“I Shot the Interviewer!”: The Effects of In-VR Interviews on Participant Feedback, Presence, and Rapport

Jacob Young
jacob.young@otago.ac.nz
School of Computing
University of Otago
Dunedin, New Zealand

Jennifer Ferreira
jennifer.ferreira@vuw.ac.nz
School of Engineering and Computer Science
Te Herenga Waka - Victoria University of Wellington
Wellington, New Zealand

Nadia Pantidi
nadia.pantidi@vuw.ac.nz
School of Design Innovation
Te Herenga Waka - Victoria University of Wellington
Wellington, New Zealand

Figure 1: The techniques for user interviews we have investigated in this research. (Left): A traditional face-to-face interview. The participant must take off their head-mounted display and talk with the interviewer in person. (Centre): The participant remains immersed within the virtual environment while interviewed. This allows the participant to remain in an environment they can still see, hear, and interact with, while keeping the interviewer hidden from them. (Right): The interview is conducted entirely within the virtual environment. The interviewer is represented as an abstract avatar which is controlled by them remotely using their own head-mounted display and controllers. This hides the identities of both parties throughout the interview. Note: All individuals in these images are not participants or paper authors.

ABSTRACT

The integration of questionnaires into virtual reality experiences has recently been proposed as a way to reduce the potential biases introduced through the negative effects of leaving VR, however there has been little attention paid to how qualitative interviews could similarly be integrated into the virtual world for the purposes of user evaluation. In this paper we explore how conducting interviews within the virtual environment may affect the outcome of the evaluation and the relationship between participant and interviewer, and how this may differ with and without visual representation of the interviewer through use of an avatar. We conclude that in-VR interviews are a valid and promising method of data collection for user evaluation with similar data quality to in-person interviews, but that the interviewer should have a visual presence in the environment to maintain their relationship with the participant and the perceived realism of the environment.

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642707

CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods**; **Virtual reality**.

KEYWORDS

Virtual/Augmented Reality, Interview, Qualitative Methods

ACM Reference Format:
Jacob Young, Jennifer Ferreira, and Nadia Pantidi. 2024. “I Shot the Interviewer!”: The Effects of In-VR Interviews on Participant Feedback, Presence, and Rapport. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3613904.3642707

1 INTRODUCTION

Advancements in interactive and immersive technologies have prompted researchers to rethink how usability and user evaluation studies can be conducted and how key data collection methods, such as questionnaires and interviews, can be optimised. For example, questionnaires that were once completed with pen and paper have been recreated digitally, sometimes resulting in higher quality data due to the extra layer of perceived anonymity by participants when there is no researcher present [4]. Equally, conducting interviews remotely via video-conferencing tools has become common practice spurred on by the Covid-19 pandemic.

This exploration is particularly alive in the field of *Virtual Reality* (VR) [2], in part due to a unique problem inherent in the evaluation

of VR technologies: participants have to take their VR headset off before they can complete a pen-and-paper questionnaire. In doing so, it has been reported that participants may experience a *break in presence* characterised by possible disorientation, surprise, and a lack of body ownership while they readjust to the real world [33], possibly affecting their perception of the virtual experience before it. By having participants complete the questionnaire within the virtual environment (referred to as *in-VR*) we can reduce or avoid this break in presence [21, 54], resulting in more consistent data [61] and a more enjoyable experience for participants [2].

Despite these benefits, there has been little research into how this in-VR approach can be adapted to another main component of usability anduser evaluation: user interviews. Interviews are a versatile and widely used technique across a range of research questions, methodologies, and study contexts — ranging from user interviews for evaluating interactive systems [56], to clinical interviews in psychological research [53], to analysing discourse in the field of applied linguistics [41], providing insight into the subjective experiences and perspectives of individuals. Subsequently, interviews are a regular part of user evaluations of virtual and augmented reality experiences [30, 33, 42, 57].

We are interested in exploring this further as it is possible there are several benefits in conducting user interviewswithin a virtual environment. Interviews as part of usability and user evaluations may be susceptible to the same breaks in presence that questionnaires are [61], and it may be possible that participants are still disoriented during the interviews. Conducting interviews within the virtual environment would allow participants to discuss things that they can still see and interact with rather than relying on their recollections of this experience, possibly leading to more reliable and detailed data. Finally, the integration of the interview process into the virtual environment would increase the possibilities and potential for remote user testing [55] and stealth assessment [64], providing a more robust study design.

In this paper we explore how conducting interviews as part of user evaluations, within and outside a virtual environment, may affect their sense of *spatial presence* within the virtual environment, their *rapport* and *psychological involvement* with the interviewer, and their willingness to critique the experience. This was explored through mock user evaluation where participants were interviewed about their experience playing a short VR shooting game either face-to-face, while the participant kept their VR headset on, or completely within the virtual world through the use of avatars. Participants were told that the study was seeking feedback on the shooting game for further development, however the game had been developed intentionally poorly in several areas and the study was actually about how their feedback and experience may change depending on how the interview was conducted. In addition to participants' feedback via the interview, we collected data using three questionnaires to ascertain their sense of spatial presence within the environment, their sense of rapport and psychological involvement with the interviewer, as well as further feedback on the game.

We found no significant difference in the spatial presence, rapport with the interviewer, or likelihood to praise or criticise the game across the three conditions. However, participants' psychological involvement with the interviewer, or empathy and mutual understanding with them, significantly increased when participants were able to see the interviewer in person or in the virtual world. This suggests that, subject to further investigations, in-VR qualitative interviews could be a promising method of usability and user evaluation, potentially on par with face to face interviews. We also found that those who established stronger rapport with the interviewer, or knew the interviewer beforehand and already had pre-established rapport with them, were more likely to give positive feedback about the game and report feeling present within it.

To our knowledge this is the first investigation into using immersive technologies for qualitative interviews, and so we conclude with a discussion and recommendations for how these interviews can be incorporated into the virtual environment in future research.

2 RELATED WORK

Interviews have several well known benefits as a research method: they allow access to rich, detailed accounts, permit exploration of issues that may be too complex to investigate through quantitative means, provide the ability to understand an individual's context and motivations, allow access to verbal and non-verbal cues, and provide opportunity for follow up and probing of participants' responses [35, 58].

There are several important factors that can determine the effectiveness of an interview. One that is paramount is the *rapport* between the participant and the interviewer, which is an engaging and harmonious connection that can exist between two people [22]. Rapport consists of three components: i) mutual attentiveness, or how interested each interactant is in the actions and words of the other; ii) positivity, or mutual friendliness; and iii) coordination, or how "in-sync" the two interactants are in their actions [69]. Two people experiencing rapport with one another will demonstrate psychological coordination, symmetry, and closeness that can facilitate free conversation [5] and thus make the interaction more enjoyable [19]. This closeness between interviewer and participant can lead to greater disclosure of sensitive topics [43], resulting in greater data quality and accuracy.

A related factor is the *psychological involvement* experienced between the participant and interviewer, which Biocca et al. [8] describe as the focal attention each interactant allocates to the other, manifesting as an increased awareness of the other's intentions and emotional states. Establishing this emotional connection can be crucial in building and sustaining rapport [73], and is also an important factor in making speakers feel as though they are sharing the same space [7].

2.1 Remote Interviews and the Effects of Visual Anonymity

Interviews are not always carried out in person. Some of the earliest examples of remote interviews are phone interviews, which have remained a popular research method as they enable geographically-dispersed populations to participate in research, potentially increasing the diversity of participants [37]. Phone interviews also foster greater uniformity across participants, are more cost-efficient, and facilitate faster data gathering than in-person studies [23, 65]. Another benefit is that separation between participant and interviewer

results in greater disclosure of sensitive information [9], and so phone interviews have been used in a wide variety of research topics where participants desire anonymity such as substance abuse [11], mental health [52], and reproductive health [24].

These benefits are also present when applied to computer-mediated experiments where anonymity is maintained, particularly when researching sensitive topics. In a series of studies comparing face-to-face to visually anonymous computer-mediated conversations, Joinson [28] found that participants disclosed significantly more sensitive information about themselves than those who spoke face-to-face or without visual anonymity, suggesting a heightened sense of self-awareness as the cause. Similarly, Greist et al. [18] found that patients are more likely to report suicidal tendencies to a computer than to a physician.

Despite these benefits, phone and computer-mediated interviews are often not opted in qualitative research as the inability for speakers to see each other could limit the development of rapport between them [46]. The use of videoconferencing solutions such as Skype or Zoom, has been suggested as possibly resulting in similar levels of rapport between participant and interviewer as in-person interviews [12, 27] while still facilitating the sharing of sensitive information [68]. Video interviews can also feel less formal or "daunting" than in-person visits [73], and thus are often preferred by shy or introverted participants [48, 63]. However, there are concerns that videoconferencing may not provide a sufficient environment for effective communication to occur, particularly when it comes to sharing visual bodily conversational cues as only the head and some of the upper body is visible [12, 25, 63].

2.2 Virtually-Mediated Interviews

The recent rise in popularity of virtual reality technologies has resulted in the development of new "telepresence" platforms that aim to make it feel as though speakers are physically co-located with one another. Rather than only capturing the user's head using a webcam, as with videoconferencing technologies like Zoom, users are represented as a virtual *avatar* that depicts the movements, and possibly appearance, of their entire body, greatly increasing the potential for non-verbal communication [17, 50].

In popular platforms such as VRChat [72] these avatars are player-created, meaning users can depict themselves however they want without regard to their actual physical appearance. Recent research has also focused on how to virtually recreate the actual appearance of users in such environments, either through pre-constructed 3D scans [3] or live capture of their entire body [49, 51], resulting in a greater sense of presence between speakers [3, 10, 74].

With these new technologies comes a new way in which we could consider conducting interviews, and indeed there has been growing interest in this space, though to our knowledge none so far have directly addressed qualitative data gathering for the purposes of user evaluation and usability.

2.2.1 ***Job Interviews.*** Adiani et al. [1] proposed the use of a virtual reality simulator for preparing autistic people for job interviews. The interviewer in this case used a rudimentary AI and text-to-speech but could respond to the interviewer's mood through physiological sensors. They conclude that the virtual interview was sufficient to prepare interviewees for a real one, though some participants expressed feeling discomfort due to the robotic nature of the interviewer [1].

Kwon et al. [36] similarly explored the use of virtual reality to simulate job interviews, and also used an AI-controlled interviewer for this purpose. They found that a virtual job interview could induce similar levels of anxiety in participants as a face-to-face one, particularly when the interviewer was depicted realistically, and that a VR interview was significantly more anxiety-inducing than one displayed on a laptop.

