (2013), Weeks et al. (2019)) Most focus on a one-on-one interview through a webcam or in person while the participant's eye movement is tracked. In Weeks et al. (2019) participants are conversing in pairs of two through a webcam, they are not told their sight is being tracked. Their findings correlated with the widely accepted notions that people with symptoms of social anxiety are more likely to avert eye contact and extended this notion by finding that a participant's self-evaluation of having social anxiety is also inversely correlated to eye contact duration. However, they found that "neither post-task state social anxiety nor self-perception of performance was significantly related to eye contact behaviour". This shows the importance of eye tracking as the participant's self-evaluation does not accurately represent how anxious a person was during the conversation. Similar findings can be found in the study of Weeks et al. (2019), where they found that participants with Severe Anxiety Disorder (SAD) held eye contact shorter, made eye contact less often and reported they felt a higher degree of anxiety over healthy participants. The study used 13 positive and 13 negative videos of actors giving statements to the participants. Both positive and negative videos have shown there to be a significant difference in all three findings between SAD and healthy participants. These findings lead to the belief that gaze tracking can be a good identifier of a person's perceived anxiety and being able to track this can help in researching Social Anxiety.

2.5 Unity Experiment Framework

To study human behaviour in a VR setting, the Unity Experiment Framework(UXF) tool was created by Brookes et al. (2019). UXF is a comprehensive tool that leverages the capabilities of Unity to streamline the creation, management, and analysis of experiments. This framework provides researchers with a robust, user-friendly environment that facilitates the setup of complex experimental protocols, efficient data collection, and seamless integration of diverse experimental components. UXF achieves this through several key features: an intuitive configuration system for defining experimental parameters, automated data handling for accurate and consistent data collection, and a flexible event system for precise control over experimental events and stimuli presentation. These features are designed to support a wide range of experimental designs, from simple reaction time tasks to intricate virtual reality (VR) and augmented reality (AR) scenarios.

From related works, a lot of important features can be obtained for influencing Social Anxiety and increasing the perceived immersion. Features such

as amount, visual expression of NPCs and verbal interaction. Other features such as real-world walking and interactive NPCs can boost the perceived immersion. Features that can help to identify Social Anxiety, like gaze tracking, have already been determined by previous research. However, many of these researches mention limitations that the current environment did not allow for further investigation and most findings are manually collected by the researcher instead of the environment. The VR tool that has been created resolves such problems by providing researchers with an easily adaptable environment which also provides automatically collected data. This makes it less likely that the functionality of the tool will limit researchers. To make sure the VR tool fulfils the needs of those researchers a preliminary user study was performed to define the wishes of these researchers. From these findings, the tool was created and tested to make sure it did indeed fulfil the goal of being an easy manipulatable VR environment, which provides researchers with enough options to manipulate the environment to their wishes.

3 Preliminary User Studies

To better understand user preferences and priorities in altering parameters, recording measurements, and determining non-player character (NPC) features/behaviour, a preliminary user study was conducted. The study aimed to gather insights into the importance of these aspects from the perspective of potential users. The purpose of the tool is to provide an environment in which researchers can easily set up experiments into SADs, as such it was important to involve potential users in the development to better understand the needs and wants of these people. Through a combination of qualitative and quantitative questions that were formed based on the Literature review(2), insight was obtained that was used to create the initial design and MoSCoW for the tool. The questionnaire can be found in Appendix A

3.1 Participants

A total of 10 participants were recruited for the study through acquaintances and recommendations. The participants had backgrounds and experiences in either Psychology or Social and Behavioral Sciences.

3.2 Procedure

Participants were presented with a series of questions and scenarios related to altering parameters, recording measurements, and defining NPC features. Each participant was asked to rate the importance of specific parameters to be altered, measurements to be recorded, and features for NPCs on either a Likert scale ranging from 1 (Not Important) to 5 (Very Important) or choose all that apply style questions. Additionally, participants were encouraged to provide qualitative feedback to elaborate on their ratings and suggest any additional parameters or features they deemed significant.

3.3 Results

3.3.1 Importance of Altered Parameters

Participants rated the importance of altering various parameters on a Likert scale, of which the results can be found in Table 1. Preliminary analysis revealed that altering the mood and conversations of the NPC were rated higher than 4 with a low variance, indicating a high priority for this aspect. Conversely, the room layout and looks of the NPCs were rated below 3.5 suggesting a lower priority. Altering the number of NPCs scored on average 3.67 with high variance, meaning that the importance is between the other 2 groups.

3.3.2 Importance Of Recorded Measurements

Participants also provided ratings for the importance of recording measurements in the virtual environment, of which the results can be found in Table 2. Initial findings demonstrated that participants considered recording "Eye tracking", "NPC Characteristics" and "The conversation transcript" as "Quite Important," underscoring the significance it might have in research. A lower importance was given to "Distance to NPC" and "Time spent on things". "Total distance travelled" had the lowest importance of the offered features.

3.3.3 NPC Features

In assessing the features deemed essential for NPCs, participants were asked to prioritize aspects such as clothing styles, gestures, and expressions, the results can be found in Table 3. Analysis of responses indicated that 100% of participants highlighted Facial expressions and Gender as a crucial aspect, while 90% emphasized the importance of body type for enhancing NPC diver-

	Minimum	Maximum	Mean	Std deviation	Variance
the amount of NPCs	3.00	5.00	3.67	0.94	0.89
the mood of NPCs	4.00	5.00	4.33	0.47	0.22
the room layout	2.00	4.00	3.00	0.82	0.67
NPC looks	3.00	4.00	3.33	0.37	0.22
NPC conversation	4.00	4.00	4.00	0.00	0.00

Table 1: Preliminary user study results: Importance of Altered Parameters

	Minimum	Maximum	Mean	Std deviation	Variance
Distance to NPC	2.00	5.00	3.67	1.25	1.56
Distance travelled	2.00	4.00	3.33	0.94	0.89
Time spend on things	3.00	5.00	3.67	0.94	0.89
Eye tracking	4.00	5.00	4.33	0.47	0.22
NPC Characteristics	4.00	5.00	4.33	0.47	0.22
Conversation Transcript	4.00	4.00	4.00	0.00	0.00

Table 2: Preliminary user study results: Importance of Recorded Measurements

sity. 70% of the participants felt that head gestures like nodding and shaking could be important, whereas hand gestures were slightly less important for 60% of the participants. There were also 60% of the participants who found skin colour important. And both clothing and hair colour and style were rated least important at only 40%.

3.4 Qualitative Feedback

Qualitative data from open-ended questions provided valuable insights into participants' reasoning behind their ratings and additional considerations beyond the predefined parameters. Common themes included combining walking in the real world and VR world, a preference for first-person perspective, a high-importance emotional expression of NPCs and posture tracking, which have been integrated into the VR tool.

4 The VR tool

The VR tool was designed to facilitate experimental research in social psychology and human behaviour. The environment is structured into three main sections: the Settings room, the Practice Room, and the Experiment Scene, ensuring a comprehensive and seamless participant experience.

