University of Twente

Section of Psychology of Conflict, Risk, & Safety (BMS)

M.Sc. Thesis

Evaluating a New Virtual Reality Environments` Effectiveness for Suicide Negotiation Training Performance Evaluation

Jonah Sauer

08.01.2024

1st Supervisor: Dr. Steven Watson

2nd Supervisor: Dr. Miriam Oostinga

Abstract

Suicide negotiations fulfil an important role in law enforcement activity, as they might save lives, and must be trained thoroughly. The recent technology of VR as a suicide negotiation training evaluation tool seems promising for improving current training methods. A new developed VR environment (VE) was tested for its effectiveness for suicide negotiation training, specifically for performance evaluation and skill fostering through meaningful experience. The present proof-of-concept, mixed-methods study tried to shed light on the relationships between *Immersion, Realism,* and aspects of *Presence* and variables important to suicide negotiation training and experiential learning (*Performance* & *Motivation*). How immersed and present participants were in the VE, and how realistic they felt the VE to be, was measured. Participants received a training about a common negotiation model before they had to apply the gained knowledge in the new VE. Subsequently, questionnaires were administered, and correlations and regression analyses were run. Additionally, a content analysis was conducted on participants answers to three open questions. Results show significant correlations between VR-variables themselves, and Social Presence and Empathy. Scores on the variables did not predict performance or motivation. Qualitative results seem to indicate sufficiency of levels of elicited VR metrics, as well as emotional involvement. Most importantly, there were indications for the VE achieving baseline-effectiveness. Based on the present results, future research should focus on comparing VR´s effectiveness against that of role-plays and differentiate between aspects of realism. Future VE´s for suicide negotiation training should expand interaction possibilities, focus design on the social aspects of realism and increase difficulty.

Keywords

Suicide intervention; crisis negotiation training; virtual reality; immersion; social presence

Introduction

Recently, Virtual Reality (VR) has become a widely accessible tool for private persons but is also more frequently used in research and for training purposes (e.g., Schmidt Mast et al., 2018). VR is already used for training of practical, social (Schmidt Mast et al., 2018), and professional skills (Checa & Bustillo, 2019). Often, these VR-enabled trainings produce more effective results than traditional trainings (Checa & Bustillo, 2019). Compared to traditional training methods, VR training enables professionals to train situations that otherwise would not be trainable adequately, since it allows for realistic depictions of high-risk situations, enables a cost-efficient training (Saghafian et al., 2020), and is controllable and reproducible (Howard & Gutworth, 2020). One such high-risk situation is suicide negotiation, a type of crisis negotiation. Generally, crisis negotiations are negotiation between a law enforcement crisis negotiator and another person that is displaying a personal crisis, which is often a potential suicide or criminal behaviour (e.g., terrorist attack or hostage negotiation) (Vecchi et al., 2005). In suicide negotiation, law enforcement correspondingly negotiates with a person displaying suicidal intentions. To achieve non-lethal outcomes, negotiators are trained to use psychological models of behavioural influence (Vecchi et al., 2005). Application of these models in real-life high risk suicide negotiations require previous training in a possibly close-to-life simulation to maximize success probability. This strongly suggests that incorporating VR into suicide negotiation training might improve learning outcomes by allowing for realistic simulation and thus possibly lead to better negotiation outcomes in the real world.

High Risk Scenarios & Role-Plays

Suicide negotiation is a ‘high risk scenario’, which are scenarios with the defining factor of having a high risk to end in detrimental or even fatal outcomes if errors are made (Vecchi et al., 2005). As high stakes are difficult to simulate, but also defining characteristic,

realistic training becomes difficult. However, following experiential learning, learners should ideally actively experiment with the acquired skills and knowledge to solidify them (Burch et al., 2019). One of experiential learnings central elements is, as the name says, the experience and experimentation with new knowledge to solidify this knowledge through engagement with it (Kolb & Kolb, 2018)..

In law enforcement, these experiences are currently mostly gathered with role-plays (Vecchi et al., 2005). They serve to test trainees' knowledge and skills, while simultaneously fostering them through experience. To improve training, van Hasselt et al. (2008) suggest implementing higher levels of detail in scenario descriptions, aiming to help participants "get into" their roles" (van Hasselt et al., 2008, p. 260) and towards a need for greater realism, to ultimately closer mirror the potentially dangerous simulation in the real world. Some authors pose, for example, enhancing realism by using trained confederates or actors (van Hasselt et al., 2008). This still bears the risk of human failure (to act properly) and does not guarantee meaningful experience, replicability, and comparability. Further, employing an actor often might be expensive. VR might pose a more cost-efficient and effective approach to enhance realism that could support role-play trainings, which are said to lack transferability to the real world, thus, ecological validity (van Hasselt et al., 2008).

VR might possibly have potential to be a platform for training with higher ecological validity. Learners could potentially feel "more in" the situation, allowing for a more memorable experience with a perceived other human, similar as in social skills training (Schmidt Mast et al., 2018). This, in turn, would be allowing for eliciting empathy and emotional responses (Gillies & Pan, 2019). Such a tool would have to be designed to the trained scenarios needs and at best convey a feeling of high risk.

Suicide Negotiation and BISM

Any new VR training should mirror the benefits of current trainings to be effective. Current trainings are allowing learners to practice applying skills needed to change the behaviour of the person in crisis. Thus, new VR trainings should be able to do so as well. To reach the aim of peaceful resolution, the suicide negotiator normally must bring the person in crisis (PiC) from an emotional state back into a state of reason, to make them susceptible to behavioural influence from the negotiator (Vecchi et al., 2005). The Behavioural Influence Stairway Model (BISM) is often taught. It was originally developed by the FBI for general crisis negotiation and delineates a process through relationship-building between negotiator and PiC by using interpersonal skills, until ultimately making the PiC susceptible to rational arguments for behavioural change (e.g., not kill themselves) (Dalfonzo, 2002; Noesner & Webster, 1997; Webster, 1998a, 1998b; as cited in Vecchi et al., 2005). Vecchi et al. (2019) adapted the model based on insights about persons with suicidal intentions specifically.

The original BISM is made up of five stages. The stages are progressed through cumulatively and sequentially, always starting at stage 1 (*Active Listening*) and aiming for reaching stage 5 (*Behavioral Change*). Stage 3 (*Rapport*) can only be reached through the onset and the upkeep of *Active Listening* and stage 2 (*Empathy*). Maintaining empathy and active listening helps with the progression into stage 4 (*Influence),* where the negotiator is expected to show concrete action perspectives that change the intention to engage in undesired behaviour (e.g., suicide) into engagement in the desired behaviour (e.g., PiC removes himself from the direct suicidal threat (Vecchi et al., 2005). The model allows for alternation between the stages (e.g., when there is a need for revisiting a previous stage or an error was made). The model was updated to change active listening skills into an underlying skill the negotiator should apply throughout the negotiation instead of only being the first step (see *Figure 1*). Also, the *Rapport* stage was extended with *Trust*. As soon as rapport

transforms into trust, negotiators should change their frame of reference. This is accounting for the fact that the building of trust can already start when starting to build rapport (Vecchi et al., 2019). However, the logical sequence of the five steps from the original model stays the same.

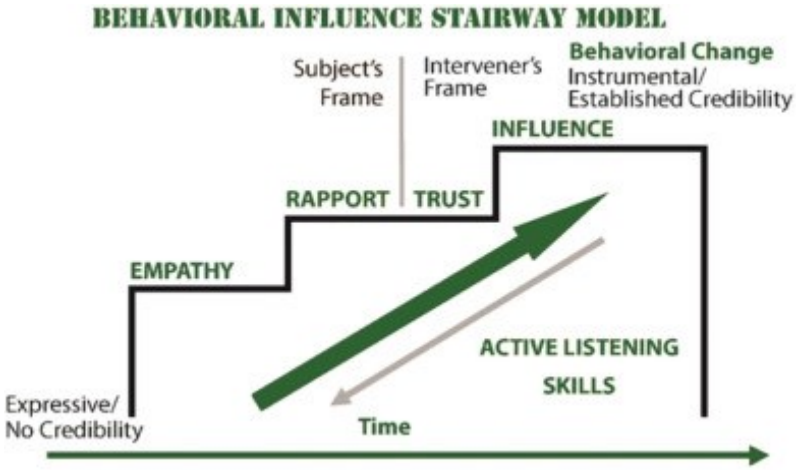


Figure 1. Behavioural Influence Stairway Model (updated version; Vecchi et al., 2019)

A basic requirement for a VE to be effective for suicide negotiation training would be that it includes an interaction where the user can apply the BISM. Apart from that, it might surpass role-plays in some qualities since it can create an exactly repeatable virtual experience. VR can depict every situation with a stereoscopic 3D-view, throwing the user 'into' the situation. As such, it would represent a new technology used for training, falling under the umbrella term *digital game-based learning.*

Digital Game-based Learning & Motivation

In the present case, the aim is to assess qualities of a new VR tool for suicide negotiation training centred around the BISM. Digital game-based learning in general is defined to be a method of instruction, merging video games with educational content, aiming at a higher engagement among learners (Bahadoorsingh et al., 2016). Central elements,

important in digital game-based learning, are the engagement of learners, their intrinsic motivation, and the aspect of experiential learning (Udeozor et al., 2021). In line with experiential learning hypothesis (Kolb & Kolb, 2018), engagement of learners aims to leave an 'impact' on the learner. Precisely, this reflects the 'Active Experimentation' phase (Kolb & Kolb, 2018). The aspect of intrinsic motivation picks directly up on this. Learners are said to be engaged in games for nothing but the intrinsic motivation to play the game itself (Garris, Ahlers, & Driskell, 2002). As such, it would benefit the training process and learning outcomes if usage of the VR is motivating learners to learn *and* motivating them to use VR.

VR, as a media of digital game-based learning, poses a possible entry for learners into a cyclical process, as it has been shown that engagement is predicting motivation and vice-versa (Martin et al., 2017). However, such an entry would largely be dependent on the VE`s potential to make users perceive the depicted situation as believable and close to reality. To assess whether a VE succeeds in making the experience believable and whether it can create meaningful, engaging experiences, different variables can be focused.

VR Training for Suicide Negotiation

Immersion

Variables relating to this quality of the VE relate to the ability of VR to fully *immerse* the user in the experience (Witmer & Singer, 1998). Immersion refers to the degree to which a person that experiences something (here: VR) is feeling like being absorbed in this experience (Witmer & Singer, 1998). More immersion is a likely to be a perquisite to a higher learning outcome since it provides a more involving experience (Checa & Bustillo, 2019). Aspects that influence immersion, defined in detail in the following, can inform the tailoring along the requirements for a VE intended for performance evaluation after users learned about the BISM.

Realism and Spatial Presence

Realism is a construct that describes the conscious evaluation of a user regarding auditory and haptic realism, as well as graphic fidelity (Saghafian et al., 2020). In other contexts of role-plays, it was already shown that VR can induce realism that is deemed appropriate for practical training application (Jouriles et al., 2009). Spatial presence refers to the perception of being present in another physical location which is differing from the one the user is in the real world (Wirth et al., 2007). As such, the conscious observation and evaluation (realism) is directly related to the feeling of being in another physical location (spatial presence) (Jung & Lindemann, 2021). In order to create an immersive experience, the VE has to be believable when it comes to surrounding appearance.

Co-Presence and Social Presence

A virtual agent, being designed appropriately, might also convey a more believable sense of high risk (e.g., fatality in the case of failing) as a confederate role-playing by showing more realistic emotional expressions, gestures, and body language. Co-presence refers to " the subjective experience of being together with others […]" (Poeschl & Doering, 2015, p. 59). Co-presence is likely important for VR suicide negotiation since it is the perquisite to building a relation with someone to first be with that person in some kind of form. Social presence describes the occurrence of a feeling of another individual being present, indicated by a behaviour, form, or sensory experience (Poeschl & Doering, 2015).

Social presence is important within the context of VR suicide negotiation because it can be understood as a direct precursor to co-presence (Bulu, 2012). Generally, high co-presence might help with the emotional bond that participants/learners form towards the virtual PiC (Wu et al., 2014). Co-presence poses a direct basis for forming a social relationship by making the learner feel to be with a social entity. The perception of a possibility to form a social relationship is crucial for training the BISM, which directly aims

at building such a relationship to change behaviour. This perceived option for relationship- or rapport-building with the virtual PiC is likely based on the levels of co-presence and social presence. Generally, both concepts are closely related.

It might thus be that VR can achieve sufficient levels of social presence and co-presence for enhanced realism. Practitioners gave information and confirmed that role-plays often lack this realism due to learners not taking the task seriously or not playing the role of the PiC realistically. Thus, factors like acting skills or learners' relationship might negatively impact on learning outcome, as they weaken emotional valence in traditional role-plays. It is further supported by research that shows that participants indeed can have negative emotional responses towards virtual agents (Wu et al., 2014). These emotional responses are, in the case of VR, reflected in the concepts of social presence and co-presence, which directly pertain to feelings of being around another human(-like) being. While eliciting empathic reactions is likely associated with co-presence, social presence should be more related to the learner's perception of being able to have a social relationship with the conversant.

Dangers of low Co-Presence or Social Presence. High co-presence scores will likely enhance emotional processes and empathy, increasing effectiveness of the VE and ecological validity of the training. A failure to depict the PiC in a way inducing co-presence and social presence might undermine the learning effect of the VE regarding the empathy skill of the BISM and have negative impact on performance. Further, it might fail to mirror the complex nature of such social interactions in the real-world. Feelings of uncanniness can be induced by characters who do not move realistically and smooth or look "nearly-but-not-quite human" (Pan & Hamilton, 2018, p. 406). This would be likely to negatively impact the degree to which emotional responses can be evoked by the VE, and thus effectiveness.

Thus, it becomes clear that a VE intended to be used for performance evaluation would at best induce high levels of co-presence to be able to induce a feeling of being with a

real PiC that participants can have empathic reactions to. High co-presence would likely facilitate social immersion and facilitate emotional reactions within users when interacting with the virtual agent. Further, the perception of being with a sentient social entity may allow for the user to feel they can build rapport with the PiC.

The Current Study

The current study tested the effectiveness of VR to be a performance evaluation tool within training programmes for suicide negotiation. While previous research identified VR to be a useful tool in training of practical professions` skills (Schmidt Mast et al., 2018), the current study addresses a suicide negotiation VR simulation`s effectiveness for the training evaluation of social suicide negotiation skills. Specifically, realism, immersion, spatial presence, co-presence, and social presence will be tested for their influence on performance, as well as to the degree they are elicited by the new VE. Additionally, attention will be given to empathy and rapport in correlations and qualitative analysis.