Noguchi et al. [45] explored how avatars could be introduced into face-to-face job interviews through the use of a virtual display over the interviewer's face. In a comparison of three different interviewers, all of which induced differing levels of nervousness in participants due to their appearance, it was found that obscuring the interviewer's face with either a "diligant" or "gentle" avatar face could reduce the variance in participant nervousness between them.

2.2.2 ***Clinical Interviews.*** Another popular application of virtual interviews has been for clinical psychology. An example of this is the SimSensei Kiosk system [13], which allows users to talk to a virtual human (either human or AI-controlled) for healthcare decision support. It was found that participants experienced similar rapport with an AI-controlled avatar as they did with a human interviewer, and even experienced more rapport with the avatar when it was human-controlled. Devault et al. [13] suggest that this may be because participants felt more comfortable revealing sensitive information to the avatar than they would with a human interviewer.

This increased willingness of participants to disclose sensitive information to virtual humans was confirmed in a follow-up study by Lucas et al. [38]. Using a similar human avatar to SimSensei Kiosk, participants were asked a series of questions about clinical symptoms, and were told that the avatar was speaking to was either human- or AI-controlled. It was found that those who believed they were talking to a computer were not only significantly more likely to disclose sensitive information about themselves, but would also visibly display more emotions such as sadness during the interview.

In a further follow-up study by Lucas et al. [39], this system was used to interview military service members about potential post-traumatic stress disorder symptoms. Participants who spoke to the AI-controlled avatar were significantly more likely to report symptoms than they would in even an anonymised questionnaire, which the authors link to the rapport developed between the participant and the virtual interviewer.

2.3 Completing Questionnaires in Virtual Environments

Though in-VR interviews for user evaluation and usability purposes, have yet to be investigated, there has been increasing interest in conducting questionnaires within virtual environments in recent years. Questionnaires have long been a popular method for VR usability testing, with popular options such as the iGroup Presence Questionnaire [59] and the Simulator Sickness Questionnaire [32] having been used in several thousand experiments to date. However, the traditional pen-and-paper approach faces a unique problem when

applied to virtual reality research: participants have to take off their head-mounted display before they can complete the questionnaire, and in doing so are subject to a *break in presence* when transitioning from the virtual world to the real one. This break in presence can result in disorientation and reduced body ownership and control [33], potentially resulting in less consistent self-reported data [61].

One method that has been proposed to reduce this break in presence between the experience and its evaluation is to integrate the questionnaire into the virtual environment [2]. Rather than use pen and paper, the questionnaire appears as a virtual element that can be interacted with through the use of controllers [20], hand gestures [60], or other virtual elements [21]. Schwind et al. [61] found that this integration can reduce the variation in self-reported presence measures by reducing the disorientation and confusion felt by participants after the experiment, and Putze et al. [54] found that in-VR questionnaires can reduce the break in presence associated with leaving the virtual environment without affecting the self-reported player experience.

Integrating the questionnaire into the virtual environment can be taken one step further by actually making it part of the experience [55], allowing for stealth assessment of the experience or the user's achievement [64]. For example, Frommel et al. [16] integrated a questionnaire into a driving game by having participants drive into their desired responses, and saw higher rates of reported enjoyment and virtual presence when the game was evaluated this way. Similarly, Gründling et al. [21] tasked participants with evaluating an archery game by shooting their desired responses with a virtual bow, and found that this method of answering may reduce the negative effects of transitioning between the virtual and real worlds without impairing the questionnaire results.

2.4 Addressing the Research Gap

Though we have seen increasing research on how to best conduct in-VR questionnaires [2, 21] for user evaluation and usability purposes, we have yet to see research addressing how qualitative interviews can likewise be conducted from within the virtual environment.

Existing research on in-VR questionnaires showed that allowing the participant to keep their VR headset on during an interview would reduce the break in presence associated with leaving virtual reality [21, 54], while increasing self-reported presence [16] and reducing the variability of results [61]. Research on phone- and computer-mediated interviews has shown that introducing a degree of separation and the perceived sense of anonymity can allow participants to be more forthcoming in their feedback (and criticism) [9, 18, 28, 38]. On the other hand, it can limit rapport and psychological involvement with the interviewer [8, 12, 27, 46], which in turn can affect disclosure.

Our work aims to address the research gap of conducting in-VR qualitative interviews for purposes of user evaluation and usability. In particular, based on the review of existing work, we are interested in investigating the effects in-VR interviews may have on spatial presence, rapport and psychological involvement between participants and the interviewer, and participants' likelihood of praising or criticising the game.

3 METHOD

We carried out a study to determine what effects the manner in which an interview is conducted will have on participants' sense of spatial presence, their rapport and psychological involvement with the interviewer, and the nature of the feedback they provide. It was crucial that participants were unaware of the study's true purpose so as not to influence their results; participants were thus told that the study was seeking feedback on a virtual reality shooting game, shown in Figure 2 and described in section 3.1. As part of this, we asked participants to play the VR game and collected their feedback through a mock interview which varied in the way it was presented, as described in subsection 3.3: either face-to-face, with the participant still immersed in the virtual environment, or with both participant and interviewer in the virtual environment. After the interview, participants also filled a series of questionnaires (see 3.5.5).

This research was approved by [name of university ethics board redacted to maintain anonymity].

3.1 Virtual Reality Game

In the virtual reality game players were tasked with fending off increasingly large waves of aliens using a light blaster; this lasted for 70 seconds and was impossible to fail or complete before this time, ensuring similar exposure times for all participants before the post-game interview.

The game was purposefully developed to contradict what is considered 'good' game design in several areas:

- **Graphics:**
  - The aliens were incorrectly made semi-transparent so that various body parts visibly clipped through each other.
  - The gun and enemy meshes were much higher fidelity than the environment, causing visual inconsistency.
  - The user interface contained several misspelled words.
- **Audio:**
  - The background music was a particularly loud "dubstep" track, chosen to be repetitive and grating.
  - The light blaster only had one sound effect that played identically every time the gun was fired, quickly becoming repetitive as the game required rapid fire.
  - The enemies didn't make any sounds other than their blaster fire.
- **Gameplay:**
  - The scoring system was inconsistent; points awarded per kill were semi-randomised, headshots were randomly given without regard to hit registration, and points would be randomly withdrawn for imaginary infractions.
  - The aliens were programmed to always miss the player with their return fire.
  - Collision detection on the player's bullets had a low update rate so that they would occasionally go through enemies.

These deficiencies were purposefully included to ensure that the participant would have negative aspects to discuss in the post-game questionnaires and interview. Still, consideration was given to make the game enjoyable and aesthetically pleasing to, likewise ensure there would be positive aspects of the experience to discuss; for

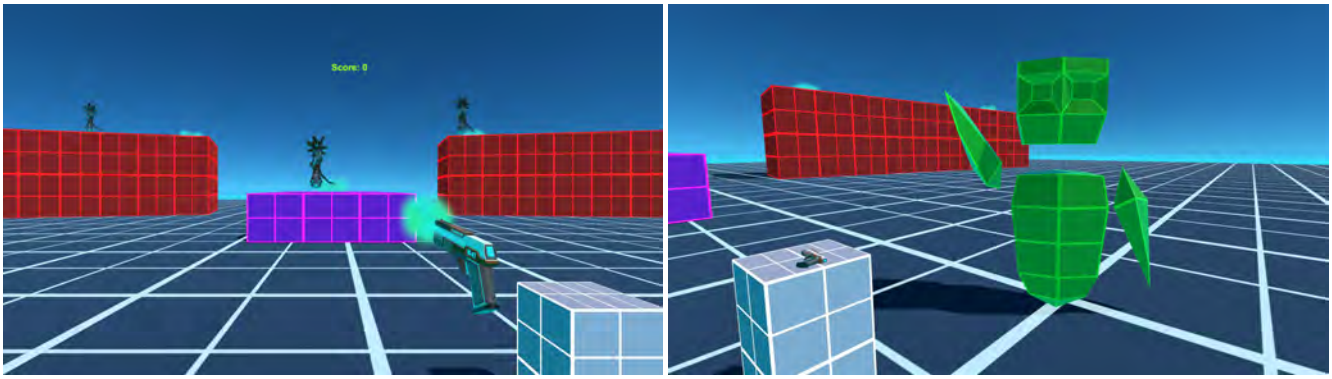

Figure 2: The virtual reality game participants played as part of our study. Using a light blaster, players had to fend off increasingly large waves of alien creatures. (Right): The avatar that would appear to interview participants after they had completed the game. This was designed to be visually coherent with the game environment.

example, much work was put into the look and feel of the gun to make it satisfying to fire.

3.2 Study Procedure

Each study session followed the phases described below .

(1) **Onboarding:** Participants were first given an information sheet explaining that the purpose of the study was to test a short VR game we had developed and provide feedback to guide future development. As mentioned earlier, our study involved incomplete disclosure (a form of deception) as we withheld information about the actual study purpose in order to prevent biasing the results. Participants were then asked to complete a short demographics questionnaire.
(2) **Demographic Questionnaire:** Participants completed a short questionnaire asking for their demographic information, including their susceptibility to simulator sickness; participants with a prior history of adverse effects were excluded from further testing.
(3) **VR Game:** After a brief explanation of the controls, participants played the short VR game described in subsection 3.1 where they were asked to fend off an increasingly large number of alien creatures. The game lasted 70 seconds and was impossible to fail or complete before this time, ensuring a consistent exposure time for all participants. Participants were told before the game started that they would be interviewed afterwards and how this interview would be conducted.
(4) **Post-Game Interview:** A short interview was then conducted, which we henceforth refer to as the *post-game interview*, to ask about participants' thoughts on the game. The manner in which each interview was conducted following one of the three conditions described in subsection 3.3, with one condition randomly assigned per participant and an equal number of participants per condition. Participants were informed when the interview was about to begin and to either take their headset off or leave it on depending on the study condition. All interviews were conducted in English.
(5) **Post-Game Questionnaires:** After the interview was completed, participants were then asked to complete a questionnaire to gather quantitative feedback about their experience with the game and during the interview. To ensure consistency between the three conditions this was administered on a standard desktop display, requiring all participants to remove themselves from the virtual environment before completing it.
(6) **Debrief:** After completing the questionnaire, but before responses were submitted, participants were verbally debriefed on the true nature of the task: that we were interested in the nature of their responses rather than looking for feedback on the game. Participants were given the option to withdraw from the study at this point if they were unhappy with this deception or the responses they gave, though none chose to do so.

3.3 Conditions

The core part of our study was a post-game interview conducted after participants had played the game. We were interested in how conducting the interview from within the virtual environment may affect participants' sense of spatial presence, their ability to establish rapport and psychological involvement with the interviewer and their feedback on the game i.e. willingness to provide criticism or praise.