4.1 The User

Based on the preliminary user study results(3.3), decisions have been made for the functionality of the user-controlled character. The character will be in first-person view and be able to walk around using either the joysticks or by walking around in real life. Given that the first-person view had been chosen there was no need to add the option to be able to customize the user-controlled character. The user will interact with the NPCs based on proximity and voice level detection. The user needs to use the controller to interact with the settings.

4.2 Settings Room

The settings room(Figure 1) is the first room you see when launching the tool. This room should be

	Frequency
Skin color	6
Eye color	3
Clothing color/style	4
Hair color/style	4
Gender	10
Body type	9
Hand gestures	6
Head gestures	7
Facial expressions	10

Table 3: Preliminary user study results: What NPC features/behaviour are a must-have

Figure 1: The settings room with a practice chair, experiment explanation, teleport pad and settings panel

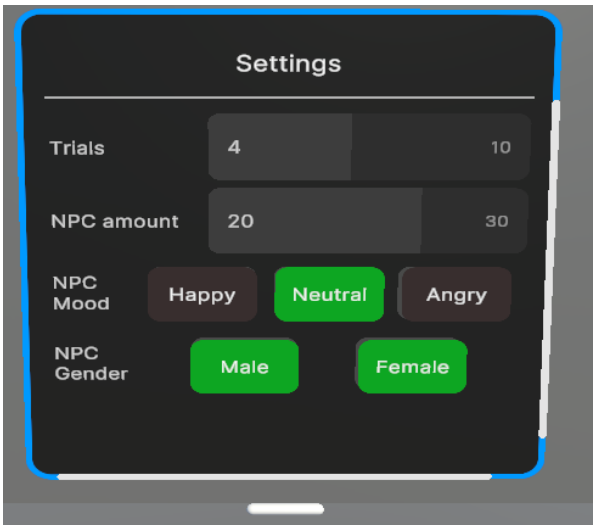

Figure 2: The settings panel

entered by the researcher and used to set up the experiment. The room is a simple square room with a panel on the wall with settings that can be altered, some additional explanations of the settings, and a teleport pad that takes you to the practice room. The following settings had been chosen based on the results of the preliminary user study(3.3) and can be seen in Figure 2: number of trials, number of NPCs, NPC gender and NPC mood. The number of trials references how often the experiment gets repeated, this number can be altered by moving a slider to the desired number. There can't be less than 1 Trial. For the experiment there was a maximum amount of trials of 10, this maximum can be altered to be as large as needed. The number of NPCs references how many NPCs spawn at the start of each Trial, there is also a button which randomizes the amount that are spawned. The number of NPCs is again controlled by a slider, with a minimum value of zero and a maximum value equal to the number of spawn points in the scene. The NPC gender references the physical appearance of the NPCs, at the moment only male and female NPCs are implemented. The genders are selected through toggleable buttons, when none are selected no NPCs are spawned. The NPC mood references the facial appearances of the NPC. For example, this can alter an NPC having a frown or a smile. At the moment there are 3 moods implemented: happy, neutral and angry, these were chosen based on the NPC emotions(2.3.1) section. The difference be-

tween a happy and angry NPC can be seen in Figure 6

4.3 Practice Room

The practice room is a mainly empty room for the participant to practice the controls in VR. The room further has an explanation panel for the experiment, a teleport pad to start the experiment and a chair to get a feel for how close you need to get to it.

4.4 Experiment Room

The experiment room is the actual room. For the experiment, a bus scene was created. The scene has a bus stop(Figure 4) and a bus(Figure 3). The bus stop is a small shelter with a path to the bus. The bus is rectangular, with a small elevation in the back and windows on all sides. A total of 28 chairs are placed in pairs on both sides of the bus, starting from around a fourth until the back end of the bus. All chairs possess the spawn attributes so that NPCs can spawn on them(Figure 5). There are also 3 sound effects in this scene: background noise for the bus, a voice recording which reminds the player to return to the bus stop when the trial is finished and a ding when the next trial starts. Additionally, a bar scene was created as well, however, this scene was not part of the experiment to lessen the duration of the experiment. The bar scene was a square room divided by a wall. In the room, a couple of tables were placed and a bar in one of the corners. NPCs had spawn points spread throughout the room, at the tables and behind the bar. The user would spawn at the entrance of the bar. The bar had a couple of sound sources, playing either music or chatter.

4.5 Important Objects

4.5.1 Chairs

At the moment chairs do 2 things, they can spawn an NPC or check if the player chooses it. A chair has a method called "Spawn", this method takes an NPC object and instantiates it based on the location and scale of the chair. There is also the method "Despawn" that removes any objects instantiated by the chair. If the chair has not instantiated an NPC and the player enters its bounding box it starts a timer for 2 seconds, if those 2 seconds pass without the player moving out of the bounding box the chair recognizes this as being chosen by the player. When a chair is chosen it calls the end Trial method and saves its number to the gathered data.

4.5.2 NPCs

The NPC have 2 types of behaviour. The first one is continuous behaviour, here the NPC spawns and chooses an animation sequence that it performs during the trial. Currently, there are 3 such animation sequences in the tool and they can be applied to any NPC with a humanoid avatar rig. The second type of behaviour is reactionary behaviour, here the NPC still chooses an animation loop but can do something else if certain requirements are met. At the moment this is when the player gets close enough, they turn to the player and say a phrase like "hello" or "hi". There is also a system in place where the NPC reacts with pre-generated phrases based on what the player asks them. During the experiment, only continuous behaviour was used, because in the bus scene, the NPC has no reason to interact with the player.

4.5.3 Teleport Pads

Teleport pads are green circles on the ground that when stood on teleport the player to a predetermined location. Their function is to provide the player with a manual way to progress between locations, without giving them the option to get to spaces where they're not yet or not anymore allowed to.

4.5.4 UXF

UXF manages all the actions that must be performed between trials and keeps track of the data that must be collected. At the end of a trial, all the NPCs are removed, the data gets saved and the sound byte that tells the player to go back to the beginning runs. At the start of a trial the NPCs get spawned in again, the sound effects of the bus start playing and a ding sound to identify that the next trial has begun. The following continuous data is being tracked: the player's position in the environment and the rotation of the VR device. Based on the results from the preliminary study(3.3), the following static data is being tracked: the seat chosen at the end of the trial, the occupancy of the other seats, the amount of NPC that the user got close to, the largest group the player got close to, time spent and the trial number.