Immersion will likely help reflect the high-risk physical environment and be a predictor of performance, based on the notion that higher engagement and involvement in experience facilitate knowledge acquisition/retention. High co-presence and social presence would support simulating the high risk concerning the social environment. Motivation would likely facilitate engagement in repeated practice. To address this, it will also be tested whether immersion etc. are impacting participants motivation to engage in further use of VR for suicide negotiation training.

Methods

Participants

The sample consisted of 47 participants, of which 37 were German, three Dutch, and seven of other nationalities. Out of all participants, 23 were male, 24 female. Participants

were largely recruited by the means of convenience sampling through the social environment of the researchers. Other participants were recruited via the internal participant-recruitment tool of the University of Twente ("SONA"). Participants from SONA received credits that they have to collect to finish their studies themselves. Participants from the social environment of the researchers did not receive incentives. The BMS Ethics Committee approved of the research (reference number: 220428).

Materials

Training

The present VE was used for performance evaluation after participants received a training. The training was a PowerPoint presentation and was created by another student, interested in similar aspects, and working simultaneously on their thesis. The key content was the Behavioural Influence Stairway Model (BISM) (Vecchi et al., 2005). Other content was error management (Oostinga et al., 2017), but is irrelevant for this study. Screenshots of the training can be seen in Appendix A. The BISM is a conversational model that gives a guideline to achieve behavioural change in another person by progressing through different stages. The training gives information on all steps through the model. Participants are instructed to move through the steps in order, which is what they need to apply in the task to successfully complete it.

There were two conditions as to which training participants received. In the *Clear Instruction* condition, the presentation contained concrete behavioural examples of each stage of the BISM (see Appendix B). In the *Without Instruction* condition, there were no concrete examples. Participants had to click through the PowerPoint presentation before entering the VE. A researcher was present to answer questions if any came up.

The VR Scenario

The present VE was built and run in Unity 3D (version 2021.3.8f1). Impressions (screenshots) of it can be found in Appendix C. The main element was a suicide negotiation conversation with the PiC. The participant-PiC interaction is more thoroughly described in the procedure section. The conversation takes place on a rooftop, where the PiC is standing because they intended to commit suicide by jumping. The conversation was designed along the BISM in such a fashion, that a decision-tree (see Appendix D) was created, with each stage of the tree representing a stage in the BISM.

Participants had to go through an interaction with six stages. On each stage, there were two possible options. Thus, participants task was to make the correspondingly correct choices at each stage, based on a previously given to them training on the BISM. Each option is framed in a positive wording to prevent the wording to give away the correct option. As an example, the two possible answers to choose between on one stage are "*Trying to get the person to step down*" and "*Trying to talk to the person*", whereby both alternatives are worded in an active tense and display a path of action that could be in line with the BISM. Remembering previously taught material gave the answer away (i.e., at this point, it is too early for suggesting an action). After making a choice for one option, the corresponding text is displayed (e.g., choosing "*Ask person why he is on the roof*" leads to the UI displaying "I didn't mean to scare you, I am sorry. I am here because I got a call […]"). A textbox displayed the words of the PiC. Under this textbox, the two possible choice options came up when the PiC finished their utterances.

Participants had to choose their answer by pointing their controller towards their choice and clicking a button (see Appendix E. 2nd picture). After having made the choice, the PiC answered, and his words were shown in the textbox again. Further, he showed corresponding movements (mimics, gestures). Movement in the VE was done via a method called teleportation locomotion, which means they had to point to the place they wanted to go

before clicking a certain button and ‘jumping’ there (meaning they just seamlessly teleported there). Participants could always move freely, but only used it for correcting their visual perspective on the UI.

Questionnaires

Game Experience and Negotiation Experience. Negotiation experience and game experience were included to act as possible controlling factors in the analysis. Game experience was measured with two questions originally stemming from research from van Sintemaartensdijk et al. (2022) and were “How many hours do you play videogames per week with a keyboard?” and “How many hours do you play videogames per week with a controller?”. Both questions had to be answered by choosing between “0”, “1-3”, or “more than 3”. Negotiation experience was investigated by simply asking whether participants have experience in negotiation, answerable with either “No”, or “Yes, namely:” followed by a free text entry field, in which they could name what negotiation experiences they had if they had any.

Open Questions. The first questions after participants finished the negotiation in the VE were open questions, asking directly for their experience. They were included in proximity to the actual experience to keep memories recent and facilitate retrieval, possibly enhancing the chance for participants to describe their experiences precisely. These were included to allow for a content analysis. The experiential insights resulting from this could later be integrated into the contextualisation of the quantitative results. The first question was “How did you experience the crisis negotiation?”, followed by “What did you perceive as positive?”, and lastly “What did you perceive as negative?”.

Performance. Performance is measured by number of correct choices in total, deducted from the Unity output data. At each stage of the decision tree, the participant will have to choose between two options, where one option is always in line with the BISM (so,

correct) and one is not in line with the BISM (so, incorrect). As an example, “Who are you? Why are you creeping up on me like that?” can be correctly answered with choosing “*Ask him why he is on the roof*” (reflecting active listening skills), or incorrectly by choosing “*Share the reason why you are on the roof*” (not reflecting active listening skills). From this, the number of correct responses will be counted for each participant, resulting in their performance scores from 0 – 6.

Immersion. The Immersive Tendencies Questionnaire (Witmer & Singer, 1998) was used to measure Immersion. It was slightly adapted to the case at hand, since the original was formulated to measure individual tendencies to becoming immersed, instead of being a measure of a current experience. Therefore, some questions were adapted, while others have been taken out. They were adapted in such a fashion, that all items referring to a person’s tendencies to feel/experience something in general were changed to directly refer to the present VR experience (e.g., “Do you easily get deeply involved in movies or tv dramas?” became “I did easily become involved in the VR experience”). The original questionnaire consisted of 18 items. However, due to inapplicability to the present case, eight items got removed. These included items like “Do you ever become so involved in a TV program or book that people have problems getting your attention?”, or “How physically fit do you feel today?”. All ten resulting items were answerable on a 5-point Likert-scale (1 = *strongly disagree* to 5 = *strongly agree*). The changed items that were used here can be seen in Table F1. The scales` final measure will be the mean score of all included items. Reliability for the scale was good (Cronbach`s $\alpha = .74$).

Realism. To measure the extent to which participants felt the spatial aspects of the VE were realistic, the German VR Simulation Realism Scale (Poeschl & Doering, 2013) was employed. The nature of all items was to ask participants about the naturality or realism in different aspects of the VE (e.g., “Proportions of the virtual space were realistic”). The item

"Virtual humans differed concerning their appearance" was excluded, due to the present VE having only one other virtual human present, as well as to prevent overlap with the social presence and co-presence scales. Also, the item "Ambient sound intensity in the virtual room was…" was excluded since its scale had anchor points that did not fit into the rating scheme (as response options were 1 = *too low* to 5 = *too loud*). The eleven remaining items had to be answered on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*). Reliability of the scale was good (Cronbach`s α = .82).

Spatial Presence. Spatial presence was measured with the Spatial Presence Experience Scale (SPES) (Hartmann et al., 2015). The measurement of spatial presence is intended to capture the feeling of 'physically feeling there' as an indication of VE fit for an immersive simulation. The nature of all items pertained to asking participants whether they perceived the virtual space to be physically present around them (e.g., "I felt like I was actually there in the environment of the presentation"). Participants had to answer the eight items on a 5-point Likert-scale (1 = *strongly disagree* to 5 = *strongly agree*). The final score was the mean score from all items. Reliability of the scale was good (Cronbach`s α = .83).

Co-Presence and Social Presence. Co-presence and social presence were both measured with an adapted version of the Social Presence Scale (Poeschl & Doering, 2015). The adapted items can be seen in Table F2. Both variables indicate social dimensions of immersion, which are likely specifically important to the present case, since it is a social interaction that should foster interpersonal skills, which are applied in social situations. The scale is divided into the two subscales of co-presence (3 items) and social presence (11 items).

The items of the social presence scale pertained to reactions of the PiC and the participant, as well as to the perception of a possibility of interaction. One item got removed, due to inapplicability for the present case ("I was easily distracted by people"). For the co-presence scale, the items pertained to the feeling of another person being present (e.g., "I was

aware that the person in crisis was with me on the rooftop"). Thus, the social presence variable was more geared towards testing the feeling of potential social interaction with the virtual agent, while co-presence is more about the perceived physical presence of another being. The final scores for both scales were the mean scores. Here, social presence and co-presence were handled as two separate scales. This was done since the concepts pertained to two different aspects of the social side of the VE. All the included items had to be answered on a 5-point Likert-scale from 1 = *strongly disagree* to 5 = *strongly agree*.

For social presence, reliability was moderate (Cronbach`s α = .59) when all items were included. After inspection of the inter-item correlation matrix and item total statistics, two items could be pointed out that were worth to exclude to increase reliability. First, the item "The person in crisis` behaviour had an influence on my mood" showed low correlations with other items, while the item-total statistics showed that excluding it results in a slightly higher alpha. Second, the item "Sometimes, the person in crisis was influenced by my mood" also showed many null or negative inter-item correlations, while it was indicated that excluding it would result in a reasonably higher alpha. Importantly, both items included the concept of "mood". Thus, the decision was made to exclude both from the social presence scale, resulting in a good reliability score (Cronbach`s α = .71) Reliability for co-presence was low (Cronbach`s α = .29). After inspection of the item-total statistics, it became clear that the item "I felt alone in the virtual environment" was causing the low reliability, while also not being correlated with the other items. Due to this, this item was excluded, resulting in a higher reliability score for co-presence **(**Cronbach`s α = .71).

Empathy. Empathy was measured with two questions. The questions were formulated based on the notion that empathy can be understood as either an emotion or as a cognitive response (Cameron et al., 2019). Thus, the first resulting item was "I understood why the person in crisis felt the way they did", based on the cognitive understanding. The second item

was “When talking to the person in crisis, I found myself feeling the same emotions they did”. Both items could be answered on a 5-point Likert-scale, ranging from 1 = *strongly disagree* to 5 = *strongly agree.* Reliability of the scale was low (Cronbach`s α = .29). Both items were not significantly correlated (*r* = .22). However, both aspects of empathy are theoretically important, mirroring crucial empathy aspects. Thus, content validity was deemed more important in the present case and the scale was used as it was.

Rapport. Measuring rapport was done with a single item. The item was “The Person in Crisis judged our interaction to be positive during the negotiation”. It could be answered on a 5-point Likert-scale, ranging from 1 = *strongly disagree* to 5 = *strongly agree.* It encapsulates the perception of having formed a good relationship with the PiC, which is central to the concept of rapport (Abbe & Brandon, 2013), and a necessity within the BISM to progress. Whether the participant sees the relationship as positive is here displayed by asking them for their perception of how the PiC would evaluate the relationship. This approach was motivated by notions of mutual understanding in humans regarding relationships, or interpersonal synchrony (Stolk, Verhagen, & Toni, 2016).

Motivation. For the measurement of motivation to use VR for training of suicide negotiation, a new questionnaire was developed. It was oriented at a definition of (intrinsic) motivation by Erhel & Janet (2013), “Inner desire to engage in a task out of interest or amusement, or because of the challenge it poses” (p. 157). Ten items were created which corresponded to either amusement (e.g., “I think using VR applications is fun”), interest (e.g., “I think VR applications are interesting.”), or the challenge aspect (e.g., “I think engaging in challenging tasks in VR is fun”) regarding a VR training for serious situations. All items can be seen in Table F3. All items were answerable on a 5-point Likert-scale from 1 = *strongly disagree* to 5 = *strongly agree*. The reliability for the scale was good (Cronbach`s α = .77).

VR Sickness. Many people feel VR sickness when using VR. It is a form motion sickness resulting from VR usage and was used as a controlling variable here, as it might decrease mental capacity of participants. It was measured with the Virtual Reality Sickness Questionnaire (Kim et al., 2018) and contained 9 items, all corresponding to symptoms of VR sickness. Scoring was done via a certain mathematic formular (see Figure F1). All items were answered on a 4-Point scale from 1 = *none* to 4 = *severe*. Reliability of the scale was good (Cronbach`s α = .74).

Coding of Measures

Due to the recency of VR and the reactive technology assessment field, it was difficult to find cut-off points for all variables (Oh, Bailenson, & Welch, 2018; Brandkamp, 2000). Additionally, effectiveness is here understood as baseline-effectiveness, as the VE is a first prototype and improvements are expected to be made in the future. Therefore, the approach will be to test whether the current VE is producing enough immersion, realism, spatial presence, social presence, and co-presence to achieve dispersed scores that do not mostly stick to the lower end of the corresponding scales, indicating said baseline-effectiveness in creating a meaningful experience. If possible, these scores will be compared to other VE´s or the qualitative results to make more differentiated claims about effectiveness.

Hardware

The VE was displayed on the Oculus Meta Quest 2, which is a head-mounted display (HMD) by Meta Platforms. It has six degrees of freedom and uses a stereoscopic view. The resolution is 1832 x 1920 pixels per eye and the field of view is 85-to-97 degrees. The built-in displays are LCD displays. Everything was run on was an Acer PREDATOR HELIOS 300 gaming laptop with a NVIDIA GeForce ® GTX1060 graphics card and an intel core i7 (8^{th} generation) processor. The data collection took place in a standard experiment room of the University of Twente. Within the room, there was a corner desk in one corner. While going

through the VE, the participant stood before this desk, wearing the HMD. On the other side of the room, about two to three meters away from the corner desk, there was another table, where the researcher was sitting while the participant was in the VE (see Appendix G). Questionnaires and consent form were administered on a different, unspecified Laptop.

Procedure

Pre-VR

First, participants were verbally told about the procedure and asked whether they had concerns or questions, next to being told that they could withdraw at any time. Second, participants were given the consent form where they read about their right to withdraw, possible related risks, and got information on their possibility to take off the HMD at any time if they felt motion sick (see Appendix H). Subsequently, they were asked to answer the first page of the questionnaire (demographics, questions about gaming experience, negotiation experience, participant number). Followingly, participants were given the training corresponding to their condition, meaning either a training with or without concrete examples for each BISM stage. Subsequently, the researchers gave each participant the same instructions about how to navigate in the VE (see Appendix I for the complete guideline that researchers used) and how to select options in the interaction with the PiC. Also, it was mentioned that they could take off the HMD at any point if they feel VR sickness. Lastly, participants were reminded to use their knowledge from the training and that they will play the role of a suicide negotiator. Afterwards, participants were equipped with the HMD and the VE was started. The researcher remained in the room, in case of any problems, participants struggling with VR Sickness, or questions.