We varied the manner in which the post-game interview was conducted in three different conditions. These were as follows:

- **The *Face-to-Face* Condition:** The post-game interview was conducted in person with a human interviewer.
- **The *Immersed* Condition:** The participant remained immersed within the virtual environment throughout the interview. The interviewer was physically present and speaking to the participant normally, however there was no representation of the interviewer within the virtual environment.
- **The *Avatar* Condition:** The interview was conducted completely within the virtual environment. The interviewer wore their own HMD and were shown as a character within the game, pictured in Figure 2, which had an articulated head

and arms to allow for basic body gestures. The participant was likewise represented by their own avatar. The participant and interviewer communicated through integrated voice chat using the microphones and speakers embedded in their respective head-mounted displays.

The study utilised a between-subjects design and so each participant was only exposed to one condition, and thus only completed one interview. The same interviewer was used for all participants across all conditions, while another member of the research team (the 'facilitator') greeted participants in person and guided them through all phases of the study session.

In all conditions the interview was the participants' first encounter with the interviewer in the session. To ensure the interviewer was not seen outside the VR environment, the interviewer entered the room after the participant put on their head-mounted display and left as soon as the interview had concluded in the *Immersed* condition, or was located in a separate room for the *Avatar* condition.

3.4 Research Questions

In conducting this research we aimed to answer the following questions:

- **(RQ1):** Would conducting interviews within the virtual environment, as in the *Immersed* and *Avatar* conditions, affect participants' sense of *spatial presence*?

Based on existing research on conducting in-VR questionnaires we expected that having the participant keep their VR headset on during the interview, as in the *Immersed* and *Avatar* conditions, would reduce or remove the break in presence associated with leaving virtual reality [21, 54] while increasing self-reported presence [16] and reducing the variability of results [61].

- **(RQ2):** Would visually obscuring the interviewer from the participant in the *Immersed* and *Avatar* conditions affect the participant's ability to establish *rapport*?

By having the participant keep their VR headset on during the interview, their view of the interviewer is obscured, making conversation similar to a phone- or computer-mediated interview. As with these mediated interviews, we thus expected that participants in the *Immersed* condition would have a limited ability to develop rapport with the interviewer [12, 27, 46]. We also expected that the *Avatar* condition would alleviate this loss of rapport by reintroducing a visual representation of the interviewer, resulting in similar levels of rapport as that of a face-to-face conversation [13].

- **(RQ3):** Would visually obscuring the interviewer from the participant in the *Immersed* and *Avatar* conditions affect the participant's ability to establish *psychological involvement*?

As with rapport, we expected that it would be difficult for the participant and interviewer to gauge each other's emotional and attentional state in the *Immersed* and *Avatar* conditions when they cannot see each other [8], resulting in reduced psychological involvement in these conditions, though this may be mitigated in the *Avatar* condition which reintroduces several non-verbal cues such as gestures [38].

- **(RQ4):** Would visually obscuring the interviewer from the participant in the *Immersed* and *Avatar* conditions result in the participant providing more *criticism* about the game than if the interview was conducted face-to-face?

Prior research has shown that participants are more forthcoming about sensitive issues when interviews are conducted via phone or a computer due to an increased sense of anonymity [9, 18, 28, 38]. Based on this, we expected that participants will be more forthcoming with criticism about the game in the *Immersed* and *Avatar* conditions than in the *Face-to-Face* condition.

3.5 Measures

This research used a mixed-methods design, utilising both quantitative and qualitative data gathering methods to obtain a detailed understanding of participants' overall experience and in particular with regards to the interview process and its impact.

3.5.1 ***Demographic Information.*** Participants were asked for their age, gender, prior experience with video games (Weekly / Monthly / Yearly / Rarely / Never) and prior experience with virtual reality (Weekly / Monthly / Yearly / Rarely / Never).

The demographic questionnaire also asked participants if they had prior history of simulator sickness or significant motion sickness; if so, they would have been excluded from further testing, though this was found not to be necessary.

3.5.2 ***Post-Game Interview.*** Each interview was recorded and transcribed, and was also timed to determine if their length differed between conditions.

We used a semi-structured approach where the interviewer was free to pursue additional lines of inquiry, but always followed an interview script with a set of questions that were asked to all participants:

(1) How was your overall experience playing the game?
(2) What did you think of the gameplay?
(3) What did you think of the graphics?
(4) What did you think of the audio?
(5) Do you have anything additional to add?

The interviewer's demeanour was polite and informal but to the point. They always introduced themselves at the beginning of each interview and thanked the participants for playing the game and offering their feedback. At the end of each interview, they always asked if there was anything else the participants wanted to add and then thanked the participants for their time.

3.5.3 ***Sentiment Analysis.*** SentiStrength [66] was used to analyse the sentiment expressed in the open feedback on the game within the questionnaire, as well as individual statements in the interview transcript. No training or adjustments were made to the SentiStrength algorithm or corpus. Accordingly, two sentiment strengths were reported on a scale of 1 (not negative) to -5 (extremely negative), and 1 (not positive) to 5 (extremely positive). A strength of [-1, 1] was considered neutral.

3.5.4 ***Praise and Criticism.*** Phrases from the participant responses were coded either as 'praise', 'criticism' or 'neutral' with respect to the feedback they were providing on the game. While sentiment analysis provided insights about the mood or emotional tone of the interviews and open questions on the questionnaire,

analysing praise and criticism accounted for the specialised terminology of VR technology and VR games used by participants, which is not accurately analysed with off-the-shelf sentiment analysis software such as SentiStrength. Sentiment analysis software must be specifically designed and trained to identify praise and criticism, requiring additional semantic rules and a curated lexicon drawn from the relevant domain (e.g., [67]).

If a participant described an aspect of the game that they 'liked', 'enjoyed', or mentioned that something was 'good', 'fun', or 'cool' we coded that phrase as praise, for example: "That was fun," or "... I like that there's not too much movement involved". If a participant described feeling uncertain about a feature in the game, or that something was too taxing, or mentioned something that they did not like, we coded that phrase as criticism. For example: "my finger started hurting," or "you can't tell whether they're dying."

There were many instances of participants making suggestions about alternative designs or extensions to the existing game design. These were coded as 'neutral', unless the statement was very clearly addressing what was seen as a shortcoming in the game, for example, "it's better to have the two guns to shoot in different ways, so it's more kind of active" was coded as criticism, whereas "two guns or something would be quite nice," was not. To ensure reliability and mitigate researcher bias each author independently coded randomly assigned interviews for praise and criticism feedback. Points of difference between the researchers' codes were discussed to enhance consistency.

Praise and criticism comments were then tallied for each interview. If an interview consisted of an equal proportion of praise and criticism comments then the interview was considered 'balanced'. If an interview consisted of more than 50% praise (or criticism) then the interview was considered complimentary (or critical), with 70% or more praise (or criticism) comments considered strongly complimentary (or critical) respectively.

3.5.5 ***Post-Game Questionnaires.*** The post-game questionnaires consisted of the iGroup Presence Questionnaire (IPQ) [59] to evaluate participants' sense of spatial presence (or sense of "being there") within the virtual environment; the Networked Minds Measure [8] to measure their psychological involvement (empathy and mutual understanding) with the interviewer; and a questionnaire by Hall and Bernieri [22] to measure the participants' rapport with the interviewer.

Participants also rated the game's graphics, audio, and gameplay to determine if their impression of each was consistent with what they told the interviewer. Finally, participants were asked to rate how well they knew the interviewer before this study ("Not at all"-"Very well").

All questions used a seven-point Likert scale. The full set of questionnaires is available in the supplemental material.

3.5.6 ***Interviewer Reflection.*** The interviewer took note of their own experiences in terms of rapport, benefits and challenges and overall process of conducting the interviews, both in VR and outside VR. Their reflections on the interview process have been provided in subsection 4.4.

3.6 Data Analysis

The responses to the post-game questionnaires and interviews were analysed quantitatively and qualitatively. All 45 post-game interviews were transcribed and analysed for sentiment and praise and criticism.

3.6.1 ***Quantitative Analysis.*** The data from the spatial presence, rapport and psychological involvement questionnaires were statistically analysed. As the Likert scales in our questionnaires represent ordinal data we used non-parametric Kruskal-Wallis tests in their analysis ($N = 45$, $\alpha = 0.05$). One-way ANOVA was used for other normally-distributed data, or Kruskal-Wallis tests for non-normal data, as determined by a Shapiro-Wilkes test. All reported p values have been adjusted using Bonferroni correction.

3.6.2 ***Qualitative Analysis.*** The statements made during the the interviews and the open questions in the questionnaires were coded and grouped deductively according to the categories of interest based on our research questions [6]. Our focus was on identifying patterns in participants' spontaneous comments about their first impressions of being interviewed in VR, how they reacted to the avatar, and how they reacted to remaining in VR while the interviewer was not represented in the VR environment. To complement the quantitative view (constructed through the statistical analysis, sentiment analysis, and praise and criticism analysis) the thematic analysis helps to frame some of the quantitative results and provides further context for their interpretation.

4 FINDINGS

In this section we present our quantitative results (section 4.2), qualitative results (section 4.3), and interviewer reflection on the study (section 4.4).

4.1 Participant Demographics

45 participants were recruited between the ages of 19 and 57 (M=31.29, SD=9.38), with 15 experiencing each condition. 31 were male, 10 were female, three were non-binary/gender diverse, and one preferred not to disclose their gender. Participants identified as European (26), Asian (14), Pacific (2), Latin American (1), Middle Eastern (1), or Other (3), with a further three preferring not to disclose their ethnicity.

The majority of participants (29) didn't know the interviewer at all before the study. 15 knew them well (rated between 5 and 7), with three of these knowing the interviewer "very well" (rated 7). These participants were evenly distributed between conditions (five per condition).

Almost all participants had prior experience with video games (43), with 20 playing them at least once per week. Similarly, most had tried virtual reality in the past (38), though only five used it on a weekly basis.

See Appendix A for detailed participant demographic information.