5 User Study

5.1 Study Design

In this study, the usability and completion status of the product is evaluated by having participants

Figure 3: A picture of the bus used in the experiment

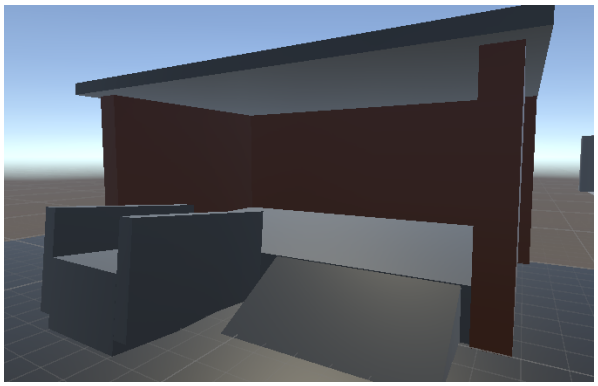

Figure 4: The bus station with ramps for when the player falls off the stage

perform specific tasks using the tool and capturing their feedback. This approach allowed me to assess whether any functionality was still lacking and to gauge whether the current design met the expectations of potential end-users. The insights gained from this evaluation were used to test RQ1 and RQ2.

5.2 Participants

All 5 participants of this research are people who are working or have experience in researching phobias and related psychological conditions. The participants come from 2 universities and have differing levels of experience in performing research, working with VR, and being familiar with IT in general. Recruitment was conducted using convenience sampling, selecting participants from a pool of colleagues connected to the product's end user, as well as personal contacts.

5.3 Materials

The First part of the study (Experiment setup) is performed on the Meta Quest 3 VR headset, made by Meta, using the corresponding Meta Quest Touch Plus-controllers to move around. The software used is created in Unity version 2022.3.21f1. During this part the participants perform 2 sets of tasks, the first set is:

- Set the number of trials to 3
- Set the number of NPCs to 4
- Choose only male NPCs
- Select "Happy," "Neutral," and "Angry" emotions for the NPCs

For the second set the following tasks needed to be performed:

- Choose a low number of Trials
- Select a large number of NPCs
- Have female NPCs in the Trials
- Have neutral NPCs in the Trials

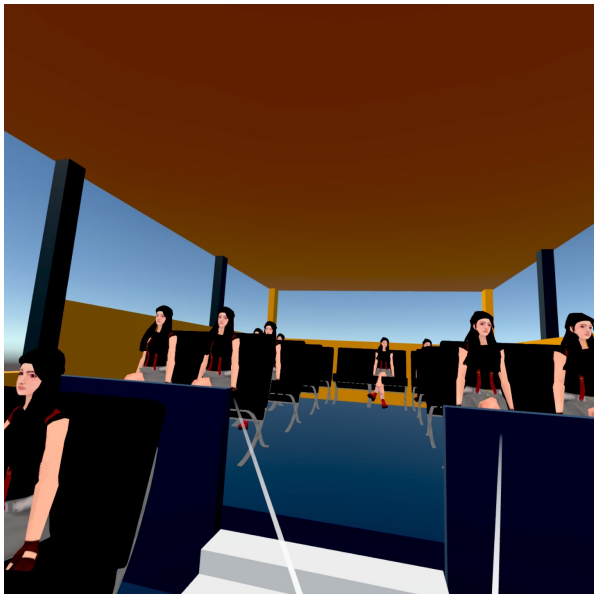

Figure 5: Inside of the bus, with a single type of female NPC

The second part of the study (Data part), is performed on a laptop(ASUS TUF GAMING F15) on which the participant has access to Excel and the directory of the VR headset. The participant also has access to a piece of paper with a guide for opening the data in Excel and how to navigate the VR headset directory, the guide can be seen in Appendix B. During this part the following tasks need to be performed:

- Retrieve the start and end times of Trial 2 for a specified user
- Identify the chair chosen by the user in all three trials
- Determine the number of NPCs present when Chair 2 was selected in a given trial for a specific user
- Extract the largest Z position from the movement file associated with Trial 1

For the third part (Semi-structured Interview) the interview was recorded using the mic-record function of a POCO X3 Pro. The questions that are asked can be found in Appendix C.

5.4 Procedure

5.4.1 VR Trial

In the VR trial, participants were tasked with setting up and configuring specific settings within a virtual reality (VR) environment, followed by executing a predefined task within the configured scene. This research component was designed to assess participants' proficiency in interacting with and manipulating the VR environment. Participants were required to configure two sets of settings for the VR scene. Upon configuring these settings, participants performed the selected scene, which involved the "Bus Scene." In this scene, participants were instructed to enter the bus, choose a seat, and then exit the bus, repeating this process according to the number of trials specified. The primary data was obtained through a post-task interview, where participants shared their experiences and reflections on the task. Secondary data was obtained by the researcher taking notes of the performance and remarks while the participant was performing the trial.

The second set of tasks was intentionally more ambiguous to evaluate participants' ability to interpret and implement less-defined instructions. After setting these parameters, participants once again performed the "Bus Scene" under the newly configured settings. As in the first task, notes were taken by the researcher, and detailed feedback was gathered through interviews to gain insights into their decision-making processes and the usability of the VR interface. The evaluation metrics included the accuracy and alignment of the scene setup with the specified instructions, and the participants' subjective experiences, as captured during the interviews. These metrics were crucial in assessing the effectiveness of the VR interface and the participant's ability to navigate and manipulate the virtual environment under varying degrees of task specificity.

Figure 6: 2 NPCs with different emotions. The left NPC is happy and the right NPC is angry

5.4.2 Data Analysis Task

Following the VR trial, participants were asked to complete a series of data analysis tasks based on the performance data collected during their interaction with the VR scene. These tasks were designed to evaluate the participant's ability to effectively work with and navigate through the data generated by the VR tool. These tasks required participants to engage with the data at both a detailed and holistic level, testing their ability to accurately extract and analyze relevant information. The primary objectives of this phase were to assess the participants' accuracy in retrieving and interpreting the data, as well as their efficiency, measured by the time taken to complete each task. Performance was evaluated based on two key indicators: Correctness, The degree to which the participants' answers matched the correct data points, and Efficiency, The time required to complete each task. These metrics provided insights into the participants' proficiency with data handling in the context of VR-based research, highlighting both their analytical capabilities and their ability to navigate and interpret complex datasets.

5.4.3 Semi-Structured Interview

Upon completion of the VR trial and subsequent data analysis tasks, participants take part in a structured qualitative interview. The purpose of this interview is to delve into their experiences and perceptions related to both the VR environment and the data analysis tasks. The interview explored several key areas:

- Ease of Use: Participants discuss how intuitive and user-friendly they found the VR interface and data analysis tools
- Perceived Utility: The interview gathered insights into how useful participants found the VR environment and data analysis tasks for their intended purposes
- Challenges Faced: Participants have been asked to identify any difficulties they encountered during the VR trial and data analysis, highlighting potential areas for improvement
- Suggestions for Improvement: Finally, participants have been invited to offer feedback on how the VR environment and data analysis tools could be enhanced to better meet user needs

This mixed-methods approach integrates quantitative performance metrics from the VR trial and data analysis tasks with qualitative insights gained from the interviews. By combining these data sources, the study aims to provide a comprehensive understanding of participants' interactions with the VR environment and their data analysis capabilities. This dual perspective helped identify the tool's strengths and pinpoint specific areas where further development is necessary to optimize its performance and usability.