During VR

Within the VE, participants were first spawned into a starting area where they could look around and received an "emergency call" ("We have an emergency. There is a man that is apparently drunk and suicidal. He wants to jump from a building close to you. Please go there quickly."). This starting sequence was intended to get participants used to being in the VE before the task started. Followingly, participants were verbally instructed by the researcher to go towards the "Bar" sign, which was big and red, placed in the VE to have an easily discernible reference point. As soon as they reached a certain point (a crossroads in front of the building), they got teleported to the rooftop. Here, they first looked in such a direction, that they had to turn by 180 degrees to see the PiC. There, participants had to turn around, then approach the PiC, which triggered the start of the interaction. From there, participants had to make six choices about how to behave. All choices mirrored the stages of the BISM. The first choice pertained to how to approach the PiC (*Approach him* vs. *Asking to approach him*). This related building empathy, as the negotiator shows respect for the PiC by asking for his permission, acknowledging his emotional situation.

Thus, the second choice (*Share the reason why you are on the roof* vs. *Ask him why he is on the roof*) mirrored the active listening skills that are important throughout the whole model but should still start to be used at the begin of the interaction. The correct option (*Ask him why he is on the roof*) should reflect a desire to understand the PiCs motives, thus aligning with the empathy stage of the BISM. Both options here appear reasonable at first. However, since the BISM puts emphasis on first understanding the PiC, an attentive learner might remember this and will opt for listening to the PiC first. In the third choice (*Try to get the person to step down* vs. *Try to talk to the person*), the correct choice (*Try to talk to the person*) reflects adherence to the BISM steps, as opposed to directly skipping to suggesting behavioural change (*Trying to get the person to step down*). Therefore, this choice also tests whether the participant learned that he should progress through the BISM sequentially.

The fourth choice (*Showing that you care for his safety* vs. *Showing empathy for his situation*) reflects the empathy stage. The correct answer (*Showing empathy for his situation*) is more deeply rooted in the BISM, as it also literally entails the empathy aspect. However, both choices lead to an unpleasant reaction from the PiC. This was done to reflect the fact that in real life, it can often occur that PiC`s show unexpected reactions, and it is important that people respond appropriately, as was trained in the error management part of the training.

Afterwards, participants had to make the fifth choice (*Mention consequences of suicide to him* vs. *Acknowledge his feelings*), which pertained to the rapport stage, as acknowledging one's feelings is an interpersonal action that builds rapport based on empathy. At this point, the framing should switch into the positive and the negotiator should align with the PiC. This is directly affecting an interpersonal relationship by possibly creating a confirming and supportive atmosphere (Redmond, 1989). The correct choice shows the negotiators understanding of the PiC's feelings.

The sixth and final choice was about the Influence/Behavioural Change stage. Here, the participant was presented with the options "*Ask him to step down*" and "*Offer him help*". While both seem to aim at a behavioural suggestion, the aim of the BISM is that the negotiator works together with the PiC to find a solution. Within the BISM, it is suggested to best come to a mutual decision about the change of behaviour, instead of posing it as a one-sided request. "*Ask him to step down"* is intended to portray the proposing of a unilateral rather than a mutually agreed on solution, contradicting, or at least not in line with the BISM.

Post-VR

After the PiC showed his reaction to the participants last choice, a researcher verbally informed the participant that the VR-part is over, and that the HMD can be taken off. As soon as the HMD was removed from the participants head, they were seated in front of a laptop to answer the questionnaire. Before they started with answering, the researcher told them that

they will be left alone in the room for this to prevent distraction. They were instructed to step out of the room (where researchers were waiting) if they need to ask questions or when they are finished.

Debrief

After participants gave the signal that they were finished, they were asked whether there are any questions regarding the research in general. Further, it was explained to participants that they can contact the researchers at any time if questions arise later or if they want to have access to their data, or want the data deleted. The researchers thanked participants for their participation and dismissed them.

Data Analysis

First, descriptive statistics will be inspected to check for levels of VR-related variables induced by the VE. It will be tested whether there is a correlation between VR-related variables and performance, motivation, empathy, and rapport in the VE. This will allow for checking the predicted associations between VR-related variables and outcome variables, assessing covariance, and testing predicted associations between co-presence and social presence and empathy. Also, a multiple regression was intended to be run through SPSS to check for statistically significant relationships from the VR-variables together and performance, as well as for their relationship with performance on their own. The multiple linear regression would have allowed for making predictions about performance based on the VR-related variables. However, as there was a ceiling effect in performance towards the higher end of the scale, the variable was later transformed into a dichotomized variable (*All Correct* vs. *Not All Correct*), and a logistic regression was carried out. Further, a multiple linear regression was carried out to investigate whether the VR-related variables on their own and together statistically significant influence motivation. An exploratory content analysis was carried out on participants open question answers to triangulate results.

Content Analysis - Coding Process

Answers from the three open questions had to be coded. The coding process in the present study was first to be conducted as inductive. However, while going over the data coding, it was noticed that many sequences of participants answers reflected the quantitative variables. To test how well the already identified predictors were perceived by participants (as reflected in their answers), it was therefore decided to opt for a deductive approach, informed by the quantitative measures. This approach seemed more promising for finding results that map onto the variables in the present case (Azungah, 2018).

First, familiarization with the data was conducted by multiple readings. Codes were developed based on the variables measured in the quantitative part before investigating how these were experienced by participants in their own words. Subsequently, with the conceptualizations in mind, the researcher went over the answers again to apply the codes. The units of analysis were sequences of participant answers to open questions that expressed a single idea relevant to the corresponding code, regardless of length. As an example, the code *Immersion* was assigned to sequences that directly mention immersion (e.g., "Immersiveness and sound design" as an answer to "What did you perceive as positive?") or answers indicating some kind of emotional involvement (e.g., "at first a bit scary, but after some time I felt relaxed and thought it was quite cool to experience.").

Codes were applied in such a fashion, that an answer from one participant to one question could contain multiple codes. This would occur when two ideas came up in proximity and close relation to each other in a participant's words. For example, "the emotions of the person in crisis were changing" got assigned with *Empathy* and *Co-Presence*. *Empathy* was applied because the words show that the participant could understand the emotions of the PiC, while *Co-Presence* was applied because the answer referred to an actual person, instead of a virtual agent. The total amount of answers to the three open questions was

146. Within those, 73 units of coding were found that could be coded with codes deducted from the quantitative measures. Many of the remaining, uncoded answers were irrelevant for the purpose of evaluating this VE and VR´s effectiveness in the context of suicide negotiation (e.g., “interesting” was often given as an answer to all the open questions, referral to the presentation).

Codes from initial inductive analysis

The initial inductive analysis run resulted in a few codes that were common among answers but were discarded after opting for the deductive approach. Some themes were participants mentioning the experience having been too short, too easy, or the previous training presentation. Important to note is that, in this scheme, the five most common themes were similar ones as those in the final deductive approach (e.g., *Empathy*).

Results

Descriptive Statistics

Descriptive statistics can be observed in Table 1. For the performance measure, two datapoints were missing, thus two participants got excluded. Overall, performance scores were high, with most participants giving the correct response option 5 or 6 times out of 6 (M = 5.50, SD = 0.77). The distribution of performance scores was displaying a ceiling effect. To account for this skew, a dichotomized variable, *Dichotomized Performance*, was created. It depicts whether a participant always picked the correct answer, showing that more than half of the participants got everything correct (n = 25) while less than half of the participants had made at least one mistake (n = 17). This dichotomized variable allowed for a logistic regression. The scores for immersion were normally distributed (M = 3.09, SD = 0.58). Comparably higher scores were observed on realism (M = 3.66, SD = 0.58), social presence (M = 3.99, SD = 0.48), and motivation (M = 4.37, SD = 0.51). VR sickness scores were a little below the midpoint of the scale (M = 42.78, SD = 9.96) and strongly skewed to the right.

Thus, most people felt a little cyber-sick, but a small minority felt very sick. Spatial presence was normally distributed (M = 3.05, SD = 0.74). Co-presence was approximately normally distributed (M = 3.44, SD = 0.84).

Table 1

Descriptive Statistics of Variables

	M	*SD*	*N*
Performance (0 – 6)	5.50	0.77	42
Immersion (1 – 5)	3.09	0.58	44
Realism (1 – 5)	3.66	0.58	44
VR Sickness (0 – 100)	42.78	9.96	44
Spatial Presence (1 – 5)	3.05	0.74	44
Co-Presence (1 – 5)	3.44	0.84	44
Social Presence (1 – 5)	3.99	0.48	44
Motivation (1 – 5)	4.37	0.51	44
Empathy (1 – 5)	3.60	0.70	44
Rapport (1 – 5)	3.70	0.82	44

Correlations

All correlations table can be observed in Table 2. Performance was significantly negatively correlated with realism (r (42) = -.36, p = .01) and motivation (r (42) = -.32, p = .02). However, after exclusion of an outlier scoring high on motivation and realism, but low on performance, and running a sensitivity analysis, this effect disappeared. Further, immersion was significantly positively correlated with the following variables: realism (r (42) = .34, p = .01), spatial presence (r (42) = .66, p < .001), co-presence (r (42) = .60, p = .002), and social presence (r (42) = .30, p = .03). Those correlations with immersion were expected, as

literature indicates that the variables feed into the concept of immersion. Unexpectedly, performance was negatively correlated with realism and motivation.

Table 2

Full Correlations Table

	1	2	3	4	5	6	7	8	9
1.Performance									
2. Immersion	-.15								
3. Realism	-.36*	.34*							
4. VR Sickness	-.17	-.12	-.33*						
5. Spatial Presence	-.25	.66*	.61*	-.12					
6. Co-Presence	-.12	.60*	.45*	.02	.68*				
7. Social Presence	-.20	.44*	.31*	.14	.41*	.44*			
8. Motivation	-.32*	.13	.33*	.00	.18	.05	-.10		
9. Empathy	-.15	.19	.13	.10	.14	.27	.37*	.02	
10. Rapport	.06	-.15	-.01	-.03	-.08	-.01	-.14	.09	-.15

Note. *= $p < .05$, $N = 44$

Mean difference between Training Conditions

To make sure there were no differences in the VR-related variables dependent on the experimental condition, a simple comparison of means between the two groups was carried out. In the "*Clear Instruction*" condition were 23 participants, while in the "*Without Instruction*" included 21 participants. an independent samples t-test was carried out. There were single outliers in realism, VR sickness, and motivation, and two outliers in co-presence, observed in boxplots. However, these outliers were determined to be genuine, and thus kept in the analysis to reflect diversity. There were some minor violations of the assumption of normality, as assessed by the Shapiro-Wilk test. Given the robustness of t-tests against minor

violations of normality, the test was still carried out. See the table below for results. No significant differences were found in the VR-related variables between the two training conditions, apart from immersion*, t(42)=2.42, p = .02.* Immersion in the clear instruction condition (*M* = 3.28, *SD* = 0.51) was higher than immersion in the without instruction condition (*M* = 2.88, *SD* = 0.60).

Table 3

Results of T-test for Mean Differences between Training Conditions

	Clear Instruction		Without Instruction		*t*	*df*	*p*
	M	*SD*	*M*	*SD*			
Immersion	3.28	.51	2.88	.60	2.42	42	.02
Realism	3.64	.53	3.68	.63	-.27	42	.79
VR Sickness	41.78	8.53	42.89	11.44	-.70	42	.49
Spatial Presence	3.16	.56	2.93	.89	1.06	42	.30
Co-Presence	3.61	.52	3.26	1.07	1.39	42	.17
Social Presence	4.00	.46	3.94	.49	.67	42	.51
Motivation	4.33	.46	4.41	.57	-.66	42	.51

Logistic Regression

Since the data violated the assumptions for the multiple linear regression, a substitute logistic regression was carried out with the aim of predicting the probability of participants ending up as having made all choices. Like the initial multiple linear regression, the logistic

regression did not result in any significant results. Based on the Box-Tidwell (1962) procedure, linearity of continuous variables was assessed with respect to the logit of the dependent variable. A Bonferroni correction with 15 terms included was performed and showed that all relations between independent variables and the dependent variable are linear. The Case-wise List showed one standardized residual with the value of 2.368. This was kept in the analysis.

A logistic regression analysis was performed to assess the effects of immersion, realism, spatial presence, co-presence, social presence, VR sickness, and motivation on the probability of participants to end up in the condition of having all choices correct. It resulted in a statistically non-significant logistic regression model, $\chi2(7) = 7.52$, $p = .38$. The logistic regression model explained 22.1% (Nagelkerke R^2) of variance in *All Correct* and was able to correctly classify 71.4 % of cases. Specificity was 58.8%, sensitivity was 80%. The negative predictive value was 66.67% and the positive predictive value was 74.07%. None of the variables was a statistically significant predictor (see Table 4).

Table 4

Table of Regression of dichotomized Performance Variable on IV's

						95% CI for Odds Ratio	
	B	*S.E.*	Wald	*p*	Odds Ratio	Lower	Upper
Immersion	.23	.90	.07	.80	1.27	.22	7.43
Realism	-1.30	.98	1.77	.18	.27	.04	1.86
VR Sickness	-.06	.04	2.01	.16	.94	.86	1.02

Spatial Presence	.12	.83	.02	.89	1.12	.22	5.67
Co-Presence	.53	.64	.70	.40	1.70	.49	5.89
Social Presence	-.41	.89	.22	.64	.66	.12	3.77
Motivation	-.86	1.00	.73	.40	.43	.06	3.03

Multiple Linear Regression – Motivation

Diagnostic tests revealed that there are linear relationships between motivation and all the other variables. There was independence of observations, as shown by the Durbin-Watson test with a statistic of 1.77. Case wise diagnostics showed one outlier. However, the case showed no hints to it being anything different than genuinely extreme data point. As the leverage the case exerts on the result (0.17) can be deemed as safe because it lies below 0.2 (Huber, 1981). Inspection of the Cook`s statistic further showed that the case was not an overly influential point (Cooks distance = .50) (Cook & Weisberg, 1982). The case was kept in the analysis. The plot of standardized predicted values against standardized residuals showed that there is homoscedasticity. Inspection of the P-P plot showed residuals are approximately normally distributed (see Appendix J). The variance inflation factor was below 4 for all variables and tolerance was above 0.25 for all variables, so there was no sign for multicollinearity (Belsley et al., 2005).