4.2 Quantitative Results

This section presents the quantitative results of the post-game questionnaires (sections 4.2.1– 4.2.4) and interviews (sections 4.2.5 and 4.2.6). Interviews with participants took four minutes and one

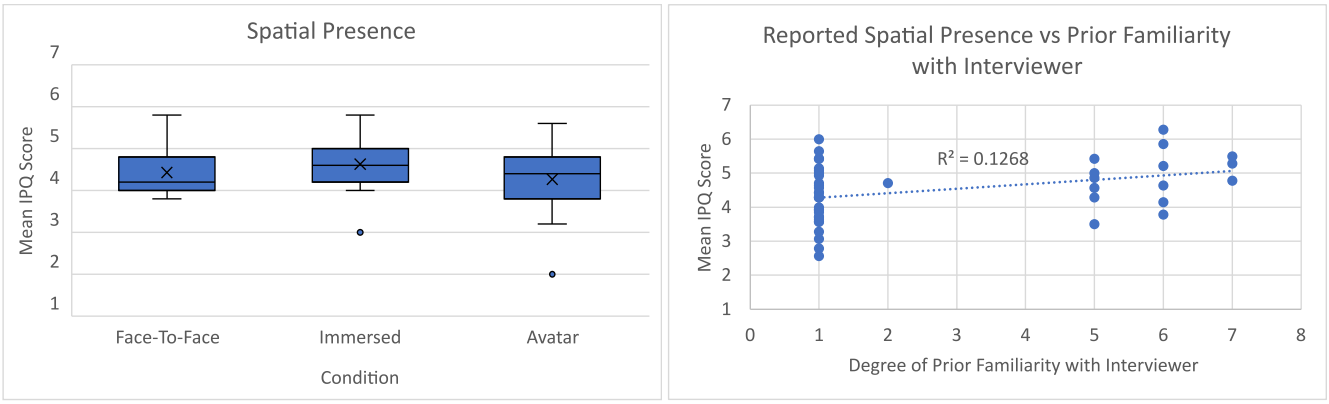

Figure 3: (Left): The spatial presence induced in each condition as represented by mean IPQ scores. No significant different in induced presence was found between conditions. (Right): The relationship between participants' prior familiarity with the interviewer and their reported spatial presence. This relationship was found to be statistically significant.

second on average (SD=1m30s), with one-way ANOVA finding no significant difference in interview length between conditions ($F(2, 42) = 0.88$, $p = 0.42$).

4.2.1 ***Spatial Presence.*** Kruskal-Wallis tests found no significant difference in the overall IPQ score per condition ($\chi^2 = 2.81$, $p = 0.25$), nor for the Presence ($\chi^2 = 0.59$, $p = 0.74$), Spatial Presence ($\chi^2 = 1.81$, $p = 0.40$), or Involvement subscales ($\chi^2 = 0.93$, $p = 0.63$).

A significant difference was found for the rated Realness between conditions ($\chi^2 = 7.19$, $p = 0.03$), with Dunn's test confirming that Realness scored significantly higher in the *Face-to-Face* condition than the *Immersed* condition ($p = 0.04$). No significant difference was found between the *Face-to-Face* and *Avatar* conditions ($p = 1.00$) or between the *Immersed* and *Avatar* conditions ($p = 0.10$), though a difference was found in the latter comparison before bonferroni correction ($p = 0.03$).

Simple linear regression found that those more familiar with the interviewer before the study were more likely to report higher spatial presence within the environment, with the mode IPQScore = 4.14 * (PriorFamiliarity) found to be statistically significant ($R^2 = 0.11$, $F(1, 43) = 6.27$, $\beta = 0.13$, $p = 0.02$).

4.2.2 ***Rapport.*** A Kruskal-Wallis test found no significant difference in rapport with the interviewer between conditions ($\chi^2 = 2.09$, $p = 0.35$). Further tests revealed that the only rapport factor in which a significant difference was present was how "Friendly" participants found the interviewer ($\chi^2 = 7.17$, $p = 0.03$), with Dunn's test showing that those in the *Avatar* condition found the interviewer significantly friendlier than those in the *Immersed* condition ($p = 0.02$).

Simple linear regression found a relationship between participants' rapport with the interviewer and how well they knew the interviewer prior to the study, with the model Rapport=5.27 * (PriorFamiliarity) being statistically significant ($R^2 = 0.09$, $F(1, 43) = 5.25$, $\beta = 0.14$, $p = 0.03$).

4.2.3 ***Psychological Involvement.*** A Kruskal-Wallis test found a significant difference in the reported Psychological Involvement between conditions ($\chi^2 = 9.61$, $p = 0.01$). Dunn's test confirmed that psychological involvement was scored significantly higher in the *Face-to-Face* condition than the *Immersed* condition ($p = 0.03$) and significantly higher in the *Avatar* condition than the *Immersed* condition ($p = 0.02$). No significant difference in psychological involvement was found between the *Face-to-Face* and *Avatar* conditions ($p = 1.00$).

Analysing this difference in psychological involvement further, another Kruskal-Wallis test found a significant difference in the empathy induced between conditions ($\chi^2 = 8.08$, $p = 0.02$), with Dunn's test finding a significant difference between the *Immersed* and *Avatar* conditions ($p = 0.02$) but not between the *Face-to-Face* and *Avatar* conditions ($p = 1.00$) or between the *Face-to-Face* and *Immersed* conditions ($p = 0.10$).

Further Kruskal-Wallis tests failed to find any significant difference in how empathetic the participant felt towards the interviewer between conditions ($\chi^2 = 14.56$, $p = 0.10$), however a significant difference was found in how empathetic the interviewer was perceived to be towards the participant between conditions ($\chi^2 = 12.48$, $p < 0.01$); Dunn's test confirmed a significant difference in perceived interviewer empathy between the *Face-to-Face* and *Immersed* conditions ($p < 0.01$) and between the *Immersed* and *Avatar* conditions ($p = 0.01$), but not between the *Face-to-Face* and *Avatar* conditions ($p = 1.00$).

Simple linear regression failed to find a significant relationship between psychological involvement and participants' prior familiarity with the interviewer ($R^2 = 0.01$, $p = 0.21$).

4.2.4 ***Game Rating.*** Participants were asked to score the game's graphics, audio, and gameplay on a 7-point Likert scale (1="very bad", 7="very good"). Kruskal-Wallis tests found no significant difference in these ratings between conditions for the graphics ($\chi^2 = 2.17$, $p = 0.34$), the audio ($\chi^2 = 1.00$, $p = 0.61$), the gameplay ($\chi^2 = 2.57$, p=0.28), or for the mean of all three ratings ($\chi^2 = 1.39$, $p = 0.50$). T

Simple linear regression found that participants' prior familiarity with the interviewer correlated positively with these scores, with the model OverallScore = 3.81 * (PriorFamiliarity) being found to

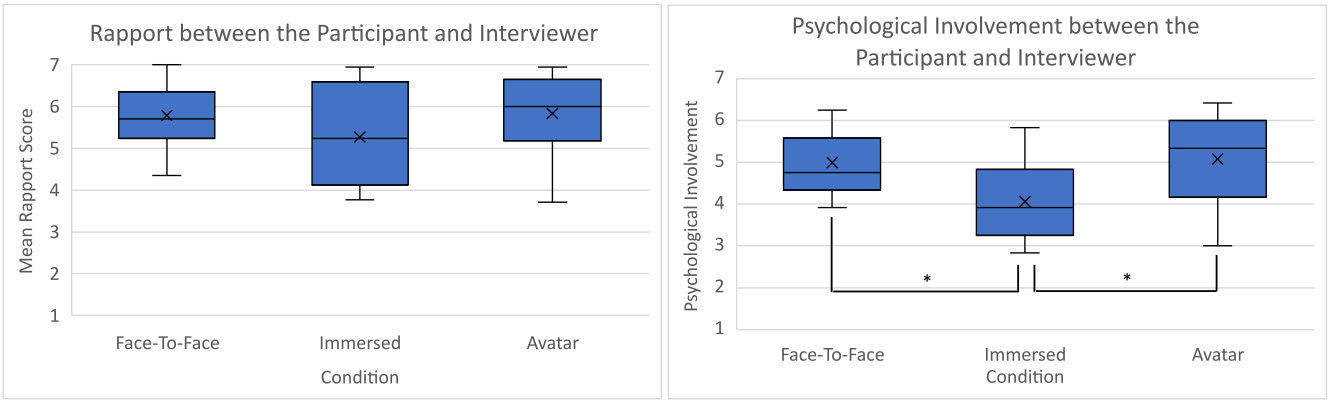

Figure 4: The participant's perception of their rapport (left) and psychological involvement (right) with the interviewer per condition. No significant differences in rapport were found, but the reported psychological involvement was significantly higher in the *Face-to-Face* and *Avatar* conditions than in the *Immersed* condition.

be statistically significant ($R^2 = 0.14$, $F(1, 43) = 8.24$, $\beta = 0.24$, $p < 0.01$).

After running SentiStrength [66] on the responses to the open questions, Kruskal-Wallis tests found no significant difference between conditions for either positive sentiment ($\chi^2 = 1.98$, $p = 0.37$) or negative sentiment ($\chi^2 = 0.05$, $p = 0.98$).

Simple linear regression found that as with the numerical rating, the relationship between participants' prior familiarity with the interviewer and the positive sentiment expressed in their comments was found to be statistically significant ($R^2 = 0.14$, $F(1, 43) = 8.24$, $\beta = 0.23$, $p = 0.006$), with those familiar with the interviewer more likely to express positive sentiment about the game.

Simple linear regression also found that a participant's overall rating of the game was significantly affected by the mean IPQ score ($R^2 = 0.43$, $F(1, 43) = 32.94$, $\beta = 0.1.05$, $p = 8.74e^{-7}$), the participant's rapport with the interviewer ($R^2 - 0.32$, $F(1, 43) = 20.51$, $\beta = 0.80$, $p = 4.65e^{-5}$), and the participant's psychological involvement with the interviewer ($R^2 = 0.19$, $F(1, 43) = 11.00$, $\beta = 0.60$, $p = 0.002$).

4.2.5 ***Post-Game Interview Praise and Criticism.*** Participants provided 9.93 pieces of non-neutral feedback per interview on average (SD = 4.46), of which 47% were praise and 53% criticism on average (SD = 26%). Kruskal-Wallis tests found no significant difference in how much feedback was provided per condition ($p = 0.83$), nor how much of this was praise ($p = 0.28$) and how much was criticism ($p = 0.23$). A Chi-Squared test also found no association between condition and the categorisation of interviews based on praise and criticism ($p = 0.62$).

Simple linear regression found no relationship between participants' prior familiarity with the interviewer and how much feedback provided ($p = 0.19$), how much of this feedback was praise ($p = 0.08$), or how much of this feedback was criticism ($p = 0.08$). However, a significant positive relationship was found between rapport and how much feedback was praise ($R^2 = 0.18$, $F(1, 43) = 10.83$, $\beta = 1.69$, $p = 0.002$).

The *Immersed* condition provided the highest number of criticism comments (n=82), followed by the *Avatar* (n=78) and *Face-to-Face* conditions (n=74). This ordering was consistent with the proportion of interviews skewed towards criticism in each condition: *Immersed* (10/15), *Avatar* (8/15), and FTF (7/15). Of the interviews skewed towards criticism in the *Immersed* condition 50% (5/10) were strongly skewed, compared to 62.5% (5/8) in *Avatar* and 57.1% (4/7) in FTF. While these results may not be statistically significant, the pattern observed does suggest that the *Immersed* and *Avatar* conditions invited more criticism than the *Face-to-Face* condition.