5.4.4 Data Collection

Data collection during the VR trials was facilitated by the UXF tool, which logs key metrics such as the time spent per trial and the specific

settings chosen by the participant for the experiment. While participants were engaged in the VR environment, the researcher also took observational notes when participants verbally expressed their experiences. These notes served to provide additional context and clarification for the participants' responses during the subsequent interview.

To minimize variability across interviews and avoid potential bias from inconsistent or leading questions, the interview is structured. This approach ensures that all participants were asked the same questions in a consistent order. Some questions include optional follow-up prompts, which were used when a participant provided a brief answer or to confirm that the participant had nothing further to add, whether positive or negative.

This comprehensive interview framework is designed to capture both qualitative insights and participant feedback, which were crucial for assessing the tool's current effectiveness and identifying areas for further development.

5.5 Data Analysis

Thematic analysis was employed as the primary method for analyzing the qualitative data obtained in this study. This approach was selected for its flexibility and its ability to uncover, analyze, and report patterns (themes) within the data. By applying thematic analysis, the study was able to produce a rich, detailed, and nuanced account of participants' experiences and perceptions, allowing for a deeper understanding of the key themes that emerged from the interviews.

For quantitative data, task performance and task duration were collected. All tasks were successfully completed by participants. Additionally, task duration data was excluded from analysis because participants employed varying approaches to completing the tasks, resulting in time measurements that were not reasonably comparable across participants. Consequently, the focus remained on the qualitative insights gathered through thematic analysis to inform the study's conclusions.

5.5.1 Audio Transcription

The audio recordings of the interviews were transcribed into text, with clear delineation between the interviewer and participant. This transcription process was carried out using Amberscript, a tool approved by Utrecht University for handling sensitive data. After the initial transcription, the resulting text files were manually reviewed for accuracy by comparing them with the original audio recordings.

During this manual review, the text files were also refined to enhance clarity. Any gestures or references made by participants that are not immediately clear in the text were annotated to provide the necessary context. Additionally, any conversational content that does not contribute value to the structured interview, such as irrelevant digressions, was removed to maintain focus and clarity in the analysis.

5.5.2 Thematic Analysis

Following transcription, the text files were systematically coded to identify and label chunks of data relevant to the research questions. This coding process employed both inductive and deductive approaches. The deductive approach was informed by the research objectives, leading to the identification of codes related to predefined topics. Concurrently, an inductive approach was utilized to allow for the emergence of codes grounded directly in the data, capturing unexpected themes and insights.

After generating a comprehensive set of initial codes, these codes were then collated into potential themes. Themes represent broader patterns of meaning that are essential to addressing the research questions. This phase involved grouping related codes into overarching themes that encapsulated significant aspects of the data, ensuring that each theme captured a coherent and meaningful pattern.

The identified themes were then reviewed and refined in a two-step process. First, the coherence of the data within each theme was evaluated, ensuring that the codes grouped together fit well. Second, the validity of the themes was assessed against the entire dataset, ensuring that the themes accurately reflected the overall data. During this review, themes were revised, merged, or discarded as necessary to develop a thematic structure that was both comprehensive and representative of the participants' experiences and perspectives.

Finally, each theme was clearly defined and named, providing a distinct and meaningful label for the patterns identified in the data. Where appropriate, sub-themes were also identified to provide further granularity and depth to the analysis, enhancing the overall understanding of the participant's interactions with the VR environment and data analysis tasks.

6 Results

The thematic analysis revealed distinct thematic areas corresponding to different phases of the experiment: the Settings Room, the Experiment Room, and the Data Analysis phase. Each of these areas was further broken down into sub-themes that reflect specific aspects of participants' experiences and perceptions. The settings room subthemes include:

- Ease of Use: How intuitive and user-friendly the settings interface was for participants
- Completeness and Clarity: If the participants felt that it was clear what all the settings did and the options they would want to be present

The experiment room sub-themes include:

- Immersion: The extent to which participants felt immersed in the VR environment
- NPC state: Participant's opinion about the implementation of the NPCs
- Task Performance: Participants' perceptions of their ability to perform the tasks within the VR environment.

The Data Analysis sub-themes include:

- Completeness: Participant's opinion if there are data types missing
- Navigation: Participant's experience finding the data with the help of the handbook snippets
- Workability: Participant's ability to perform the given tasks

6.1 Settings Room

The 2 sets of 4 tasks of the settings room, that were mentioned in 5.3, were completed by all 5 participants. The interview part of the settings room was divided according to the sub-themes. The first sub-theme is Control over settings. This is about the reaction participants gave about how they experienced manipulating the settings and things like how it felt to move the slider to alter the number of NPCs in the scene. Completeness/Clarity of the settings is the second subtheme. This sub-theme is about how well people understood how the settings would come back into the scene and how they felt about the options available to them.

6.1.1 Ease Of Use

In general, people found the settings quite intuitive to use. 4 out of 5 participants found the sliders intuitive and 5 out of 5 were positive about the buttons. Some positive reactions to the sliders were "The layout made it instantly clear it was supposed to be a slider" and "I just needed a second to look at the settings to realize that it was a slider I could manipulate". Most other positive reactions were just mentioning that it felt intuitive. The one critique on the slider was "In research, I am used to it being just a point you click for the amount instead of a slider". The main reason given for why the buttons were perceived as intuitive: "It was clear that when the button turned green it was selected". The feature that created some issues had to do with the last deselected button turning white which identified that the state of that button was changed. "I was quite unsure if they correctly turned off when they turned white", "I had no idea what the white glow meant on the button when deselected" and "I got confused by the white colour so I didn't know if I needed to do something else" were some of the frustration this feature lead to. An additional point of critique had to do with the settings screen as a whole and went as follows: "When not standing clearly in front of the screen it was hard to control the slider and press the buttons", such an issue was mentioned by 2 participants.

6.1.2 Completeness And Clarity

4 out of 5 participants gave general responses that it was clear what the settings meant, were 1 participant found the naming too vague. "I did not understand what it meant to add happy and neutral NPCs, like is it expression or behaviour?", "I did not know how many women would be added" and "I would have liked to know how the bus would be filled when more NPCs were added" These are some quotes from the participants about a lack of information about what the setting provided/did. This got further expanded upon by participants having liked features such as "A button to spread the NPCs evenly over the seats", "A slider to control the amount of female/male NPCs", "Would have liked to be able to control things like age and ethnicity" and "Some way to choose which male/female NPCs are added or what kind of behaviour they may show". 5 out of 5 participants did mention that they thought the available setting did encompass the most important features, however, 3 out of 5 thought that additional settings would be quite advantageous.

6.2 Experiment Room

The first sub-theme is perceived immersion, this is about what the participants were missing or added

to what made the bus feel like a real-world bus. NPC state is the second sub-theme and encapsulates what was good and/or bad about the NPCs. Ease of performance is the third sub-theme and is for everything that is about what went well and what went bad when performing the multiple trials.