The variable performance was excluded from this analysis, due to an apparent ceiling effect and resulting in no different result when included. A non-significant regression equation was found ($F(6, 37) = 1.60$, $p = .18$), with a R^2 of .21. There were no individual predictors observable in the result. The full regression results can be observed in the following table.

Table 5

Regression Table of Regression of Motivation on IV without Performance

	Unstandardized		Standardized	*t*	*p*	95% CI for *B*	
	B	*S.E.*	Coefficients *B*			Lower	Upper
Immersion	.23	.18	.26	1.25	.22	-.14	.60
Realism	.31	.18	.36	1.77	.09	-.05	.67
VR Sickness	.01	.01	.26	1.59	.12	-.00	.03
Spatial Presence	.08	.17	.12	.48	.63	-.27	.44
Co-Presence	-.11	.13	-.17	-.82	.42	-.37	.16
Social Presence	-.19	.19	-.18	-1.04	.30	-.57	.28

Explorative Deductive Content Analysis

There were three open questions given to participants at the end of the survey ("How did you experience the crisis negotiation?", "What did you perceive as positive?", and "What did you perceive as negative?"). An overview of the total frequency of all codes, including the ones from the previous inductive approach, can be seen in Appendix K. The five codes that were applied in the final, deductive approach were *Rapport* (frequency = 17), *Immersion* (frequency = 16), *Realism* (frequency = 15), *Empathy* (frequency = 14), and *Co-Presence*

(frequency = 13). Answers within the *Rapport* code reflected aspects of social presence (e.g., expressions of perceptions of possible relationship-building).

Rapport

All excerpts from the answers of participants that have been coded with the *Rapport* code entailed some reference to the perception of a possibility, or of the actual presence, of a social relationship between the participant and the virtual PiC. This code also included direct and indirect references to interpersonal psychological variables, like trust, influencing the relationship between participant and PiC (e.g., "[…] the PiC stayed calm and also opened up and gained trust easily"), which were pointing towards a relationship. Many answers under this code indicated a feeling of the relationship getting more positive during the interaction, for example mentioning the PiC opening up (e.g., "*[...]* he opened himself up for conversation with me.", or "The Person in Crisis was able to open up to me and shared their negative grievances […]*").*. Words coded with this code are mostly about the reactions of the virtual agent concerning the (hypothetical) relationship between virtual agent and participant.

Thus, participants mostly referred to the virtual agent as an actual entity being capable of having reactional capabilities. As such, they seemed to perceive them as being capable of feeling comfortable because of the participants choices (words). This is mirroring aspects of social presence (seeing possibility for social interaction). Further, it was shown that a critical element for many participants in referring to the relationship was the provision of information (e.g., "opening up") by the entity they were facing. Consequently, they also let shine through that the information they get provided by the entity is dependent on the quality of their reciprocal relationship. This is mirroring established theories of rapport stemming from the field of criminal psychology, stating that the quality of rapport influences the information yield (Brimbal et al., 2021). Thus, the virtual agent seems to create the illusion of a possible social relationship.

Empathy

Under the code *Empathy,* there were answers that indicated that the participant consciously used or felt empathy in the VR negotiation, as well as one that referred to the participants perception about their own empathic skills ("I was not really satisfied with how I performed, especially considering that I just finished a module on empathy and active listening."). Generally, answers coded with *Empathy* did reflect the participant understanding the emotions of the virtual agent (based off his actions), in contrast to *Rapport*, which was used when participants directly referred to actions of the PiC that might influence a relationship. For example, the excerpt "[…] the emotions of the person in crisis were changing" illustrates the participant being able to see the perspective of the PiC, not furthering what that means for the interaction. This also shows the participants perception to interact with a meaningful entity instead of just a virtual agent. As a contrast, *Rapport* excerpts (e.g., "I tried to remain calm as to not upset the Person in Crisis*"*) allure to some reciprocal exchange of not only words, but the notion that the course of interaction depends on the participants actions. Interesting within the *Empathy* coded excerpts is that human characteristics were ascribed to the virtual agent. Participants did write about the PiC like the virtual agent had actual feelings (e.g., "[…] feelings of the PiC were relatable", or "[…] the emotions of the person in crisis were changing").

Co-Presence

Answers that have been coded with *Co-Presence* included direct referring to the virtual agent as "person*"* or the usage of personal pronouns. Interestingly, there were a lot of overlap between the *Co-Presence* and the *Empathy* code. To feel empathy, a person must perceive the "thing" he interacts with as a conscious entity (e.g., a human being). Still, given the importance of a distinction between empathy and co-presence as concepts in the present study, there should be no single code for both. This is grounded in the fact that co-presence is

taken as a perquisite for empathy, making the former a design characteristic for VE´s while the latter describes the "result" of sufficiently implementing co-presence. Further, most answers did again attribute human characteristics to the virtual agent (e.g., "[…] as the emotions of the person in crisis were changing."). This does underline the notion of co-presence being a "subjective experience of being together with others" (Poeschl & Doering, 2015, p. 59) and indicates a success in achieving levels of co-presence that are high enough to elicit empathic responses. The overlap between *Co-Presence* and *Empathy* coded answers is also a strong indication for the latter, while also speaking for a blurring boundary between them.

Immersion

Most answers having been coded with *Immersion* referenced some kind of emotional reaction to the experience (e.g., "Very […] intensive […]", "[…] somewhat stressful", or "I was nervous to see his reaction and if he decided to jump"). Others directly referenced a derivative of the word "immersion" (e.g., "immersiveness", or "immersive simulation*"*). However, all of them are about some kind of involvement. Not coded here were answers that displayed a direct emotional and empathic response to the avatar. When coding, answers fitted in this code straightforward. For example, "somewhat stressful" (as an answer to first question), or "at first a bit scary" (also first question) both directly indicate a presence of emotional reaction to the experience.

Realism

The code *Realism* was given to all parts of answers that included a derivative of the word *real* (e.g., "realistic", or "reality"), and to one referencing the negotiation as "natural" ("For me it felt really natural when talking to the person*"*), which can be seen as an equivalent to *realistic* in this context. When inspecting all the answers containing *real*, most do mention it in a positive sense concerning the present VE. Some answers indicate that

participants positive experience was greatly influenced by the realism of the simulation (e.g., “It was an interesting experience and it felt like an accurate representation of an real negotiation.”, or “i thought the negotiation itself, even though it did not result in success, still provided a good experience because i experienced the negotiation as a realistic situation […]”).

Generally, the code was applied to objective statements that did not include any emotional reaction to the perceived. This is the main difference to *Immersion,* which was assigned to sequences that did not only include plainly objective statements about how *real* something is. Further, the answers showed that participants appear to judge realism more on the social interaction, instead of in terms of the graphic fidelity etc. of the environment surrounding them. Notably, the scale for realism in the quantitative measures mostly focused on just that, as five items were about optic appearance of the environment, while five were about the optical appearance of the PiC and only one item could be understood as pertaining to the PiC´s behaviour (“The person in crisis in his entirety seemed to be authentic for this occasion*”*). These answers might indicate a fit for VR as a realistic performance evaluation tool that can be used in an experiential learning training setting.

Discussion

The new VE was tested regarding its effectiveness in evaluating performance and fostering skills and knowledge that were taught in a previously administered educational resource. To do so, the present study tried to identify the influences of immersion, realism, spatial presence, social presence, and co-presence on performance in a suicide negotiation training and motivation to use VR for this. There was no evidence for a causal relationship between any of these. Further, it was tried to explore how co-presence and social presence are related to empathy and rapport and whether the present VE could generate sufficient levels of the VR-related variables to be effective in training. Significant correlations were observed

between immersion, realism, spatial, social, and co-presence, as well as between empathy and social presence. All variable mean scores lie above the corresponding scales midpoint and show dispersion rather than sticking to the lower ends of the scales, indicating baseline effectiveness. Qualitative results showed participants feeling empathy and the possibility of a social relationship with the virtual agent, while also reporting considerable amounts of immersion. Additionally, an overlap between co-presence and empathy in qualitative results hints towards an association.

Performance & Motivation

The initial plan to assess influences on performance suffered from the performance variable's display of a ceiling effect. Likely, the reason for this was the simplicity of the task, which is further underlined by the results of the initial deductive content analysis, with some participants directly commenting on the interaction being too easy. Thus, future research should include tasks with heightened difficulty (e.g., more interactions or more answering possibilities). For all demographics, a duration of 10 minutes in VR is suggested to be unproblematic in terms of VR sickness (Petri et al., 2020). Regarding motivation, a high mean score supports the notion high motivation is associated to the usage of this VE. This might possibly lead to the proposed entry into the cyclical learning process (Martin et al., 2017) but remains subject to further investigation. As the present operationalization of motivation was a self-made questionnaire centred around intrinsic motivation to use VR *and* motivation to learn about suicide negotiation training, a more differentiated conceptualization in future research is advisable.

Immersion, Realism, & Spatial Presence

The present study did discover that the new VE is able to elicit reasonable levels on immersion, realism, and spatial presence to be considered having baseline effectiveness. Immersion showed levels that were typical, as compared to another study about VR training

that depicted immersion scores around the mid-point of corresponding scales (Narciso et al., 2019). This comparison gives ground for assuming that participants were immersed enough to have emotions elicited by the VE. Frequent referral to being immersed from participants in the qualitative results further underlines this. The remaining question is whether this immersion provides an advantage over traditional training methods, as comparisons have been shown to result in unexpected ties (Shen et al., 2021). The same is true for realism, as scores laid above the mid-point of the scale, and previous research deemed this to be high enough for practical application in another VR training evaluation (Narciso et al., 2019). Narciso et al. (2019) evaluated a firefighter training, which can be assumed to be more dependent on physical features rather than social features. The present training was focusing on social skills and connected with the notion of too much realism possibly acting distracting (Moser & Bergamin, 2020), it can be assumed that realism (as measured here) was enough for this social-centric training to be effective. This is supported by previous research indicating realism can never be as high in classrooms as in VR (Bowman et al., 2007), which is logical if realism is understood as set out in the methods. This, however, gets challenged when considering the social aspects of immersion. While the association between spatial presence and realism comes apparent when inspecting correlations and was expected, qualitative results show a stronger overlap between social aspects and realism.

Social Presence & Co-Presence

Inspecting correlations, realism ratings were correlated with all types of presence. Triangulating with qualitative results, it becomes clear that many participants put more importance on the appearance and behaviour of the PiC, as opposed to the realism of the environment. This points into the direction of these social aspects being of more importance for the present training type, which needs to be further investigated in future research to set up clear design and evaluation guidelines for suicide negotiation training. Human-computer

interaction methods could help establishing the core aspects of suicide negotiation training to inform future designs (Jeon et al., 2019).

Previous research about learning scenarios had difficulties categorizing the found amount of social presence as low, moderate, or high (Hostetter, 2013). Compared with the other variables, relatively high scores were observed. Levels can be assumed to be enough to support effectiveness. Supporting this are qualitative results, where many participants expressed their perception of a (possible) social relationship to the PiC (*Rapport* code), which indicates they felt social presence. Other research did highlight social presence as a perquisite for emotional experience within VR trainings (Pfaller et al., 2021) and as strengthening relationship-building through non-verbal and verbal cue-transmission (Bickle et al., 2019). This can be observed in the qualitative results, where some participants commented on the PiC's cue-transmission (e.g., "he opened himself for conversation with me"). Found within the code *Rapport*, these are signs for the perception of the possibility of social relationship-building.

While the code *Realism* comes up very frequently across participants answers, most of those answers include a reference to the PiC and his behaviour, which shows the importance of the social aspects regarding realism ratings. Further, the code *Rapport* included many overlaps to the quantitative measure of social presence, as participants often referred to a possibility of building a relationship. This stands in contrast to the missing correlation between rapport and any of the other variables. However, the limitation might have been that the rapport measure was just a single-item scale, which is inferior to multiple-item scales in reliability and validity (Sarstedt & Wilczynski, 2009). Taken together, it becomes clear that the concept of rapport closely maps onto social presence.

Similarly, *Co-Presence* showed a lot of overlap with E*mpathy* in the qualitative results, despite no correlations in quantitative results. Most answers attributing human characteristics

to the virtual agent (e.g., "[…] as the emotions of the person in crisis were changing.*"*) imply that the present VE succeeded in achieving high enough levels of co-presence to be at least baseline effective in eliciting emotional responses. These findings show that during design and evaluation of future VR suicide negotiation trainings, more focus should be laid onto the depiction of the PiC and the interaction, particularly on PiC appearance and behaviour (Yoon et al., 2019), rather than on environmental realism considering the surroundings. Apart from this, the overlaps between social presence and rapport and co-presence and empathy, respectively, might inform future definitions of these concepts. Empathy scores showed that participants had empathic reactions to the virtual agent, consistent with previous research (Paiva et al., 2017). Social presence being the only variable correlated with empathy indicates that for having empathic reactions, it is not only important to see the virtual agent as a human, but a sentient being with which a social bond-forming is possible. As such, present scores and qualitative results give ground for assuming the amount of social presence and co-presence elicited is high enough supporting effectiveness.

Limitations

The number of participants was rather small. A higher number would have likely resulted in higher analysis power (Kreamer & Blasey, 2015). As there were multiple independent variables, a bigger sample size would have likely yielded more reliable results, as literature indicates that a regression equation with six or more predictor variables requires a minimum amount of ten participants for each included predictor variable (Van Voorhis & Morgan, 2007). However, even if the 60 participants per variable were achieved, the low variance in the performance measure would still pose problems to analytical power (Liu & Wang, 2020). Another limitation of the present study pertains to the task duration and task complexity. A longer duration, or more complex task, would have likely prevented the ceiling effect by causing more dispersion in scores. Future research should take this into account.

Most participants were university students. Transference of the present findings onto professional suicide negotiators could thus be difficult, as they are likely past the point of having to memorize the BISM. Negotiators that are new to the field might be a more fitting population for this kind of training. Lastly, it must be said that the empathy measure was included despite low reliability, which might hurt validity of the present findings. Thus, the present findings should be taken cautiously.