4.2.6 ***Post-Game Interview Sentiment.*** Interviews were positive overall in general, with a mean positive sentiment of 1.92 (SD=0.31) and a mean negative sentiment of -1.56 (SD=0.22).

Kruskal-Wallis tests found no significant difference in the overall sentiment expressed between conditions (pos:($\chi^2 = 1.57$, $p = 0.46$), neg:($\chi^2 = 0.04$, $p = 0.98$)). A significant difference was found in the positive sentiment expressed by the interviewer between conditions ($\chi^2 = 9.20$, $p = 0.01$), with Dunn's test confirming that the interviewer's sentiment was significantly more positive in the *Avatar* condition than the *Face-to-Face* condition ($p < 0.01$). No significant difference in positive sentiment was found between the *Avatar* and *Immersed* conditions ($p = 0.19$) or between the *Face-to-Face* and *Immersed* conditions ($p = 0.75$).

4.3 Qualitative Results

Thematic analysis revealed several trends in participant responses across the interviews and questionnaires.

First, the wealth of criticism we received about the game confirmed that there was sufficient content for them to discuss during the interview:

> *"I... wouldn't... uh... choose to play this because... yeah in general. Yeah"* - P15 Interview

However, it does appear that some participants felt social pressure not to criticise the game too much:

> *"I don't really wanna... badmouth the game"* - P10 Interview
> *"I don't know if I wanna give you advice"* - P38 Interview

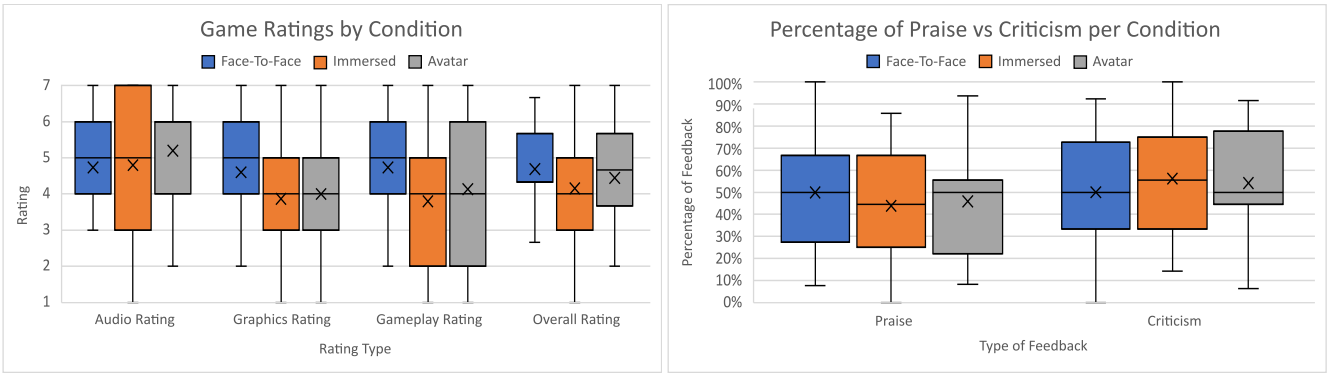

Figure 5: (Left): The numerical rating given to the evaluated game's gameplay, graphics, and audio. (Right): The percentage of feedback given in the interviews which was either praise or criticism.

Overall, the participants seemed to exhibit a positive response to the in-VR interview experience, particularly in the *Avatar* condition:

> *"Good way to ask questions rather than talking in person"* - P3 Questionnaire
> *"this is the first time I've ever done person-to-person avatar... that is quite fun as well. Nice little additional aspect."* - P2 Interview
> *"this is kind of the coolest part actually is being able to talk to someone that is a green person in VR who is not a computer"* - P45 Interview
> *"The ability to have a real-time conversation was a new experience for me in VR, and even having a basic avatar to interact with and focus my attention on went a long way towards giving a sense of presence and shared space."* - P2 Questionnaire

A benefit to this approach was that as participants were still immersed in the virtual environment during the interview they were able to discuss things they were still seeing and interacting with; for example, several participants would hold the virtual gun up to their face while discussing its visuals, fire it while commenting on the sound it made, or start scanning their surroundings when they couldn't think of something to talk about. In several instances this allowed participants to make deictic references to aspects of the environment:

> *"you can put things in to hide behind so they can't shoot you... like that block in front of us"* - P25 Interview
> *"I'm very tempted to go over towards that wall and shoot back in this direction"* - P2 Interview
> *"The text in the middle was saying stuff to me... right now it just says 'Weapon"'* - P28 Interview
> *"I don't know if I could've ducked down and hidden behind this thing?"* - P45 Interview

A common issue we encountered was that participants often assumed that the interviewer was part of the game world. This led to several instances where the participant did not know that they were supposed to respond to the interviewer's questions:

> *"Does it want me to talk back?"* - P45 Interview
> *"Do I answer or is it just pre-recorded?"* - P38 Interview
> *"Am I supposed to respond? Oh, shit. Cool. Sorry."* - P36 Interview

One participant was still confused as to whether they had talked to a real person, even after the interview had finished:

> *"I am still unsure on whether it was a person or not"* - P44 Questionnaire

The appearance of the avatar seemed to encourage more interaction and commentary on each other's representations, often serving as an icebreaker to initiate conversations. Examples of such interactions included:

> *"So is the green thing supposed to be your avatar or...?"* - P12 Interview
> *"We need a better avatar for you; it doesn't suit you."* - P39 Interview

Participants still had access to their virtual ray gun during the interview which also led to several instances of the interviewer being shot. This usually happened at the start of the interview before they realised that the avatar was human-controlled, but in one case the interviewer was shot halfway through the interview, indicating a possible dehumanising effect of the avatar:

> *"Normally in VR games, I enjoy messing with NPCs during cutscenes, and during this one, I started to but then immediately stopped once the conversation started as I began to feel embarrassed for doing so."* - P44 Questionnaire

Participants in the *Immersed* condition expressed feeling a disconnect from the interview experience arising from being "in the other world". Their remarks included:

> *"As I started to get asked about the game, I answered questions about the game which caused me to stop interacting... I felt disconnected from the interviewer, however, not truly in the game world either."* - P15 Questionnaire
> *"rather distracting too having the headset on still feeling like i was in the other world"* - P40 Questionnaire

There was also some disconnect felt by participants in the *Avatar* condition as they could not see the facial expressions or judge the mood of the interviewer other than by listening to their voice:

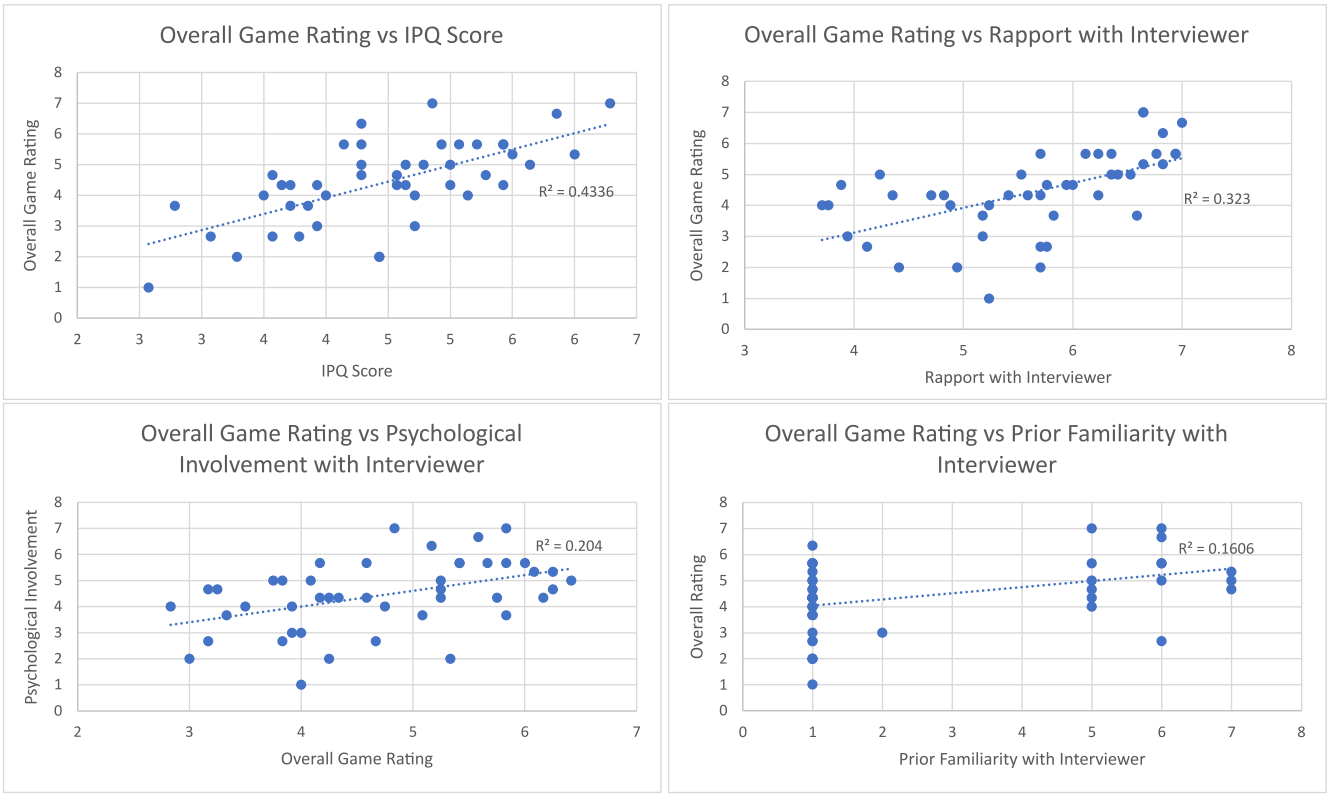

Figure 6: How the participants' overall rating of the game was affected by the induced spatial presence (top left), rapport with the interviewer (top right), psychological involvement with the interviewer (bottom left), and prior familiarity with the interviewer (bottom right). All four correlations are statistially significant, implying that the participant's interaction with the interviewer has a significant effect on the outcome of the study.

> *"hard to gauge impressions and things"* - P8 Questionnaire

They also felt less obliged to make eye contact with the interviewer:

> *"A lack of facial expressions made me feel less required to commit to eye contact."* - P38 Questionnaire

4.4 Interviewer Reflection

Here we turn to the reflection provided by the interviewer, who is also one of the authors. Due to the reflective nature this section is written in first person.