6.2.1 Immersion

Immersion in the environment was increased by the combination of their being different NPCs and that they had different behaviours, as mentioned by 4 participants signified by quotes like "I like that I decide who I want to sit next to based on their gender and how calm they are, which feels similar as I would do in real life". The bus is stylized similarly to a real bus, mentioned by 4 participants. How the NPCs spread out over the chairs, mentioned by 3 participants, "It is just as a bus, some seats are empty some people sit on the outside chair and others sit together". "I miss the sound of the bus and people", "Why was there no bus driver?" and "The setting was quite simple" were some of the aspects which took away from the perceived immersion being mentioned by 3, 2 and 2 participants accordingly.

6.2.2 NPC State

The most occurring critique of the NPCs was that "One of the NPC had no face which made it quite creepy", the existence of this NPC was mentioned by all 5 participants. Other problems with the NPCs were "It is hard to distinguish between the emotion between NPCs", "The NPCs were not correctly sitting on the chairs" and "Some of the NPCs were too expressive with their movement". More positive reactions about the state of the NPCs were quotes as "Most NPCs do feel different" and "The NPCs do give a somewhat human feeling, except for the one without a face".

6.2.3 Task Performance

During the trials only 1 participant needed to use the ramps to get back to the starting point after falling, all other participants performed the trials without any problems. 3 participants mentioned that moving around during the trial was easier as they could practice during the settings stage. 4 participants explicitly stated how it felt walking around, "Even though I am sitting down it does not feel weird to walk through the stage" and "It felt so natural that it really felt like I was walking around". There has been one mention of motion sickness, explained as "I think the amount of 180-degree turns are starting to make me motion sick ... when you need to go back to the beginning and then back into the bus".

6.3 Data-part

All 5 tasks were completed by all 5 participants. Participants were timed on their completion of the assigned tasks, with the recorded times (in seconds) as follows: 223, 238, 242, 249, and 262. The average time taken by participants to complete the tasks was calculated to be 242.8 seconds ($M = 242.8$). The standard deviation of the task completion times was 15.0 seconds ($SD = 15.0$), indicating a moderate level of variability in the time taken by participants. These results suggest that while there was some variation in task completion time, most participants completed the tasks within a similar time frame. Most of the time was spent on navigating and setting up the data files. The first sub-theme of the interview is the completeness of data, encompassing mentions about the data types given and whether the participants were missing data types. Data navigation is the second sub-theme and is about how well the participants found the right data files and how well the snippets of the guidebook helped them with it. The third sub-theme is Data workability, this is about how well the participants felt they could manipulate and use the data.

6.3.1 Completeness

2 participants felt like they lacked data about where every NPC was sitting, specifics like gender, emotion and NPC type were mentioned. "even though you choose the settings you want yourself, I feel it would still be helpful to have the chosen settings come back into the data", this want for the settings to be part of the data was the other data type that was missed by the participants. 3 participants also mentioned that it is hard to realize at the moment what they might be missing. All 5 participants agreed that most of the stuff they noticed came back into the data.

6.3.2 Navigation

All 5 participants opened the data files through File Explorer instead of Excel, and 3 participants noted that they read the part which mentioned opening it through Excel but autopiloted and still opened it through File Explorer. It was mentioned by all participants that the text snippets from the handbook were easy to use. Most participants mentioned something along the lines of "the first time finding the file took a little bit, but it was quite easy to do it again". 1 participant noted

that "it got a little confusing in which file what data is".

6.3.3 Workability

4 out of 5 participants found the data easy to work with and 1 participant found it somewhat more difficult. The reason that it was more difficult was that "I would have liked some more explanation for certain columns, to understand what they meant". The other participants did not mention having difficulty with understanding the naming, except for the XYZ-rot, which only 1 participant understood the meaning of. "If I want the data in another format it would be easy to do that in Excel" and "It was easy to read the data, but the trials were also low so that might have been part of it" were two additional remarks about how the data was experienced.

6.4 Tool Usage

3 out of 5 participants see themself setting up research using the tool, whereas the other 2 want the tool to have a couple of extra features before they think this is feasible. 5 out of 5 participants do feel they can come to a significant conclusion using the data given by the tool, 1 participant does think that there is still a lot of extra information that would be valuable that can be obtained. 2 out of 5 participants would like to have a room which explains the controls of VR for the participants to practice in. In total 3 bugs had been encountered during the testing of which 2 bugs the participant mentioned that it didn't affect their performance and 1 bug which needed a restart of the trial.

7 Discussion

The primary aim of this study was to evaluate the usability and effectiveness of a VR tool designed for setting up an experiment for people with Social Anxiety Disorder by researchers with there being no to nearly no need for IT skills. The findings from the thematic analysis of the Settings Room phase offer valuable insights into participants' experiences and highlight several key themes: Control over Settings and Completeness and Clarity of Settings. This discussion interprets these findings, compares them with existing literature, and considers their implications for future tool development.

7.1 Settings Configuration

Participants generally found the settings interface intuitive, particularly appreciating the straightforward design of the sliders and buttons. The positive feedback regarding the ease of use suggests that the tool's interface successfully lowers the barrier to entry for researchers who may not be technologically proficient. However, the concerns raised about certain UI elements—such as the confusion caused by the white glow of deselected buttons and the difficulty in interacting with the interface when not standing directly in front of it—indicate areas where the user experience could be improved. These issues, although minor, could affect the tool's accessibility, particularly for users who are less familiar with VR environments.

The feedback on the completeness and clarity of the settings also offers valuable insights. While most participants found the available settings sufficient, the desire for additional features, such as more detailed control over NPC characteristics (e.g., gender, age, and ethnicity), indicates that expanding these options could enhance the tool's flexibility and relevance for a broader range of research scenarios. The need for clearer explanations of what the settings control, as highlighted by the participants, suggests that the tool could benefit from more detailed in-tool guidance or tooltips to help users understand the impact of their choices on the experimental environment.

7.2 Experiment Environment

The experiment room, particularly the immersion experienced by participants, received mixed feedback. On the one hand, the realistic behaviour of the NPCs and the overall setup were praised for contributing to a sense of immersion, which is crucial in research on SADs where the authenticity of social interactions is paramount. On the other hand, the lack of environmental details, such as the absence of sound or a bus driver, and the simplicity of the setting were noted as factors that detracted from the immersive experience. These findings suggest that while the tool provides a good foundation for creating realistic social scenarios, further refinement is needed to enhance the realism and thus the ecological validity of the research conducted using this tool.

The issues with NPCs, particularly the unsettling appearance of an NPC without a face, underscore the importance of ensuring that all elements within the VR environment are polished and free of such distractors. While some participants appreciated the diversity in NPC behaviour, the difficulty in distinguishing between NPC emotions and the improper seating of some NPCs suggests that these elements need to be more carefully calibrated. Improving these as-

pects could significantly enhance the realism and emotional impact of the VR scenarios, making the tool more effective for studying SADs.