Conclusion

The present research assessed a new VE regarding how much immersion, realism, spatial presence, social presence, and co-presence are elicited, and how they are related to empathy and rapport. It was shown that social presence can be reflected in empathy (as shown by correlations and qualitative results) and rapport (as shown by qualitative results). Further, it was investigated if these variables influence performance and motivation. Taken together, immersion, realism, and the different types of presence did not directly influence performance or motivation. However, future research should re-evaluate this with a task of longer duration or a more complex task. This way, the complex social nature of the situation will be captured more fully. The VE did achieve scores on immersion, realism, spatial presence, social presence, and co-presence that do not stick around the lower end of the scale, which can be taken as indicators of baseline-effectiveness. Participants were immersed enough to experience emotional and empathic reactions that might support fostering of skills and knowledge, as shown by qualitative results. Lastly, for the task at hand, especially social aspects of realism seem to be more important in creating meaningful experience in this context, as well as in participants evaluation of how realistic the situation appears.

References

Abbe, A., & Brandon, S. E. (2012). The role of rapport in investigative interviewing: A Review. *Journal of Investigative Psychology and Offender Profiling*, *10*(3), 237–249. doi: 10.1002/jip.1386

Azungah, T. (2018). Qualitative research: Deductive and inductive approaches to data analysis. *Qualitative Research Journal*, *18*(4), 383–400. doi: 10.1108/qrj-d-18-00035

Bahadoorsingh, S., Dyer, R., & Sharma, C. (2016). Integrating serious games into the engineering curriculum - a game-based learning approach to power systems analysis. *International Journal of Computational Vision and Robotics*, *6*(3), 276. doi: 10.1504/ijcvr.2016.077372

Belsley, D. A., Kuh, E., & Welsch, R. E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons.

Bickle, J. T., Hirudayaraj, M., & Doyle, A. (2019). Social presence theory: Relevance for HRD/VHRD Research and Practice. *Advances in Developing Human Resources*, *21*(3), 383–399. doi: 10.1177/1523422319851477

Bowman, D. A., & McMahan, R. P. (2007). Virtual reality: How much immersion is enough? *Computer*, *40*(7), 36–43. doi: 10.1109/mc.2007.257

Box, G. E. P., & Tidwell, P. W. (1962). Transformation of the independent variables. *Technometrics*, *4*(4), 531–550. doi: 10.1080/00401706.1962.10490038

Brandkamp, M. (2000). Verfahren der Technologiebewertung. *Technologien Für Innovative Unternehmensgründungen*, 7–15. doi: 10.1007/978-3-322-91512-2_2

Brimbal, L., Meissner, C. A., Kleinman, S. M., Phillips, E. L., Atkinson, D. J., Dianiska, R. E., Rothweiler, J. N., Oleszkiewicz, S., & Jones, M. S. (2021). Evaluating the benefits of a rapport-based approach to investigative interviews: A training study with law enforcement investigators. *Law and Human Behavior*, *45*(1), 55–67. doi: 10.1037/lhb0000437

Bulu, S. T. (2012). Place presence, social presence, co-presence, and satisfaction in Virtual Worlds. *Computers & Education*, *58*(1), 154–161. doi: 10.1016/j.compedu.2011.08.024

Burch, G. F., Giambatista, R., Batchelor, J. H., Burch, J. J., Hoover, J. D., & Heller, N. A. (2019). A meta-analysis of the relationship between experiential learning and learning outcomes. *Decision Sciences Journal of Innovative Education, 17*(3), 239-273. doi: 10.1111/dsji.12188

Cameron, C. D., Hutcherson, C. A., Ferguson, A. M., Scheffer, J. A., Hadjiandreou, E., & Inzlicht, M. (2019). Empathy is hard work: People choose to avoid empathy because of its cognitive costs. *Journal of Experimental Psychology: General*, *148*(6), 962–976. doi: 10.1037/xge0000595

Checa, D., & Bustillo, A. (2019). A review of immersive virtual reality serious games to enhance learning and training. *Multimedia Tools and Applications*, *79*(9–10), 5501–5527. doi: 10.1007/s11042-019-08348-9

Cook, R. D., & Weisberg, S. (1982). Criticism and influence analysis in regression. *Sociological Methodology, 13,* 313. doi: 10.2307/270724

Erhel, S., & Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education, 67,* 156–167. doi: 10.1016/j.compedu.2013.02.019

Gillies, M., & Pan, X. S. (2019). Virtual reality for social skills training. *Proceedings of the Virtual and Augmented Reality to Enhance Learning and Teaching in Higher Education Conference 2018*, 83–92. doi: 10.1255/vrar2018.ch9

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and Practice Model. *Simulation & Gaming*, *33*(4), 441–467. doi: 10.1177/1046878102238607

Hall, M. A., Camacho, F., Dugan, E., & Balkrishnan, R. (2002). Trust in the medical profession: Conceptual and measurement issues. *Health Services Research*, *37*(5), 1419–1439. doi: 10.1111/1475-6773.01070

Hartmann, T., Wirth, W., Schramm, H., Klimmt, C., Vorderer, P., Gysbers, A., Böcking, S., Ravaja, N., Laarni, J., Saari, T., Gouveia, F., & Sacau, A. M. (2015). The spatial presence experience scale (SPES). *Journal of Media Psychology.*

Hostetter, C. (2013). Community matters: Social presence and learning outcomes. *Journal of the Scholarship of Teaching and Learning, 13*(1), 77-86.

Howard, M. C., & Gutworth, M. B. (2020). A meta-analysis of virtual reality training programs for Social Skill Development. *Computers & Education*, *144*, 103707. doi: 10.1016/j.compedu.2019.103707

Huber, P. (1981) *Robust statistics*. New York: John Wiley.

Jeon, S.-G., Han, J., Jo, Y., & Han, K. (2019). Being more focused and engaged in firefighting training: Applying user-centered design to VR System Development. *25th ACM Symposium on Virtual Reality Software and Technology*. doi: 10.1145/3359996.3364268

Jouriles, E. N., McDonald, R., Kullowatz, A., Rosenfield, D., Gomez, G. S., & Cuevas, A. (2009). Can virtual reality increase the realism of role plays used to teach college women sexual coercion and rape-resistance skills? *Behavior Therapy*, *40*(4), 337–345. doi: 10.1016/j.beth.2008.09.002

Jung, S., & Lindeman, Robert. W. (2021). Perspective: Does realism improve presence in VR? suggesting a model and metric for VR experience evaluation. *Frontiers in Virtual Reality, 2*. doi: 10.3389/frvir.2021.693327

Kim, H. K., Park, J., Choi, Y., & Choe, M. (2018). Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. *Applied Ergonomics*, *69*, 66–73. doi: 10.1016/j.apergo.2017.12.016

Kolb, A., & Kolb, D. (2018). Eight important things to know about the experiential learning cycle. *Australian Educational Leader*, *40*(3), 8–14. doi: 10.3316/192540196827567

Kraemer, H. C., & Blasey, C. (2015). *How many subjects?: Statistical power analysis in research*. Sage Publications.

Liu, Q., & Wang, L. (2020). T-test and ANOVA for data with ceiling and/or floor effects. *Behavior Research Methods*, *53*(1), 264–277. doi: 10.3758/s13428-020-01407-2

Martin, A. J., Ginns, P., & Papworth, B. (2017). Motivation and engagement: Same or different? does it matter? *Learning and Individual Differences*, *55*, 150–162. doi: 10.1016/j.lindif.2017.03.013

Moser, I., & Bergamin, P. (2020). Work-in-progress—designing a multi-user virtual reality environment for conversational skills training of dietetic students. *2020 6th International Conference of the Immersive Learning Research Network (iLRN)*. doi: 10.23919/ilrn47897.2020.9155189

Narciso, D., Melo, M., Raposo, J. V., Cunha, J., & Bessa, M. (2019). Virtual reality in training: An experimental study with firefighters. *Multimedia Tools and Applications*, *79*(9–10), 6227–6245. doi: 10.1007/s11042-019-08323-4

Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, *5*. doi: 10.3389/frobt.2018.00114

Oostinga, M. S. D., Giebels, E., & Taylor, P. J. (2017). Communication Error Management in law enforcement interactions: A receiver's perspective. *Psychology, Crime &; Law*, *24*(2), 134–155. doi: 10.1080/1068316x.2017.1390112

Pan, X., & Hamilton, A. F. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395–417. doi: 10.1111/bjop.12290

Paiva, A., Leite, I., Boukricha, H., & Wachsmuth, I. (2017). Empathy in virtual agents and Robots. *ACM Transactions on Interactive Intelligent Systems*, *7*(3), 1–40. doi: 10.1145/2912150

Petri, K., Feuerstein, K., Folster, S., Bariszlovich, F., & Witte, K. (2020). Effects of age, gender, familiarity with the content, and exposure time on cybersickness in immersive head-mounted display based Virtual Reality. *American Journal of Biomedical Sciences*, *12*(2) 107–121. doi: 10.5099/aj200200107

Pfaller, M., Kroczek, L. O., Lange, B., Fülöp, R., Müller, M., & Mühlberger, A. (2021). Social presence as a moderator of the effect of agent behavior on emotional experience in social interactions in virtual reality. *Frontiers in Virtual Reality, 2.* doi: 10.3389/frvir.2021.741138

Poeschl, S., & Doering, N. (2013). The German VR Simulation Realism Scale-Psychometric Construction for Virtual Reality Applications with Virtual Humans. *Annual Review of Cybertherapy and Telemedicine, 11*, 33-37.

Poeschl, S., & Doering, N. (2015). Measuring co-presence and social presence in virtual environments–psychometric construction of a german scale for a fear of public speaking scenario. *Annual Review of Cybertherapy and Telemedicine 2015*, 58-63.

Redmond, M. V. (1989). The functions of empathy (decentering) in human relations. *Human Relations*, *42*(7), 593–605. doi: 10.1177/001872678904200703

Saghafian, M., Laumann, K., Akhtar, R. S., & Skogstad, M. R. (2020). The evaluation of virtual reality fire extinguisher training. *Frontiers in Psychology*, *11*. doi: 10.3389/fpsyg.2020.593466

Sarstedt, M., & Wilczynski, P. (2009). More for less? A comparison of single-item and multi-item measures. *Die Betriebswirtschaft*, *69*(2), 211.

Schmid Mast, M., Kleinlogel, E. P., Tur, B., & Bachmann, M. (2018a). The future of Interpersonal Skills Development: Immersive virtual reality training with virtual humans. *Human Resource Development Quarterly*, *29*(2), 125–141. doi: 10.1002/hrdq.21307

Schmid Mast, M., Kleinlogel, E. P., Tur, B., & Bachmann, M. (2018). The future of Interpersonal Skills Development: Immersive virtual reality training with virtual humans. *Human Resource Development Quarterly*, *29*(2), 125–141. doi: 10.1002/hrdq.21307

Shen, S., Chen, H.-T., Raffe, W., & Leong, T. W. (2021). Effects of level of immersion on virtual training transfer of Bimanual Assembly tasks. *Frontiers in Virtual Reality*, *2*. doi: 10.3389/frvir.2021.597487

Stolk, A., Verhagen, L., & Toni, I. (2016). Conceptual alignment: How brains achieve mutual understanding. *Trends in Cognitive Sciences*, *20*(3), 180–191. doi: 10.1016/j.tics.2015.11.007

Udeozor, C., Toyoda, R., Russo Abegão, F., & Glassey, J. (2021). Perceptions of the use of virtual reality games for Chemical Engineering Education and professional training. *Higher Education Pedagogies*, *6*(1), 175–194. doi: 10.1080/23752696.2021.1951615

Van Hasselt, V. B., Romano, S. J., & Vecchi, G. M. (2008). Role playing. *Behavior Modification*, *32*(2), 248–263. doi: 10.1177/0145445507308281

Van Sintemaartensdijk, I., van Gelder, J.-L., van Prooijen, J.-W., Nee, C., Otte, M., & van Lange, P. (2022). Assessing the deterrent effect of symbolic guardianship through neighbourhood watch signs and police signs: A virtual reality study. *Psychology, Crime & Law, 30*(1) 1–21. doi: 10.1080/1068316x.2022.2059480

Van Voorhis, C. R., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology*, *3*(2), 43–50. doi: 10.20982/tqmp.03.2.p043

Vecchi, G. M., Van Hasselt, V. B., & Romano, S. J. (2005). Crisis (hostage) negotiation: Current strategies and issues in high-risk conflict resolution. *Aggression and Violent Behavior*, *10*(5), 533-551.

Vecchi, G. M., Wong, G. K. H., Wong, P. W. C., & Markey, M. A. (2019). Negotiating in the skies of Hong Kong: The efficacy of the behavioral influence stairway model (BISM) in suicidal crisis situations. *Aggression and Violent Behavior*, *48*, 230–239. doi: 10.1016/j.avb.2019.08.002

Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., Saari, T., Laarni, J., Ravaja, N., Gouveia, F. R., Biocca, F., Sacau, A., Jäncke, L., Baumgartner, T., & Jäncke, P. (2007). A process model of the formation of spatial presence experiences. *Media Psychology*, *9*(3), 493–525. doi: 10.1080/15213260701283079

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. doi: 10.1162/105474698565686

Wu, Y., Babu, S. V., Armstrong, R., Bertrand, J. W., Jun Luo, Roy, T., Daily, S. B., Cairco Dukes, L., Hodges, L. F., & Fasolino, T. (2014). Effects of virtual human animation on emotion contagion in simulated inter-personal experiences. *IEEE Transactions on Visualization and Computer Graphics*, *20*(4), 626–635. doi: 10.1109/tvcg.2014.19

Yoon, B., Kim, H., Lee, G. A., Billinghurst, M., & Woo, W. (2019). The effect of avatar appearance on social presence in an augmented reality remote collaboration. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. doi: 10.1109/vr.2019.8797719