Before I provide my reflection, I would like to offer a short positionality statement. I have an interdisciplinary academic background (redacted for anonymous review) and based on my research for the past 15 years, I consider myself an HCI researcher with a primary focus on (redacted for anonymous review). I have been conducting interviews (face to face as well as over zoom/teams/skype etc.) as part of designing and evaluating digital and non-digital interventions for more than a decade. While I don't feel I am the best interviewer out there, I would say that I have a very good sense of when an interview has been successful, i.e. rapport was established, participants were comfortable, forthcoming and engaged and line of questioning was non-biased and informative/relevant to the research topic. Below are my (summarised) reflections from conducting the 45 post-game interviews across the three conditions. I made notes after each individual interview about how I felt the interview went and any notable events I could remember from it.

Wearing the headset, I could not consult my interview guide and/or make notes. Having a piece of paper, a clipboard or a notepad with our interview questions and prompts usually allows me to check if all questions have been covered, make notes of key words or non-verbal cues noted during the interview and acts as a prop for when I need to create a moment of 'personal space' either for myself or the participant, and this felt lacking.

When I conduct interviews I always consider how I present myself, not just in terms of introductions, role description and self-disclosure, but also in terms of overall appearance (choice of attire, formality etc.). I noted that it was weird, but also nice and convenient in some way, that I didn't have to worry about any of the physical appearance aspects in neither the immersed nor the avatar conditions (in the latter, I could be doing the interview from my house in my pyjamas and no one would know), as participants could not see me. While this felt freeing in some manner, at the same time, I noticed that it took away from some of the mental preparation that I do before I start an interview.

In the avatar condition I couldn't see how I looked to the participants. I was aware that my avatar looked more or less the same as the participants' avatar but for example, I hadn't realised that when I didn't hold the controllers (note that holding them was not required to conduct the interview), I was showing armless to the participants. I only realised this several interviews later when one participant pointed it out. I found that this made me quite self-conscious and from there on I started thinking much more about my body posture and making sure, that e.g. I hold the controllers, that I am not crossing my hands or putting them behind my back as I wasn't sure if it would look weird.

I found myself at several times really struggling with the immersed condition. I was standing or sitting next to the participant but had no way to make eye contact or get a sense of their facial expressions. Their head and body were often oriented away from me and several participants continued shooting or looking around the VR environment during the interview, which made it quite awkward and felt as though they were disengaged and disconnected from me. In addition, I was not able to tell whether they heard or even understood my questions.

At least in the avatar condition I could maintain face gaze; participants would look around but almost always turn and face me as they answered the questions. Both our avatars were equally handicapped by the lack of facial features which in some way put us on equal footing. Finally, I noted several occasions when my avatar suddenly appeared in the VR environment for the interview, a participant was startled or surprised, some offered compliments (e.g. "this is so cool") and showed genuine interest to find out how we had made this happen. I noted how this operated as an ice-breaker with participants that I had not met before. I found myself nervous and excited, especially in the first few interviews, to appear as an avatar and pleasantly surprised with people's reactions and overall how smoothly the interview went in that condition.

5 DISCUSSION

This work aimed to address the lack of research on conducting in-VR qualitative interviews for purposes of user evaluation and usability. It was informed by existing work on integrating questionnaires into virtual environments [61], which has been validated as an appropriate method of data collection across a wide variety of studies [21, 47, 60, 62]; and the review of existing literature on remote and virtually-mediated interviews. Based on these, we aimed to investigate the effects in-VR interviews may have on spatial presence, rapport and psychological involvement between participants and the interviewer, and participants' likelihood of praising or criticising the game. Our findings and analysis reveal a number of nuanced effects on how immersion affects the interview process and several areas for further exploration.

5.1 Spatial Presence

Our first research question was whether conducting the interview within the virtual environment, as in the *Immersed* and *Avatar* conditions, would affect participants' perception of spatial presence. We expected our results here to align with prior research on in-VR questionnaires which saw increased self-reported presence [16] and lessened breaks in presence [21, 54] by evaluating the experience before participants left virtual reality in the *Immersed* and *Avatar* conditions. This was not supported by our questionnaire results as no differences in presence were found between conditions.

Despite the lack of difference in reported presence between conditions, our qualitative results and the interviewer's reflections indicate a noteworthy sense of "disconnect" felt between the interviewer and participants, particularly in the *Immersed* condition where the realness of the environment was rated significantly lower. This suggests that introducing an aspect from outside the game world — in this case, the interviewer's disembodied voice — may have affected participants' impression of the game world's realism and fractured their experience of the virtual environment [40]. The interesting implication of this result is that the manner in which the interview is conducted may have some effect on the questionnaire results.

5.2 Rapport and Psychological Involvement

Our second and third research questions related to whether participants in the *Immersed* and *Avatar* conditions would develop less rapport and psychological involvement with the interviewer due to an inability to see them, as has previously been seen in phone- and computer-mediated interviews [12, 27, 46]. This was partially supported by our results; while rapport was consistent across conditions, participants in the *Immersed* condition experienced less psychological involvement with the interviewer. Our qualitative analysis and the interviewer reflection suggests this could be due to an inability to gauge emotions and reactions, which is supported by prior research on computer-mediated interaction [38]. In contrast to the *Immersed* condition, psychological involvement in the *Avatar* condition was similar to the face to face interview which is a promising finding with regards to the ability to build rapport. This opens up questions around how in-VR interviewers and interviewees can be visually represented to ensure rapport can be developed.

A possible explanation for the similarity in rapport across conditions is that our interviews were simply not long enough to build rapport, which typically develops in four stages: apprehension, exploration, co-operation, and participation [14]; given our average interview time of four minutes, it is possible that there was insufficient time for participants to progress past the "apprehension" phase, particularly in the *Avatar* condition where participants were often taken aback by the avatar's appearance. The interviewer's reflection suggests, however, that this initial surprise was beneficial as an 'icebreaker' with participants that were not known to them.

An interesting result of our work was that only the perceived empathy of the *interviewer* differed between conditions, but not the participants' empathy towards the interviewer. The implication here seems to be that having a visual representation of the interviewer makes participants seem more capable of empathy, and that even an abstract avatar is enough to "humanise" an interviewer in this way.

However, the avatar's appearance alone was likely not enough. We noted incidents of participants shooting the interviewer until the heard their voice. Some participants also suggested that facial expressions would have helped with their perception of the interviewer's mood. This is also supported by prior research showing preference for more realistic representations of interlocutors in

virtual environments [3]. It thus seems that avatars may serve as an acceptable substitute for when mutual empathy is desired, but both voice and appearance are important for establishing psychological involvement between participants and the interviewer. Future research should investigate the effects that different voices and avatar representations have on participant rapport, for example if photorealistic avatars induce more rapport than the abstract one we chose.

5.3 Critique and Negativity

Our fourth research question was whether conducting the interview within VR would result in more critical feedback about the game, which was closely related to our expectations on how rapport would develop. Our assumption was by not being able to see the interviewer, participants would feel less pressured to praise the game and would be more forthcoming with socially sensitive criticism, as has been seen in phone- and computer-mediated interviews [9, 18, 28, 38]. While there was no significant effect between conditions, we did see the *Immersed* condition provide the highest number of critical comments overall (followed by the *Avatar* condition, then *Face-to-Face*). We also found that participants who reported higher rapport were more likely to praise the game during the interview and rate it highly in the questionnaire.

This interestingly contradicts previous research in clinical interviews where higher rapport results in increased disclosure [13, 39]; in our study, rapport resulted in *decreased* disclosure of game defects as honesty in this case had the potential to negatively affect the social bond that had developed [44]. This raises questions on whether increased rapport is always appropriate in qualitative research, and how this can be mediated by the appearance (or lack of appearance) of the avatar [3].

Since no significant results were found with respect to sentiment (as discussed in section 4.2.6), we do not see any condition in which the interview provided a decidedly more negative sentiment towards the game. Further, our results do not show a relationship between criticism feedback and negative sentiment in the interviews, which implies that in user interviews for usability studies, for example, the interview could contain critical feedback, without being a negative experience for the participant and interviewer.

5.4 Effects of Prior Familiarity with the Interviewer

Though it was not part of our original set of research questions, an existing relationship between the participant and the interviewer had a significant effect on the results of the study. The effect of a prior familiarity was found in the reported spatial presence, rapport, and the overall game rating, with those more familiar with the interviewer reporting significantly higher scores in these areas. Familiarity seems also to affect sentiment of the interviews and open comments on game feedback: those familiar with the interviewer exhibited a more positive sentiment about the game. This could be related to how we introduced our study to participants, giving them the impression that we were seeking feedback for a game *we* had developed.

Prior research has shown that participants are more forthcoming if they believe they will never cross paths with the researcher again [9], however, we found that familiarity did not significantly affect how much feedback was provided or how much of this feedback was praise or criticism. Our qualitative results did show that some participants were hesitant about giving critical feedback, which might suggest that user interviews for usability feedback, for example, may require more work on behalf of the interviewer to reassure the participant and make them feel safe about providing criticism.

6 LIMITATIONS AND FUTURE WORK

Our results suggest that there are still many issues to explore with regards to conducting qualitative interviews within virtual environments. Here we discuss the limitations and avenues for future research.

6.1 Representation of the Virtual Interviewer

In our study all participants were interviewed by the same person to ensure that any results across our conditions were not affected by the personality of the chosen interviewer. However, it may be interesting to determine what effects the interviewer's personality and appearance may have had on participants' rapport with them and their subsequent responses, and especially the appearance of their avatar [3].

It may also be worth exploring if altering the interviewer's voice has any effect on participant responses. Despite our attempts to hide the interviewer's identity in the *Immersed* and *Avatar* conditions, there were instances where the participant immediately recognised them based on their voice, which as seen from our results may have pressured them to respond more favourably towards the game. Masking the interviewer's voice, for example using a robotic filter, could provide interesting results by further anonymising them or altering the participant's perception of who they are.

6.2 Analysing the Sentiment of Violent Game Content

SentiStrength's algorithm was trained on product and movie reviews [66]. Using SentiStrength for analysing texts from domains other than what the algorithm has been trained on, e.g., software engineering or games, does bring some limitations [31], and yet it remains one of the dominant tools for this work [29].

SentiStrength can be trained on user-defined data, however the amount of effort required to achieve this in the context of this study, outweighs the benefits. We encountered similar issues in our data to those found by Viggiato et al. [71], in that sentiment in our dataset could be classified incorrectly when a participant provided positive and negative feedback in a single sentence, or when words such as 'gun' and 'killed' were used (which are automatically assigned strong negative sentiment by SentiStrength). We checked these instances and found that they had little impact on the overall sentiment score of the interview and so these were ignored. A more focused corpus of virtual reality game terms for training the SentiStrength algorithm may yield more accurate results for interview sentiment.