Taking the participants' flawless performance into account during the experiment setup and the generally positive reaction on how natural it felt to use the settings it feels safe to say that RQ1, "The VR Tool will be able to be used by the researcher to set up an experiment without any additional help", can be answered positively for the part about not needing additional help. Most participants did seem to be positive about using the tool to set up research, however, there were also several mentions of extra features that would have been liked. As such, it can be stated that the tool can be used to set up research, but there is certainly room for further improvement.

7.3 Data Analysis

The data analysis phase revealed that participants were generally able to complete the tasks within a reasonable time frame, indicating that the tool is accessible to users with varying levels of experience. The feedback on the completeness of the data, particularly the desire for more detailed information on NPC characteristics and the experimental settings, suggests that enhancing the data outputs could make the tool more powerful and versatile for researchers. with this RQ2, "The data acquired by the tool can be navigated and worked with, with only the help of the textbook snippet", can be positively answered.

Participants' experiences with data navigation and workability highlight the tool's current strengths and weaknesses. While the majority found the data easy to work with, the confusion caused by the format and labelling of certain data columns indicates a need for clearer documentation or more intuitive data structures. Ensuring that the data outputs are easily interpretable is crucial for enabling researchers to draw meaningful conclusions from their experiments.

8 Limitations

The sample size of this study was relatively small, consisting of 5 participants from Universities in the Netherlands. As a result, the findings may not be fully generalizable to larger, more diverse populations. Future studies with larger and more varied samples could enhance the generalizability of the results.

The study relied on an interview, which can introduce bias due to social desirability, recall inaccuracies, or misunderstanding of the questions. Although steps were taken to minimize these issues, such as creating a semi-structured interview to lower the odds of inconsistencies between the interviews, the potential for bias cannot be completely ruled out. Future research could utilize more objective or validated measurement tools to corroborate these findings and provide additional data to which the current findings can be compared.

For the study, it was not possible to obtain a room for every participant in which they could walk around when using the VR headset. Therefore, it has been chosen for all participants to perform the experiment in a stationary position using the controllers to walk around. This is done to limit the amount of variability between participants' experiences in the VR environment. For further research, it is important to get data on how walking around the VR environment is experienced.

Due to not having a visual designer when developing the tool the assets used are either simple or free assets from the Unity store leading to differing art styles. The inconsistencies between assets, primarily NPCs and the low quality of other assets might influence the feeling of immersion during the study. The influence of immersion might also affect the analysis of other mechanics or aspects of the environment. Making sure the assets use a similar art style to each other and are of decent quality is important for future research to make sure it does not affect the experience.

9 Future work

The participants' willingness to consider using the tool for their research, despite some reservations, is a positive indication of its potential utility. However, the feedback also suggests that further development is needed to fully meet the needs of researchers, particularly in terms of adding more features and improving the clarity of the interface. The suggestion to include a VR control tutorial room is particularly noteworthy, as it highlights the need to make the tool more accessible to participants who may be unfamiliar with VR technology.

The identification of bugs during testing, while not significantly hindering the experiment, points to the importance of ongoing refinement and testing to ensure the tool's reliability. Addressing these issues and incorporating the suggested improvements could significantly enhance the tool's

usability and effectiveness, making it a more valuable resource for researchers studying SADs.

One compelling direction for future work is the integration of conversational AI systems to enable dynamic, real-time interactions with NPCs. Unlike traditional scripted dialogues, AI-driven NPCs could engage players in more natural and contextually relevant conversations, adapting to the player's actions and decisions in real time. This could significantly enhance the immersion and realism of the experience, allowing NPCs to respond more intelligently to a broader range of player inputs. This however needs to be done using an external server due to the current load it would put on the hardware of the VR headset.

Another approach worth exploring is the development of pre-created conversational encounters. These would involve carefully designed dialogue trees and interaction scenarios that, while not as flexible as AI-driven conversations, could still provide deep and meaningful interactions based on a wide range of player choices. This method could ensure high-quality, narrative-driven interactions leading to the NPCs seem more like an actual person.

another promising direction for future work involves the development of a scene-creator tool. This tool would enable users to design and customize environments within the game or simulation through a user-friendly interface. The proposed Scene Creator would feature a drag-and-drop interface, allowing users to select from a library of pre-designed components such as terrain pieces, buildings, vegetation, and other environmental assets. By simply dragging these components onto a workspace, users could quickly and easily construct detailed scenes without requiring extensive knowledge of 3D modelling or coding.

10 Conclusion

Overall, this study demonstrates that the developed VR tool has significant potential to aid researchers in conducting studies on SADs, in particular by those with no programming experience. Both research questions have been accomplished, showing that the tool already fulfils its goals. However, while the tool's current design is largely intuitive and functional, the feedback gathered suggests several areas for improvement. By addressing the identified issues related to usability, immersion, and data analysis, and by incorporating the additional features requested by participants, the tool could become a robust and widely adopted resource in the field of SAD research. Continued development and user testing are essential in refining the tool to better meet the needs of its intended users.

References

Arnfred, B., Svendsen, J. K., Adjourlu, A., and Horthøj, C. (2023). Scoping review of the hardware and software features of virtual reality exposure therapy for social anxiety disorder, agoraphobia, and specific phobia. *Frontiers in Virtual Reality*, 4.

Bergsnev, K. and Sánchez Laws, A. L. (2022). Personalizing virtual reality for the research and treatment of fear-related disorders: A mini review. *Frontiers in Virtual Reality*, 3.

Biggs, B. K., Vernberg, E. M., and Wu, Y. P. (2011). Social anxiety and adolescents' friendships. *The Journal of Early Adolescence*, 32(6):802–823.

Brookes, J., Warburton, M., Alghadier, M., Mon-Williams, M., and Mushtaq, F. (2019). Studying human behavior with virtual reality: The unity experiment framework. *Behavior Research Methods*, 52(2):455–463.

Center, N. S. A. (2023). liebowitz-sa-scale. *nationalsocialanxietycenter.com*, 32:19/02/2024.

Coelho, C. M., Waters, A. M., Hine, T. J., and Wallis, G. (2009). The use of virtual reality in acrophobia research and treatment. *Journal of Anxiety Disorders*, 23(5):563–574.

Emmelkamp, P. M., Meyerbröker, K., and Morina, N. (2020). Virtual reality therapy in social anxiety disorder. *Current Psychiatry Reports*, 22(7).

Fernandes, C., Silva, S., Pires, J., Reis, A., Ros, A. J., Janeiro, L., Faísca, L., and Martins, A. T. (2017). Eye-tracking evidence of a maintenance bias in social anxiety. *Behavioural and Cognitive Psychotherapy*, 46(1):66–83.

Frommel, J., Dechant, M. J., and Mandryk, R. L. (2021). The potential of video game streaming as exposure therapy for social anxiety. *Proceedings of the ACM on Human-Computer Interaction*, 5(CHI PLAY):1–25.

Garcia-Palacios, A., Hoffman, H. G., Kwong See, S., Tsai, A., and Botella, C. (2001). Redefining therapeutic success with virtual reality exposure therapy. *CyberPsychology amp; Behavior*, 4(3):341–348.