Appendix A

Screenshots Training

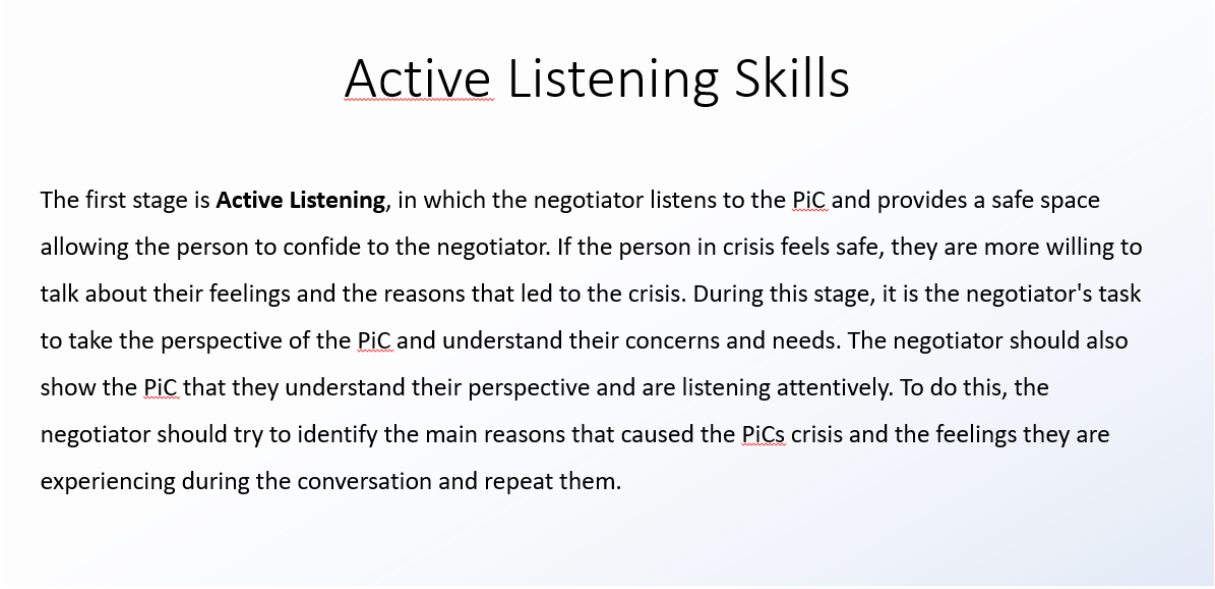


Figure A1. Training without concrete examples

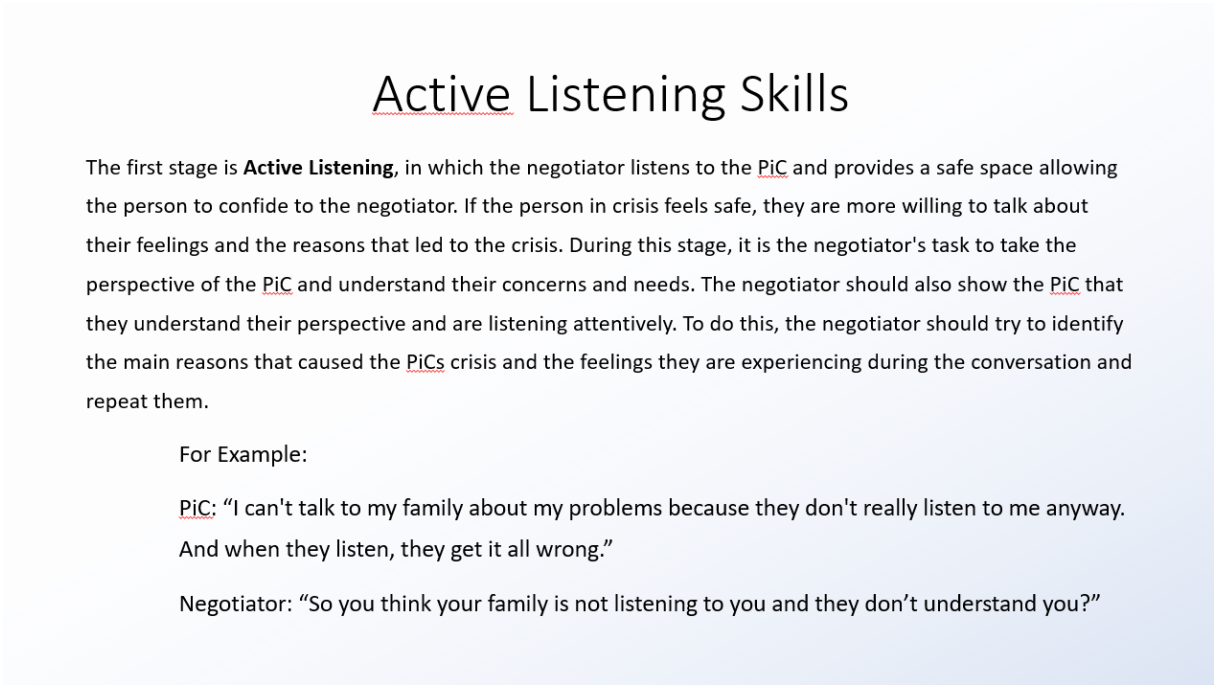


Figure A2. Training with concrete examples

Appendix B

Training with concrete examples

Active Listening Skills

The first stage is **Active Listening**, in which the negotiator listens to the PiC and provides a safe space allowing the person to confide to the negotiator. If the person in crisis feels safe, they are more willing to talk about their feelings and the reasons that led to the crisis. During this stage, it is the negotiator's task to take the perspective of the PiC and understand their concerns and needs. The negotiator should also show the PiC that they understand their perspective and are listening attentively. To do this, the negotiator should try to identify the main reasons that caused the PiCs crisis and the feelings they are experiencing during the conversation and repeat them.

For Example:

PiC: “I can't talk to my family about my problems because they don't really listen to me anyway. And when they listen, they get it all wrong.”

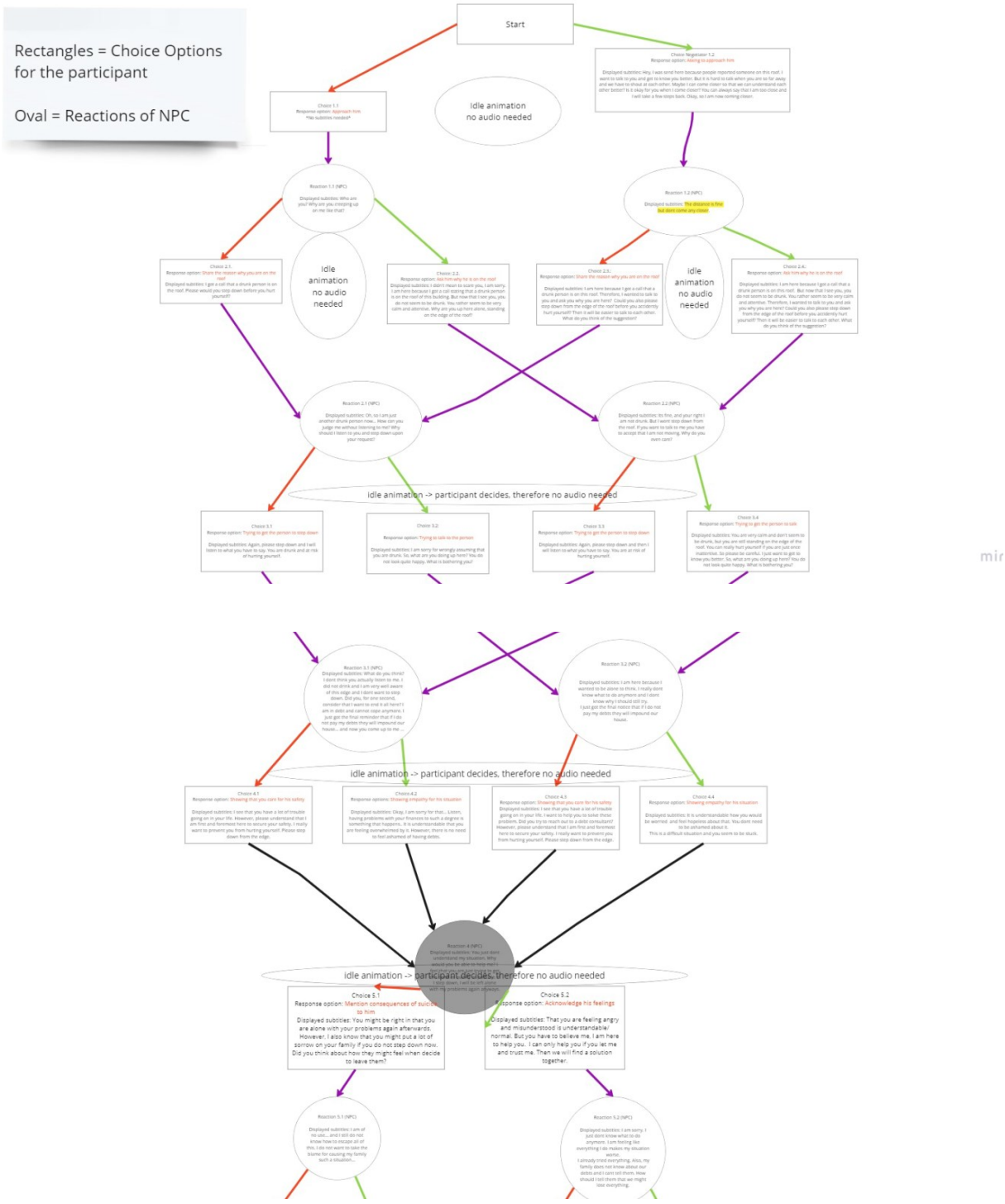
Negotiator: “So you think your family is not listening to you and they don’t understand you?”

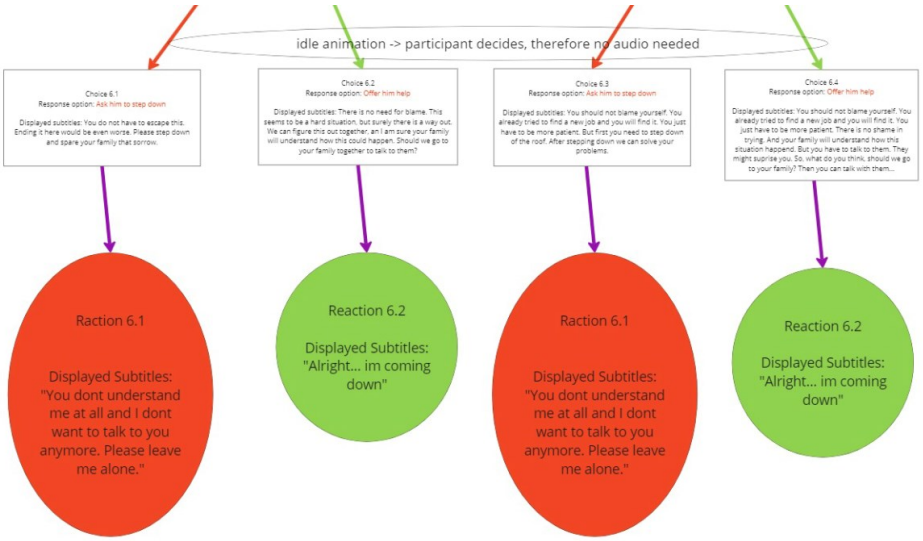
Appendix C

Impressions of the VE (Screenshots)

Appendix D

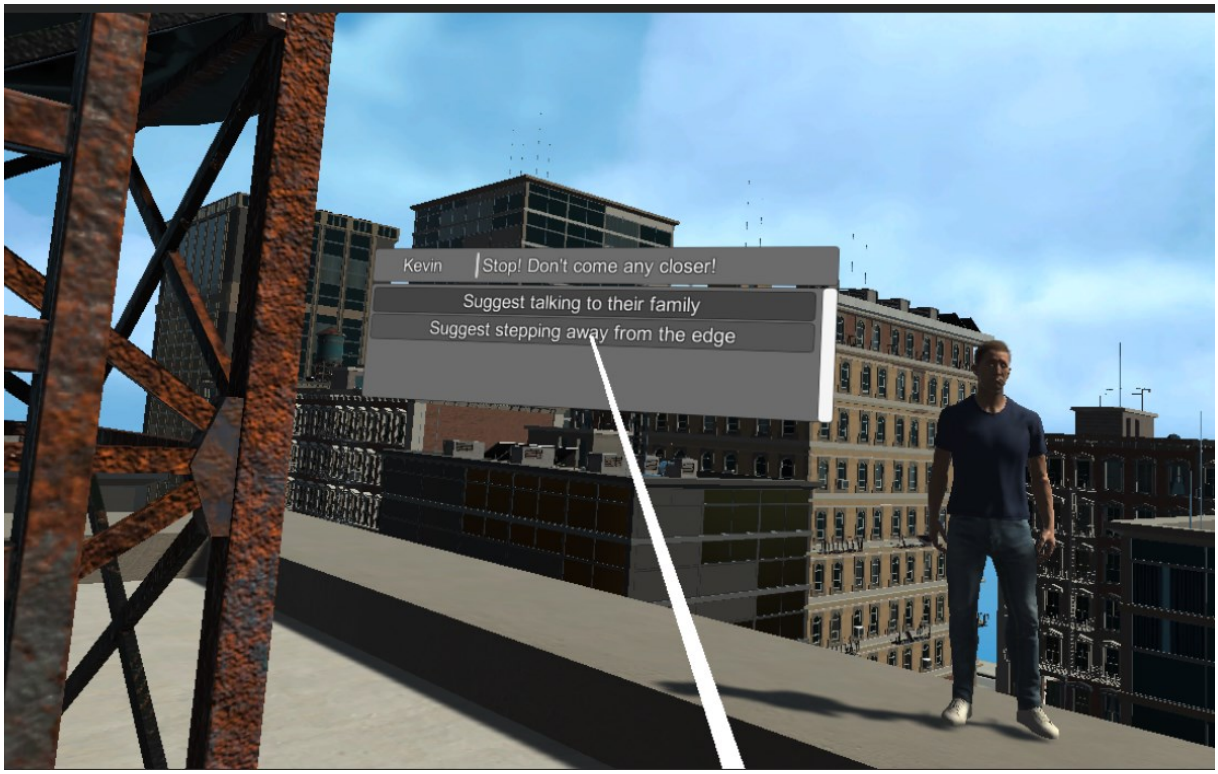
Decision-Tree




idle animation -> participant decides, therefore no audio needed
Choice 6.1
Response option: Ask him to step down
Displayed subtitles: You do not have to escape this. Ending it here would be even worse. Please step down and spare your family that sorrow.
Choice 6.2
Response option: Offer him help
Displayed subtitles: There is no need for blame. This seems to be a hard situation, but surely there is a way out. We can figure this out together, an I am sure your family will understand how this could happen. Should we go to your family together to talk to them?
Choice 6.3
Response option: Ask him to step down
Displayed subtitles: You should not blame yourself. You already tried to find a new job and you will find it. You just have to be more patient. But first you need to step down of the roof. After stepping down we can solve your problems.
Choice 6.4
Response option: Offer him help
Displayed subtitles: You should not blame yourself. You already tried to find a new job and you will find it. You just have to be more patient. There is no shame in trying. And your family will understand how this situation happend. But you have to talk to them. They might suprise you. So, what do you think, should we go to your family? Then you can talk with them...
Raction 6.1
Displayed Subtitles: "You dont understand me at all and I dont want to talk to you anymore. Please leave me alone."
Reaction 6.2
Displayed Subtitles: "Alright... im coming down"
Raction 6.1
Displayed Subtitles: "You dont understand me at all and I dont want to talk to you anymore. Please leave me alone."
Reaction 6.2
Displayed Subtitles: "Alright... im coming down"