6.3 Study Design

It is possible that any potential differences in presence across conditions were limited by our study design. As all participants completed

the questionnaire in the real world, all experienced the same break in presence resulting from taking off their headset that is already known from the literature [33]. Future research could consider comparing between conducting the entire study within VR (including the questionnaire *and* the interview) and the traditional real-world approach. Headsets with 'passthrough' capability that incorporate cameras and sensors that allow the user to 'see' the real world from inside their VR environment [34] may offer a more seamless transition between VR and the real world as another point of comparison for future work.

A further limitation with regards to our study design was the short exposure time to the virtual environment. It is possible that the effects in presence were impacted due to this, but equally it may be that longer gameplay or interview time could make participants more susceptible to the negative impacts of VR such as simulator sickness [15]. Our study also did not consider users with varied abilities or other needs such as older adults [26]. Further research is required to investigate the effects and best practice for qualitative interviews of diverse user groups.

6.4 Further Applications for In-VR Interviews

As this paper was exploratory in nature it would be interesting to see how the in-VR approach could be applied to different domains. For example, there has been increasing interest in conducting usability studies remotely in recent years, driven in part by the COVID-19 pandemic [70]. These have typically seen any interviews conducted by videoconferencing applications such as Skype or Zoom, however the use of in-VR avatars would provide a more interactive way to conduct this research more connected to the experience being evaluated. For several of our interviews the interviewer and participant were separated by several kilometers, proving the feasibility of this approach, however it is yet to be seen what affects this might have on data quality compared to a videoconferencing approach.

Another interesting application of this research is stealth assessment, which is the act of gathering feedback from a participant without their direct knowledge [64]. Many of our participants in the *Avatar* condition did not immediately realise that the avatar was human-controlled, with several asking if it was an AI. Future studies could consider capitalising on this by insisting that the interviewer is computer-controlled, and possibly further this illusion by presenting them as a non-playable character, to determine the effects this may have on self-disclosure and data quality.

7 CONCLUSION

In this paper we explored how qualitative interviews can be conducted from within the virtual experience they are evaluating. This was done in two ways: with the participant keeping their virtual reality headset on while the interviewer was in the room, and by integrating the interviewer into the virtual experience through the use of a remotely-controlled avatar.

The effects that in-VR interviewing may have were explored through the evaluation of a virtual reality shooting game which included several intentional oversights in its design. Through this we discovered that in-VR interviews result in similar data quality and interviewer rapport to face-to-face interviews, but that having a visual representation of the interviewer in the environment is important for maintaining the emotional connection with the participant and to avoid fracturing the experience and resulting in a poorer impression of the virtual environment. We also found that those who established stronger rapport with the interviewer were more likely to give positive feedback about the game, and that this was especially true for participants who knew the interviewer beforehand and already had pre-established rapport with them.

Through this research we have shown that conducting interviews virtually is a valid alternative to face-to-face conversations, either when in-person meetings aren't available or when participant anonymity is desired. This opens up a promising avenue for future research in a variety of fields, however further research is still required to determine how virtual interviews can best be supported and ensure the best possible experience for participants and researchers alike.

REFERENCES

[1] Deeksha Adiani, Aaron Itzkovitz, Dayi Bian, Harrison Katz, Michael Breen, Spencer Hunt, Amy Swanson, Timothy J Vogus, Joshua Wade, and Nilanjan Sarkar. 2022. Career interview readiness in virtual reality (CIRVR): a platform for simulated interview training for autistic individuals and their employers. *ACM Transactions on Accessible Computing (TACCESS)* 15, 1 (2022), 1–28.

[2] Dmitry Alexandrovsky, Susanne Putze, Michael Bonfert, Sebastian Höffner, Pitt Michelmann, Dirk Wenig, Rainer Malaka, and Jan David Smeddinck. 2020. Examining Design Choices of Questionnaires in VR User Studies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (, Honolulu, HI, USA,) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–21. https://doi.org/10.1145/3313831.3376260

[3] Sahar Aseeri and Victoria Interrante. 2021. The influence of avatar representation on interpersonal communication in virtual social environments. *IEEE transactions on visualization and computer graphics* 27, 5 (2021), 2608–2617.

[4] Andreas Beckenbach. 1995. Computer-assisted questioning: the new survey methods in the perception of the respondents. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 48, 1 (1995), 82–100.

[5] Frank J Bernieri, Janet M Davis, Robert Rosenthal, and C Raymond Knee. 1994. Interactional synchrony and rapport: Measuring synchrony in displays devoid of sound and facial affect. *Personality and social psychology bulletin* 20, 3 (1994), 303–311.

[6] Andrea J Bingham and Patricia Witkowsky. 2021. Deductive and inductive approaches to qualitative data analysis. In *Analyzing and interpreting qualitative data: After the interview*, C Vanover and J. Mihas, P. Saldaña (Eds.). Newbury Park, California: SAGE Publications, Newbury Park, California, 133–146.

[7] Frank Biocca, Chad Harms, and Judee K Burgoon. 2003. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & virtual environments* 12, 5 (2003), 456–480.

[8] Frank Biocca, Chad Harms, and Jenn Gregg. 2001. The Networked Minds Measure of Social Presence: Pilot Test of the Factor Structure and Concurrent Validity. In *4th Annual International Workshop on Presence*. ACM, MIT Press, 55 Hayward St., Cambridge, MA, United States, 1–9. http://astro.temple.edu/~lombard/ISPR/Proceedings/2001/Biocca2.pdf

[9] Julia Brannen. 1988. The Study of Sensitive Subjects. *The Sociological Review* 36, 3 (1988), 552–563. https://doi.org/10.1111/j.1467-954X.1988.tb02929.x arXiv:https://doi.org/10.1111/j.1467-954X.1988.tb02929.x

[10] SungIk Cho, Seung wook Kim, JongMin Lee, JeongHyeon Ahn, and JungHyun Han. 2020. Effects of volumetric capture avatars on social presence in immersive virtual environments. In *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE, Atlanta, USA, 26–34. https://doi.org/10.1109/vr46266.2020.00020

[11] Gina M Coon, David Pena, and Paul A Illich. 1998. Self-efficacy and substance abuse: assessment using a brief phone interview. *Journal of Substance Abuse Treatment* 15, 5 (1998), 385–391.

[12] Hannah Deakin and Kelly Wakefield. 2014. Skype interviewing: Reflections of two PhD researchers. *Qualitative research* 14, 5 (2014), 603–616.

[13] David Devault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, Gale Lucas, Stacy Marsella, Fabrizio Morbini, Angela Nazarian, Stefan Scherer, Giota Stratou, Apar Suri, David Traum, Rachel Wood, Yuyu Xu, Albert Rizzo, and Louis-Philippe Morency. 2014. SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. ACM, Paris, France, 1061–1068. http://simsensei.ict.usc.edu/

[14] Barbara DiCicco-Bloom and Benjamin F Crabtree. 2006. The qualitative research interview. *Medical Education* 40, 4 (2006), 314–321. https://doi.org/10.1111/j.1365-2929.2006.02418.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2929.2006.02418.x

[15] Natalia Dużmańska, Paweł Strojny, and Agnieszka Strojny. 2018. Can simulator sickness be avoided? A review on temporal aspects of simulator sickness. *Frontiers in psychology* 9 (2018), 2132.

[16] Julian Frommel, Katja Rogers, Julia Brich, Daniel Besserer, Leonard Bradatsch, Isabel Ortinau, Ramona Schabenberger, Valentin Riemer, Claudia Schrader, and Michael Weber. 2015. Integrated questionnaires: Maintaining presence in game environments for self-reported data acquisition. In *CHI PLAY 2015 - Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*. Association for Computing Machinery, Inc, London, UK, 359–368. https://doi.org/10.1145/2793107.2793130

[17] Scott W Greenwald, Zhangyuan Wang, Markus Funk, and Pattie Maes. 2017. Investigating social presence and communication with embodied avatars in room-scale virtual reality. In *Immersive Learning Research Network: Third International Conference, iLRN 2017, Coimbra, Portugal, June 26–29, 2017. Proceedings 3*. Springer, Springer, Coimbra, Portugal, 75–90.

[18] John H Greist, Thomas P. Laughren, David H. Gustafson, Fred F. Stauss, Glen L. Rowse, and John A. Chiles. 1973. A computer interview for suicide-risk prediction. *American Journal of Psychiatry* 130, 12 (1973), 1327–1332.

[19] Dwayne D Gremler and Kevin P Gwinner. 2000. Customer-employee rapport in service relationships. *Journal of service research* 3, 1 (2000), 82–104.

[20] Nathan Navarro Griffin, James Liu, and Eelke Folmer. 2018. Evaluation of Hands-busy vs Handsfree Virtual Locomotion. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play* (Melbourne, VIC, Australia,) *(CHI PLAY '18)*. Association for Computing Machinery, New York, NY, USA, 211–219. https://doi.org/10.1145/3242671.3242707

[21] Jan P Gründling, Daniel Zeiler, and Benjamin Weyers. 2022. Answering With Bow and Arrow: Questionnaires and VR Blend Without Distorting the Outcome. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, IEEE, Christchurch, New Zealand, 683–692.

[22] Judith A Hall and Frank J Bernieri. 2001. *Interpersonal sensitivity: Theory and measurement*. Psychology Press, Mahwah, New Jersey, USA.

[23] B Hensen, CRS Mackworth-Young, M Simwinga, N Abdelmagid, J Banda, C Mavodza, AM Doyle, C Bonell, and HA Weiss. 2021. Remote data collection for public health research in a COVID-19 era: ethical implications, challenges and opportunities. *Health policy and planning* 36, 3 (2021), 360–368.

[24] Patricia E Hershberger and Karen Kavanaugh. 2017. Comparing appropriateness and equivalence of email interviews to phone interviews in qualitative research on reproductive decisions. *Applied Nursing Research* 37 (2017), 50–54.

[25] Valeria Lo Iacono, Paul Symonds, and David H.K. Brown. 2016. Skype as a Tool for Qualitative Research Interviews. *Sociological Research Online* 21, 2 (2016), 103–117. https://doi.org/10.5153/sro.3952 arXiv:https://doi.org/10.5153/sro.3952

[26] Kiran Ijaz, Tram Thi Minh Tran, Ahmet Baki Kocaballi, Rafael A Calvo, Shlomo Berkovsky, and Naseem Ahmadpour. 2022. Design considerations for immersive virtual reality applications for older adults: a scoping review. *Multimodal Technologies and Interaction* 6, 7 (2022), 60.

[27] Brandy M Jenner and Kit C Myers. 2019. Intimacy, rapport, and exceptional disclosure: A comparison of in-person and mediated interview contexts. *International Journal of Social Research Methodology* 22, 2 (2019), 165–177.