Hart, T. A., Turk, C. L., Heimberg, R. G., and Liebowitz, M. R. (1999). Relation of marital status to social phobia severity. *Depression and Anxiety*, 10(1):28–32.

Herbert, J. D., Brandsma, L. L., and Fischer, L. (2014). Assessment of social anxiety and its clinical expressions. *Social Anxiety*, page 45–94.

Howell, A. N., Zibulsky, D. A., Srivastav, A., and Weeks, J. W. (2015). Relations among social anxiety, eye contact avoidance, state anxiety, and perception of interaction performance during a live conversation. *Cognitive Behaviour Therapy*, 45(2):111–122.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., and Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of dsm-iv disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, 62(6):593.

Liou, W.-K., Lin, W.-H., Lee, Y.-T., Chen, S., and Liang, C. (2023). The distinction between first-person perspective and third-person perspective in virtual bodily self-consciousness. *Virtual Reality*, 28(1).

McTeague, L. M., Shumen, J. R., Wieser, M. J., Lang, P. J., and Keil, A. (2011). Social vision: Sustained perceptual enhancement of affective facial cues in social anxiety. *NeuroImage*, 54(2):1615–1624.

Nason, E. E., Trahan, M., Smith, S., Metsis, V., and Selber, K. (2019). Virtual treatment for veteran social anxiety disorder: A comparison of 360° video and 3d virtual reality. *Journal of Technology in Human Services*, 38(3):288–308.

North, M. M., North, S. M., and Coble, J. R. (1998). Virtual reality therapy: An effective treatment for the fear of public speaking. *International Journal of Virtual Reality*, 3(3):1–6.

Parrish, D. E., Oxhandler, H. K., Duron, J. F., Swank, P., and Bordnick, P. (2016). Feasibility of virtual reality environments for adolescent social anxiety disorder. *Research on Social Work Practice*, 26(7):825–835.

Pertaub, D.-P., Slater, M., and Barker, C. (2001). An experiment on fear of public speaking in virtual reality. *PsycEXTRA Dataset*.

Pertaub, D.-P., Slater, M., and Barker, C. (2002). An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators and Virtual Environments*, 11(1):68–78.

Porter, E. and Chambless, D. L. (2017). Social anxiety and social support in romantic relationships. *Behavior Therapy*, 48(3):335–348.

Price, M. and Anderson, P. (2007). The role of presence in virtual reality exposure therapy. *Journal of Anxiety Disorders*, 21(5):742–751.

Ranta, K., Junttila, N., Laakkonen, E., Uhmavaara, A., La Greca, A. M., and Niemi, P. M. (2012). Social anxiety scale for adolescents (sas-a): Measuring social anxiety among finnish adolescents. *Child Psychiatry amp; Human Development*, 43(4):574–591.

Ratan, R. and Sah, Y. J. (2015). Leveling up on stereotype threat: The role of avatar customization and avatar embodiment. *Computers in Human Behavior*, 50:367–374.

Rheu, M. M., Ratan, R., Sah, Y. J., Cherchiglia, L., and Day, T. (2022). Jogging in your avatar's footsteps: The effects of avatar customization and control intuitiveness. *Frontiers in Virtual Reality*, 3.

Rodebaugh, T. L., Holaway, R. M., and Heimberg, R. G. (2004). The treatment of social anxiety disorder. *Clinical Psychology Review*, 24(7):883–908.

Rubo, M. and Munsch, S. (2024). Social stress in an interaction with artificial agents in virtual reality: Effects of ostracism and underlying psychopathology. *Computers in Human Behavior*, 153:107915.

Safran, E. B., Nilsen, E., Drake, P., and Sebok, B. (2024). Effects of video game play, avatar choice, and avatar power on motivation to prepare for earthquakes. *International Journal of Disaster Risk Reduction*, 101:104184.

Selzer, M. and Castro, S. M. (2023). A methodology for generating virtual reality immersion metrics based on system variables. *Journal of Computer Science and Technology*, 23(2).

Servotte, J.-C., Goosse, M., Campbell, S. H., Dardenne, N., Pilote, B., Simoneau, I. L., Guillaume, M., Bragard, I., and Ghuysen, A. (2020). Virtual reality experience: immersion, sense of presence, and cybersickness. *Clinical Simulation in Nursing*, 38:35–43.

Slater, M., Pertaub, D.-P., and Steed, A. (1999). Public speaking in virtual reality: Facing an audience of avatars. *IEEE Computer Graphics and Applications*, 19(2):6–9.

Tcha-Tokey, K., Christmann, O., Loup-Escande, E., and Richir, S. (2016). Proposition and validation of a questionnaire to measure the user experience in immersive virtual environments. *International Journal of Virtual Reality*, 16(1):33–48.

ten Have, M., Tuithof, M., van Dorsselaer, S., Schouten, F., Luik, A. I., and de Graaf, R. (2023). Prevalence and trends of common mental disorders from 2007-2009 to 2019-2022: Results from the netherlands mental health survey and incidence studies (nemesis), including comparison of prevalence rates before vs. during the covid-19 pandemic. *World Psychiatry*, 22(2):275–285.

U.S. Department of Health and Human Services (2022). social anxiety disorder more than just shyness. *National Institute of Mental Health*, page 08/02/2024.

Watson, D. and Friend, R. (1969). Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology*, 33(4):448–457.

Weeks, J. W., Howell, A. N., and Goldin, P. R. (2013). Gaze avoidance in social anxiety disorder. *Depression and Anxiety*, 30(8):749–756.

Weeks, J. W., Howell, A. N., Srivastav, A., and Goldin, P. R. (2019). "fear guides the eyes of the beholder": Assessing gaze avoidance in social anxiety disorder via covert eye tracking of dynamic social stimuli. *Journal of Anxiety Disorders*, 65:56–63.

WITTCHEN, H.-U., STEIN, M. B., and KESSLER, R. C. (1999). Social fears and social phobia in a community sample of adolescents and young adults: Prevalence, risk factors and co-morbidity. *Psychological Medicine*, 29(2):309–323.

Wong Sarver, N., Beidel, D. C., and Spitalnick, J. S. (2013). The feasibility and acceptability of virtual environments in the treatment of childhood social anxiety disorder. *Journal of Clinical Child amp; Adolescent Psychology*, 43(1):63–73.

Zaider, T. I. and Heimberg, R. G. (2003). Non-pharmacologic treatments for social anxiety disorder. *Acta Psychiatrica Scandinavica*, 108(s417):72–84.

A VR requirements for conducting experiments in social science scenarios

VR requirements for conducting experiments in social science scenarios

Start of Block: Introduction

Q25 The goal of this project is to develop a VR environment for conducting experiments in social science scenarios, e.g., for studying social anxiety. For this, we are interested in the importance of different features for such experiments. When answering the following questions, please imagine you are conducting an experiment with such a VR platform. Please rate how interested you would be in gathering the different measures. The following page is the Participant information sheet and the confirmed consent form, please read through these and click the "i consent" button on the end if you have no issues with what you consent to.