Appendix E

Choosing of option



Appendix F

Items

Table F1

Immersion Items – Immersive Tendencies Questionnaire (Witmer & Singer, 1998), adapted to the present case

Original Item	Adapted Item
Do you become easily involved in movies or TV dramas?	I did easily become involved in the VR experience
How mentally alert do you feel at present?	I felt mentally alert during the VR experience
Do you ever become so involved in a movie that you are unaware of things happening around you?	I was so involved in the VR experience that I was not aware of the things happening around me
How frequently do you find yourself closely identifying with the characters in a storyline?	I felt myself identified closely with the character I played in the VR experience
Do you ever become so involved in a video game that it is like you are inside the game rather than moving a joystick and watching the screen?	I felt as if I was really inside the virtual environment, rather than participating in a study
While watching sports, do you ever become so involved in the game that you react as if you were one of the players?	I became so involved in the VR experience that I reacted as if I was actually negotiating a suicide scenario
Do you ever have dreams that are so real that you feel disoriented when you wake up?	The VR experience felt so real that I felt disoriented when putting down the VR headset

While playing sports, do you become so involved in the game that you lose track of time?	I lost track of time during the VR experience
Have you ever got excited during a chare or fight scene on TV?	I felt excited due to what happened during the VR experience
Have you ever got scared by something happening in a TV show or a movie?	I got scared by what was happening during the VR experience

Table F2

Social Presence and Co-Presence: Changed Items

Original Item	Adapted Item
The people's behaviour influenced by style of presentation	The person in crisis behaviour influenced my choice of answer
The people's behaviour had an influence on my mood	The person in crisis behaviour had an influence on my mood
I reacted to the people's behaviour	I reacted to the person in crisis behaviour
Sometimes, the people were influenced by my mood	Sometimes, the person in crisis was influenced by my mood
Sometimes, the people were influenced by my style of presentation	Sometimes, the person in crisis was influenced by my answers
The people reacted to my actions	The person in crisis reacted to my answers
I was able to interpret the people's reactions	I was able to interpret the person in crisis reactions
I had the feeling to interact with other human beings	I had the feeling to interact with another human being
I felt connected to other people	I felt connected to the person in crisis
I had the feeling that I was able to interact with people in the virtual room	I had the feeling that I was able to interact with the person in crisis
I had the impression that the audience noticed me in the virtual room	I had the impression that the person in crisis noticed me in the virtual environment

I was aware that other people were with me in the virtual room	I felt like the person in crisis was really with me on the rooftop/in the VE
I had the feeling that I perceived other people in the virtual room	I had the feeling that I perceived the person in crisis in the virtual environment
I felt alone in the virtual environment	I felt alone in the virtual environment

Table F3

Motivation Scale Items (5-Point Likert)

Item Number	Item Content
1	I think VR Applications are interesting
2	I wanted to do more practice in the virtual space
3	I would use the VR application again to get better at crisis negotiation
4	I think using VR applications is fun
5	I think engaging in challenging tasks in VR is fun
6	I think crisis negotiation is an interesting topic
7	I think that engaging in a “serious” task in a VR environment is fun
8	I want to learn to be better at crisis negotiation
9	If I had the chance, I would want to do more training to become a crisis negotiator
10	I think being challenged by a crisis negotiation situation in VR is fun

A = Sum of questions 1-4: _____ C = A/12 x 100: _____

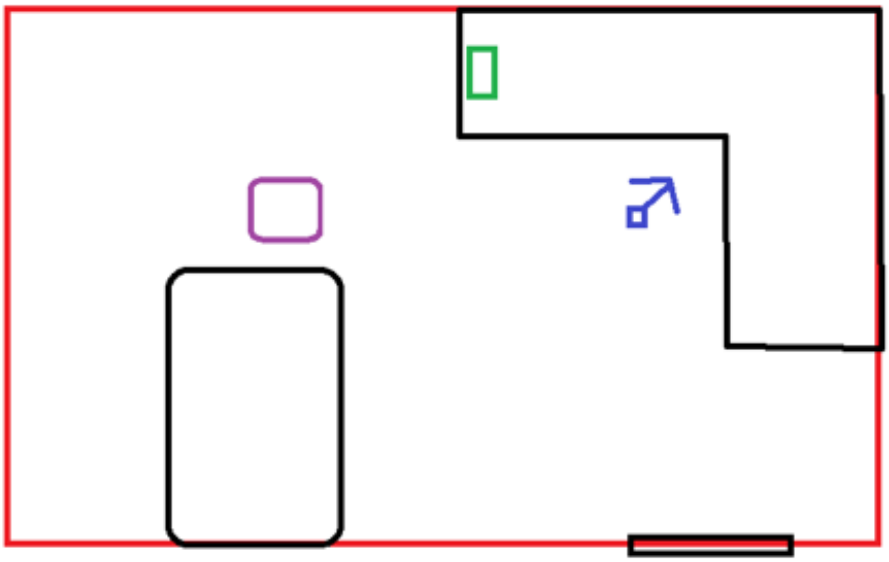
B = Sum of questions 5-9: _____ D = B/15 x 100: _____

VRSQ Score = (C + D)/ 2: _____

Figure F1. Scoring form for the VRSQ (Kim et al., 2019)

Appendix G

Room Layout



Legend:

Purple = Researcher

Black = Door / Tables

Blue = Participant position with indication of looking direction (arrow) while in VR

Green = Laptop for Training

Red = Walls (approx.. 4 x 7 m)

Appendix H

Informed Consent

Consent Form for VIRTUAL REALITY: TRAINING YOUR CRISIS NEGOTIATION SKILLS

YOU WILL BE GIVEN A COPY OF THIS INFORMED CONSENT FORM

	Yes	No
Please tick the appropriate boxes		
Taking part in the study		
I have read and understood the study information dated [DD/MM/YYYY], or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	( )	( )
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	( )	( )
I understand that taking part in the study involves a digital role-play in a Virtual Reality (VR) environment and that the depicted behaviours might cause psychological distress. Further, I understand that my actions within the VR will be captured and used for research.	( )	( )
Risks associated with participating in the study		
I understand that taking part in the study involves the following risks: 1. Usage of VR might cause nausea or cybersickness for some people. 2.The scenario will depict a scene in which law enforcement and people with "disturbed behaviour" play a key role. I understand this, and that this might induce psychological distress.	( )	( )
Use of the information in the study		
I understand that information I provide will be used for the analysis of the performance of the VR environment, as well as for corresponding improvements.	( )	( )
I understand that personal information collected about me that can identify me, such as [e.g. my name or where I live], will not be shared beyond the study team.	( )	( )

I agree that my information can be quoted in research outputs	( )	( )

Future use and reuse of the information by others

I give permission for the data concerning the VR performance and my actions within it that I provide to be archived in UT Cloud so it can be used for future research and learning.	( )	( )

UNIVERSITY OF TWENTE.

I agree that my information may be shared with other researchers for future research studies that may be similar to this study. The information shared with other researchers will not include any information that can directly identify me. Researchers will not contact me for additional permission to use this information.

( ) ( )

Signatures

X **15.06.2023**

_____	_____	_____
Name of participant [printed]	Signature	Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

[Name]
02.07.2023

_____	_____	_____
Researcher name [printed]	Signature	Date

Study contact details for further information:
[DELETED FOR PRIVACY]

UNIVERSITY OF TWENTE.

Appendix I

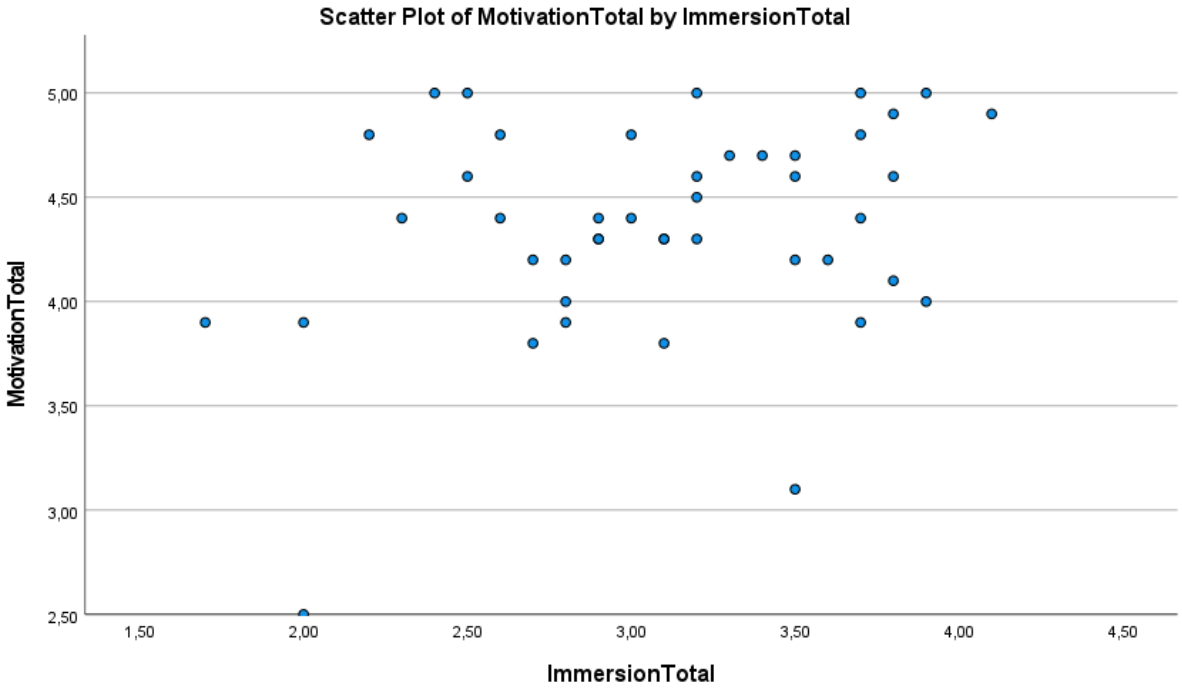
Guideline for the researcher to instruct participant before entering the VE

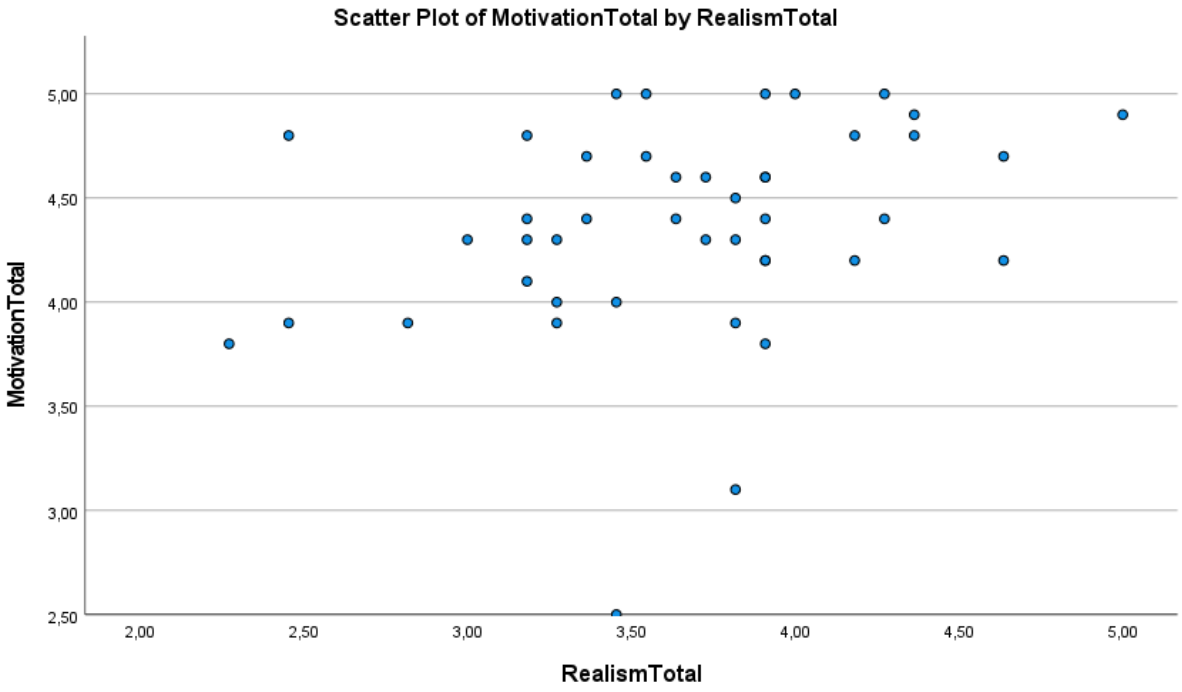
Guideline

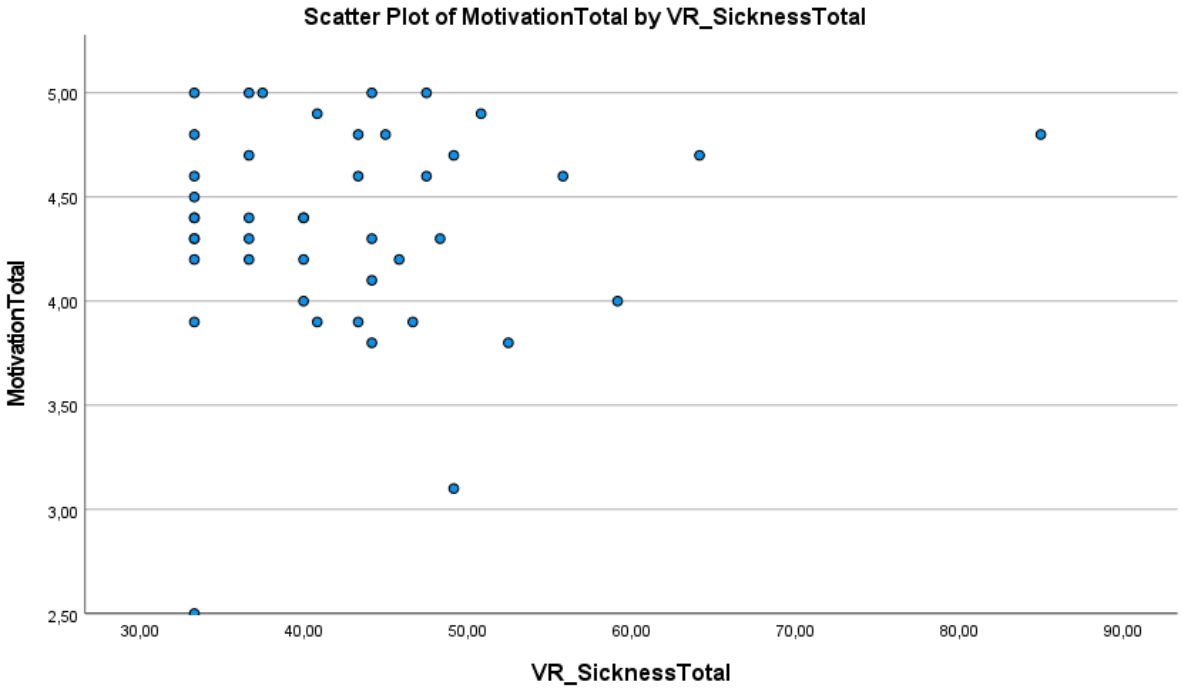
1. Did you ever use VR before?
    - ➔ Might get motion sick. If so: You can always remove glasses to pause or quit the study
2. Controls:
    a. Look around by turning head or use the stick (better against motion sickness)
    b. To move to a position, point there with your right controller, then click button to go there (teleport)
    c. You will have to choose answers in a conversation – do this by pointing to the answer and click the other trigger button on the back
3. Intro: First, you will be in a street where you can get accustomed to the feeling. A radio call will come in. After the call, you are free to approach the crossroads whenever you feel ready. As soon as you step on the crossroad, you will be teleported to actual location of the interaction
4. Upon arrival:
    a. Look around to find PiC
    b. Approach him to start the interaction
    c. You can change your positioning during the interaction as you like, in the same fashion as used before on the street
5. I will remain here to assure nothing unusual happens

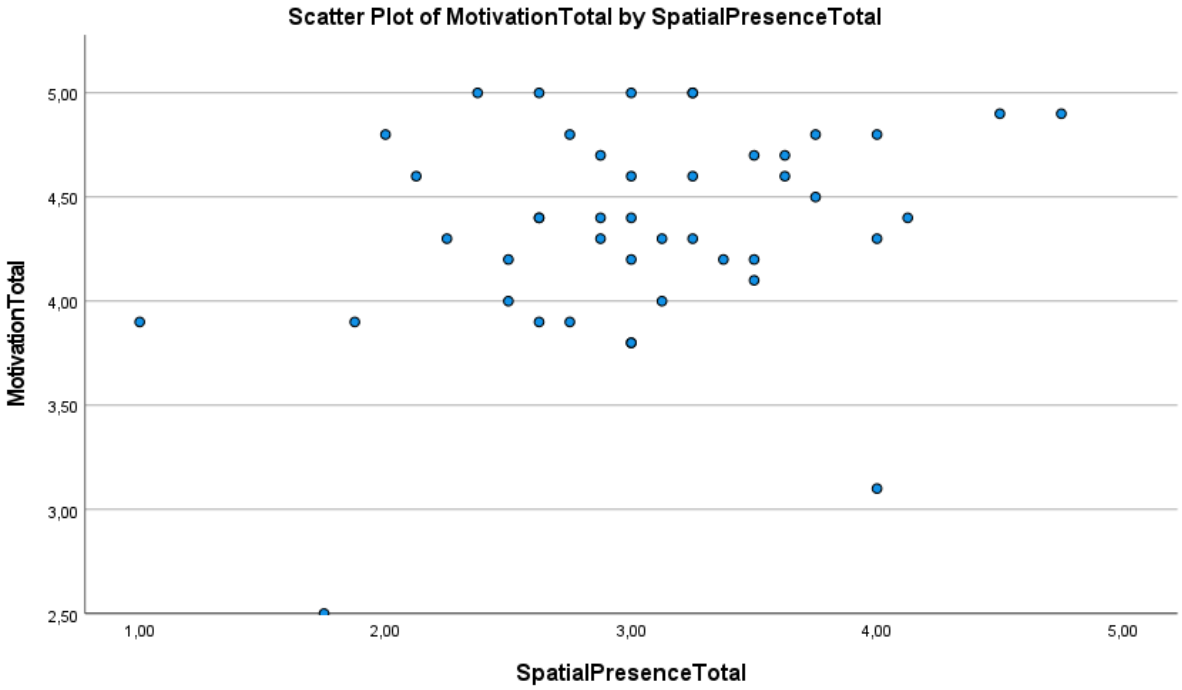
Appendix J

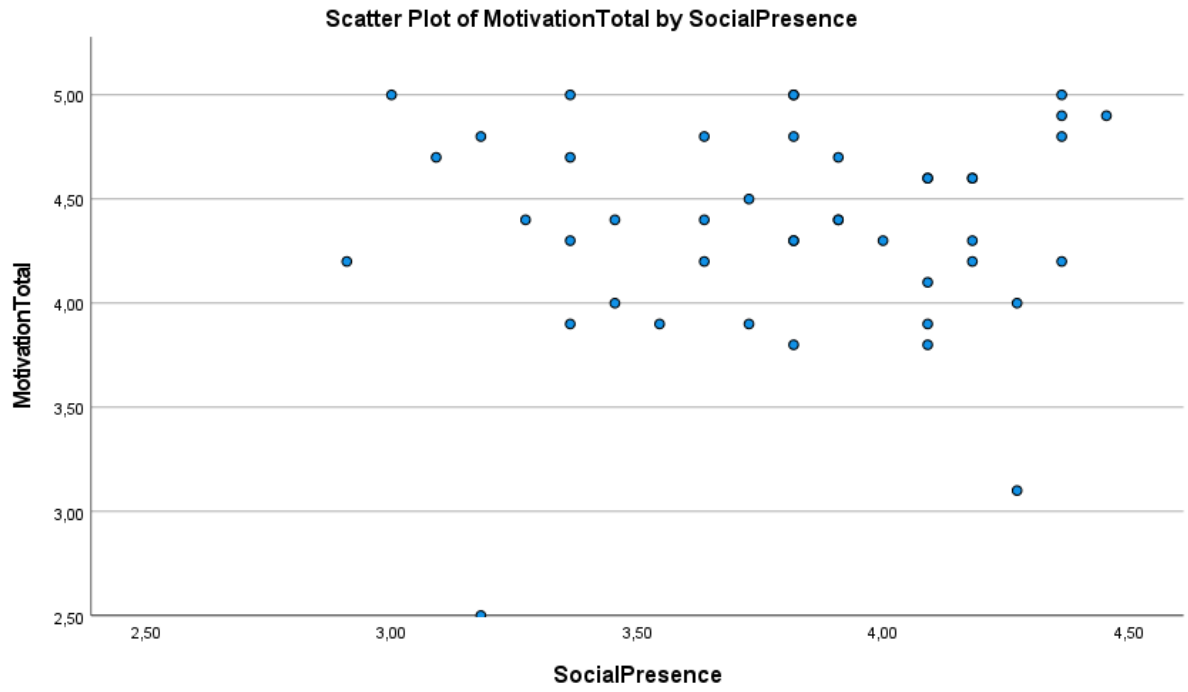
Diagnostic Tests Multiple Linear Regression






Scatter Plot of MotivationTotal by VR_SicknessTotal
MotivationTotal
5,00
4,50
4,00
3,50
3,00
2,50
30,00
40,00
50,00
60,00
70,00
80,00
90,00
VR_SicknessTotal


Scatter Plot of MotivationTotal by SpatialPresenceTotal
MotivationTotal
5,00
4,50
4,00
3,50
3,00
2,50
1,00
2,00
3,00
4,00
5,00
SpatialPresenceTotal



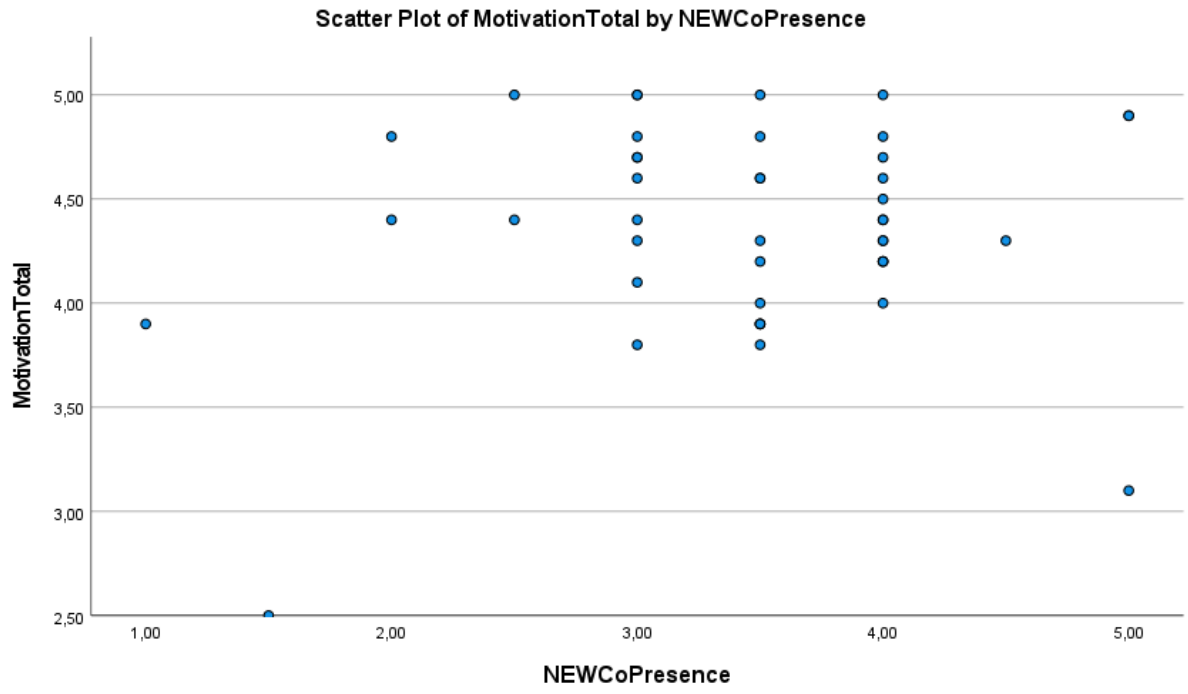


Figure J1. Linearity Plots

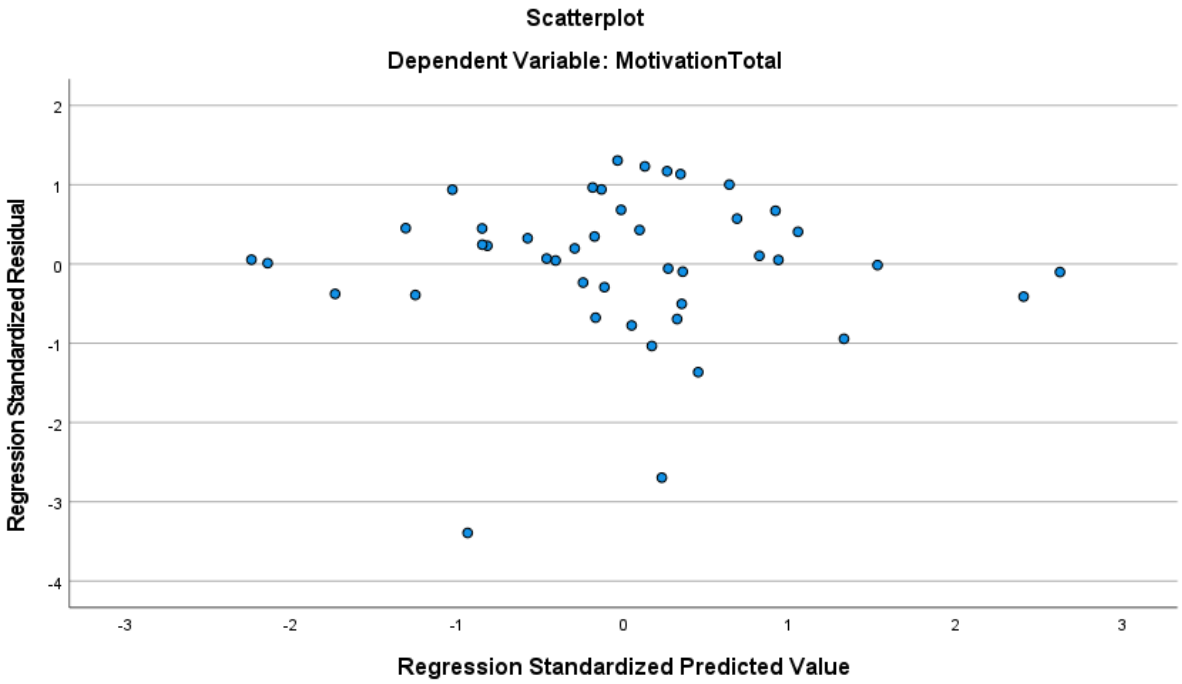


Figure J2. Homoscedasticity assessment

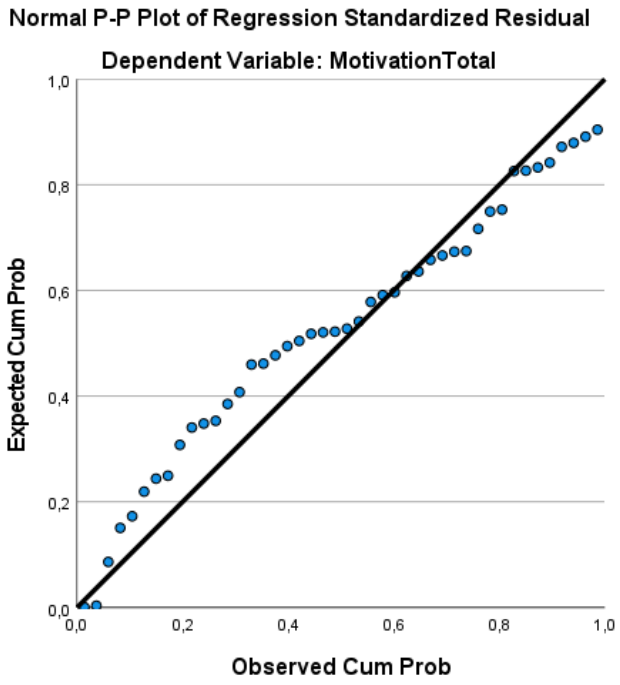


Figure J3. P-P Plot Normality of Residuals

Appendix K

Coding Overview.