Importance In the following questions, we will explore various elements from which data can be collected. We ask of you to assess the relevance and significance of this data for research purposes.

	not important (1)	little important (2)	important (3)	Very important (4)	Indisputable (5)
Distance to NPC's when walking/talking (1)	○	○	○	○	○
Total distance travelled (2)	○	○	○	○	○
Time spend Talking/Walking (3)	○	○	○	○	○
Eye tracking (4)	○	○	○	○	○
NPC Characteristics (5)	○	○	○	○	○
Transcript of conversation (6)	○	○	○	○	○

Explanation If you wish to elaborate on specific decisions or include additional elements that you find useful for data collection, please feel free to provide them in this section

_____

End of Block: Measurement Elements

Start of Block: VR Tool functionality Part 1: NPC

Multiple choice To create variety in the NPCs, what kind of looks are important (impact the research) to vary for the NPCs? (Choose all that apply)

- ☐ skin color (1)
- ☐ eye color (2)
- ☐ clothing color (3)
- ☐ hair color (4)
- ☐ clothing style (5)
- ☐ hair style (6)
- ☐ body proportions (7)
- ☐ gender (8)
- ☐ Accessories (Piercings, Tattoo and so forth) (9)
- ☐ Other... (10) _____

Most applicable What kind of conversation style should the AI/NPC possess?

- ○ Pre-determined (All answers and follow-up questions are created by hand) (1)
- ○ Semi-reactive (There is a premade set of questions/answer and the NPC chooses one based on keywords/phrases said by the participants) (2)
- ○ Conversational AI (Answers are given based on a trained AI model, think of things like ChatGPT)) (3)
- ○ Other... (4) _____

Multiple choice What kind of physical actions should the NPC be able to perform in conversation? (Choose all that apply)

- ☐ Wave/shake/Box (1)
- ☐ Facial expressions (2)
- ☐ Gestures when speaking (3)
- ☐ Nodding, Looking around, shaking head (4)
- ☐ Other... (5) _____

Most applicable How should the NPC voice be implemented

- ○ Text-to-Speech (1)
- ○ pre-recorded Voice clips (2)
- ○ Other... (3) _____

End of Block: VR Tool functionality Part 1: NPC

Start of Block: VR Tool functionality Part 2: User

Multiple choice In what perspective should the participant be in the VR environment?

- ○ 1st Person Perspective (1)
- ○ 3rd Person Perspective (2)

Explain Should the participants avatar/character be customizable? If yes, please explain what should be customizable

_____

Most applicable How should the participant move around the space?

- ○ Teleportation (Look and point somewhere with the controller and you teleport in that direction) (1)
- ○ Walking using the controller on an joystick (2)
- ○ Walking towards where the participant is looking/pointing at (3)
- ○ Real world walking (your character moves like the participant in the real world) (4)
- ○ Other... (5) _____

Multiple choice What kind of physical actions should the participant be able to perform? (Choose all that apply)

- [ ] Stand/Sit/Lay (1)
- [ ] Wave/shake/box (2)
- [ ] run/jump (3)
- [ ] throw/shove (4)
- [ ] Other... (5) _____

Multiple choice How should the participant decide when he is talking to someone?

- ○ The participant looks at the person he wants to talk to (1)
- ○ The closest NPC when the target start talking (2)
- ○ By making physical contact with an NPC (3)
- ○ By using the controller to click on the NPC (4)
- ○ Other... (5) _____

End of Block: VR Tool functionality Part 2: User

Start of Block: VR Tool functionality Part 3: UI/Game world

Choose amount How prevalent should the background noise be?

	0 10 20 30 40 50 60 70 80 90 100
during walking ()	
during conversation ()	

Importance Is it more important that the environment runs well (Efficiency) or that it looks good (Quality)

	Full Efficiency (1)	Efficiency focused (2)	equal (3)	Quality focused (4)	Full Quality (5)
Design focus (1)	○	○	○	○	○

Most applicable Functionality (Time, Tasks, etc. are visible) or Immersion (Close to reality)

- ○ Realistic (All information is immersed into the environment) (1)
- ○ Functional (Information is given in the top right of the screen) (2)
- ○ Other... (3) _____

Explain Are there particular environments that sound especially useful to design?

_____

End of Block: VR Tool functionality Part 3: UI/Game world

Start of Block: Manipulative variables

B The Guide Of The Data Part

Load File into Excel:

Opening the files directly often leads to the data not being split up by excel, to get the data accurately represented follow the steps below:

1. Open Excel.
2. Click New, then click on Blank workbook.
3. Click on the Data tab.
4. Click Get External Data From Text.
5. Navigate to the CSV-formatted file saved on your system, select it, then click Import
    - Read "How to find the Data" to know where the data is saved

How to find the Data:

To gain access to the data obtained by the VR perform the following steps:

- Plug the headset into the pc using a cable
- On the VR headset a pop-up asking for permission to share content should appear, click on this (if you were not fast enough, click on the bell icon and search for the notification in the notifications list)
- Find the data at This PC\Quest 3\
  Internal shared storage\Android\data\com.unity.vrtemplate\files\Bus_Scene
- In the Bus_Scene folder you find multiple folders with randomized names for each performed session, it should like this:

- Select one of the participants
- select the "S001" folder

- In here you find 2 folders (session_info and trackers) and an excel type file (trial_results)
    - Session_info: The files in this folder are useful for debugging the unity application and serve no further purpose as of now.
        - Log: Possesses the log of when the application is running
        - participant_details: empty
        - settings: empty
    - trackers: possesses multiple excel files, each excel file hosts the movement of the participant for each trial performed. Location is based on an pox_x, pos_y, pos_z and the rotation of the headset as rot_x, rot_y and rot_z
    - trial_results: Holds additional data of the experiment, like the amount of NPC's in the trial and the chair chosen by the participant
- To learn how to load the excel type files into excel, read "Load File into Excel" Topic

Figure 7: Guide for navigating the VR directory and opening files in Excel

C Semi-strcutured interview Questions

- What was your overall experience setting up the experiment?
- How intuitive were the different settings to alter?
  - Why did setting x feel intuitive?
  - What made setting x not feel intuitive?
- Did you want to have more control over the settings you had? if yes, what would that be?
- Are there additional settings you would want to have access to?
- How did you experience performing the experiment?
- When walking through the scene, did you see your settings come back?
  - How did you expect setting x to come back in the scene?
- Did you miss something else in the scene?
- Did the text before entering the scene provide you with enough information to confidently perform the experiment?
  - What information did you miss for this to have been sufficient?
- What was your general experience in going over the data?
- Did you have any difficulty in performing the tasks?
- Did you find the data visualized in an understandable manner?
- Did you miss any types of data you would have liked to obtain?
- How clear were the snippets of the handbook you got access to?
- Are you able to set up an experiment with the tool in its current state? if not, what would still be needed?
- Are you able to come to a meaningful conclusion using the data that is gathered by the tool? if not, what would still be needed?