[28] Adam N. Joinson. 2001. Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology* 31 (3 2001), 177–192. Issue 2. https://doi.org/10.1002/ejsp.36

[29] Robbert Jongeling, Subhajit Datta, and Alexander Serebrenik. 2015. Choosing your weapons: On sentiment analysis tools for software engineering research. In *2015 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. IEEE, Bremen, Germany, 531–535. https://doi.org/10.1109/ICSM.2015.7332508

[30] Timothy Jung, M. Claudia tom Dieck, Natasha Moorhouse, and Dario tom Dieck. 2017. Tourists' experience of Virtual Reality applications. In *2017 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, Berlin, Germany, 208–210. https://doi.org/10.1109/ICCE.2017.7889287

[31] Arvinder Kaur, Amrit Pal Singh, Guneet Singh Dhillon, and Divesh Bisht. 2018. Emotion Mining and Sentiment Analysis in Software Engineering Domain. In *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. IEEE, Coimbatore, India, 1170–1173. https://doi.org/10.1109/ICECA.2018.8474619

[32] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology* 3, 3 (1993), 203–220.

[33] Jarrod Knibbe, Jonas Schjerlund, Mathias Petraeus, and Kasper Hornbæk. 2018. The Dream is Collapsing: The Experience of Exiting VR. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (, Montreal QC, Canada,) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174057

[34] Grace Kuo, Eric Penner, Seth Moczydlowski, Alexander Ching, Douglas Lanman, and Nathan Matsuda. 2023. Perspective-Correct VR Passthrough Without Reprojection. In *ACM SIGGRAPH 2023 Conference Proceedings* (, Los Angeles, CA, USA,) *(SIGGRAPH '23)*. Association for Computing Machinery, New York, NY, USA, Article 15, 9 pages. https://doi.org/10.1145/3588432.3591534

[35] Steinar Kvale. 2007. *Doing Interviews*. SAGE Publications, Inc, London, UK.

[36] Joung Huem Kwon, John Powell, and Alan Chalmers. 2013. How level of realism influences anxiety in virtual reality environments for a job interview. *International journal of human-computer studies* 71, 10 (2013), 978–987.

[37] Sally Lindsay. 2022. A comparative analysis of data quality in online zoom versus phone interviews: An example of youth with and without disabilities. *Sage Open* 12, 4 (2022), 21582440221140098.

[38] Gale M. Lucas, Jonathan Gratch, Aisha King, and Louis Philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37 (2014), 94–100. https://doi.org/10.1016/j.chb.2014.04.043

[39] Gale M. Lucas, Albert Rizzo, Jonathan Gratch, Stefan Scherer, Giota Stratou, Jill Boberg, and Louis Philippe Morency. 2017. Reporting mental health symptoms: Breaking down barriers to care with virtual human interviewers. *Frontiers Robotics AI* 4 (10 2017), 1–9. Issue OCT. https://doi.org/10.3389/frobt.2017.00051

[40] Paul Luff, Christian Heath, Hideaki Kuzuoka, Jon Hindmarsh, Keiichi Yamazaki, and Shinya Oyama. 2003. Fractured Ecologies: Creating Environments for Collaboration. *Human–Computer Interaction* 18, 1-2 (2003), 51–84. https://doi.org/10.1207/S15327051HCI1812_3 arXiv:https://doi.org/10.1207/S15327051HCI1812_3

[41] Steve Mann. 2010. A Critical Review of Qualitative Interviews in Applied Linguistics. *Applied Linguistics* 32, 1 (11 2010), 6–24. https://doi.org/10.1093/applin/amq043 arXiv:https://academic.oup.com/applij/article-pdf/32/1/6/14093793/amq043.pdf

[42] Henry Matovu, Dewi Ayu Kencana Ungu, Mihye Won, Chin-Chung Tsai, David F. Treagust, Mauro Mocerino, and Roy Tasker. 2023. Immersive virtual reality for science learning: Design, implementation, and evaluation. *Studies in Science Education* 59, 2 (2023), 205–244. https://doi.org/10.1080/03057267.2022.2082680 arXiv:https://doi.org/10.1080/03057267.2022.2082680

[43] Tina Miller. 2017. Telling the difficult things: Creating spaces for disclosure, rapport and 'collusion' in qualitative interviews. *Women's Studies International Forum* 61 (2017), 81–86. https://doi.org/10.1016/j.wsif.2016.07.005

[44] Divya Natesan, Morgan Walker, and Shannon Clark. 2016. Cognitive Bias in Usability Testing. *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care* 5, 1 (2016), 86–88. https://doi.org/10.1177/2327857916051015 arXiv:https://doi.org/10.1177/2327857916051015

[45] Kureha Noguchi, Yoshinari Takegawa, Yutaka Tokuda, Yuta Sugiura, Katsutoshi Masai, and Keiji Hirata. 2021. Study of Interviewee's Impression Made by Interviewer Wearing Digital Full-face Mask Display During Recruitment Interview. In *Proceedings of the 9th International Conference on Human-Agent Interaction*. ACM, Online, 323–327.

[46] Gina Novick. 2008. Is there a bias against telephone interviews in qualitative research? *Research in nursing & health* 31, 4 (2008), 391–398.

[47] Sebastian Oberdörfer, David Heidrich, and Marc Erich Latoschik. 2019. Usability of Gamified Knowledge Learning in VR and Desktop-3D. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300405

[48] Lisa J Orchard and Chris Fullwood. 2010. Current perspectives on personality and Internet use. *Social Science Computer Review* 28, 2 (2010), 155–169.

[49] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. 2016. Holoportation: Virtual 3D Teleportation in Real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 741–754. https://doi.org/10.1145/2984511.2984517

[50] Ye Pan and Anthony Steed. 2017. The impact of self-avatars on trust and collaboration in shared virtual environments. *PloS one* 12, 12 (2017), e0189078.

[51] Jung-Woo Noel Park, Steven Mills, Hēmi Whaanga, Paora Mato, Robert W Lindeman, and Holger Regenbrecht. 2019. Towards a māori telepresence system. In *2019 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, Vol. Dec. IEEE, IEEE, Dunedin, New Zealand, 1–6.

[52] Paul N Pfeiffer, Adrian J Blow, Erin Miller, Jane Forman, Gregory W Dalack, and Marcia Valenstein. 2012. Peers and peer-based interventions in supporting reintegration and mental health among National Guard soldiers: A qualitative study. *Military Medicine* 177, 12 (2012), 1471–1476.

[53] Jonathan Potter and Alexa Hepburn. 2005. Qualitative interviews in psychology: problems and possibilities. *Qualitative Research in Psychology* 2, 4 (2005), 281–307. https://doi.org/10.1191/1478088705qp045oa arXiv:https://doi.org/10.1191/1478088705qp045oa

[54] Susanne Putze, Dmitry Alexandrovsky, Felix Putze, Sebastian Höffner, Jan David Smeddinck, and Rainer Malaka. 2020. Breaking The Experience: Effects of

Questionnaires in VR User Studies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (, Honolulu, HI, USA,) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376144

[55] Rivu Radiah, Ville Mäkelä, Sarah Prange, Sarah Delgado Rodriguez, Robin Piening, Yumeng Zhou, Kay Köhle, Ken Pfeuffer, Yomna Abdelrahman, Matthias Hoppe, et al. 2021. Remote VR studies: A framework for running virtual reality studies remotely via participant-owned HMDs. *ACM Transactions on Computer-Human Interaction (TOCHI)* 28, 6 (2021), 1–36.

[56] Eeva Raita. 2012. User Interviews Revisited: Identifying User Positions and System Interpretations. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design* (Copenhagen, Denmark) *(NordiCHI '12)*. Association for Computing Machinery, New York, NY, USA, 675–682. https://doi.org/10.1145/2399016.2399119

[57] Bernhard E. Riecke and Daniel Feuereissen. 2012. To Move or Not to Move: Can Active Control and User-Driven Motion Cueing Enhance Self-Motion Perception ("vection"') in Virtual Reality?. In *Proceedings of the ACM Symposium on Applied Perception* (Los Angeles, California) *(SAP '12)*. Association for Computing Machinery, New York, NY, USA, 17–24. https://doi.org/10.1145/2338676.2338680

[58] Kathryn Roulston. 2010. Considering quality in qualitative interviewing. *Qualitative Research* 10, 2 (2010), 199–228.

[59] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments* 10 (2001), 266–281. Issue 3. https://doi.org/10.1162/105474601300343603

[60] Valentin Schwind, Pascal Knierim, Lewis Chuang, and Niels Henze. 2017. "Where's Pinky?": The Effects of a Reduced Number of Fingers in Virtual Reality. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (, Amsterdam, The Netherlands,) *(CHI PLAY '17)*. Association for Computing Machinery, New York, NY, USA, 507–515. https://doi.org/10.1145/3116595.3116596

[61] Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. 2019. Using Presence Questionnaires in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300590

[62] Valentin Schwind, Pascal Knierim, Cagri Tasci, Patrick Franczak, Nico Haas, and Niels Henze. 2017. "These are not my hands!": Effect of Gender on the Perception of Avatar Hands in Virtual Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1577–1582. https://doi.org/10.1145/3025453.3025602

[63] Sally Seitz. 2016. Pixilated partnerships, overcoming obstacles in qualitative interviews via Skype: a research note. *Qualitative Research* 16, 2 (2016), 229–235. https://doi.org/10.1177/1468794115577011 arXiv:https://doi.org/10.1177/1468794115577011

[64] Valerie J Shute. 2011. Stealth assessment in computer-based games to support learning. *Computer games and instruction* 55, 2 (2011), 503–524.

[65] Roger W Shuy. 2003. In-person versus telephone interviewing. *Handbook of interview research: Context and method* 0 (2003), 175–193.

[66] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. 2010. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology* 61, 12 (2010), 2544–2558. https://doi.org/10.1002/asi.21416

[67] Mike Thelwall, Eleanor-Rose Papas, Zena Nyakoojo, Liz Allen, and Verena Weigert. 2020. Automatically detecting open academic review praise and criticism. *Online Information Review* 44, 5 (2020), 1057–1076.

[68] Sara Thunberg and Linda Arnell. 2022. Pioneering the use of technologies in qualitative research–A research review of the use of digital interviews. *International Journal of Social Research Methodology* 25, 6 (2022), 757–768.

[69] Linda Tickle-Degnen and Robert Rosenthal. 1990. The nature of rapport and its nonverbal correlates. *Psychological inquiry* 1, 4 (1990), 285–293.

[70] Livia Tomás and Ophélie Bidet. 2023. Conducting qualitative interviews via VoIP technologies: reflections on rapport, technology, digital exclusion, and ethics. *International Journal of Social Research Methodology* 0 (2023), 1–13.

[71] Markos Viggiato, Dayi Lin, Abram Hindle, and Cor-Paul Bezemer. 2022. What Causes Wrong Sentiment Classifications of Game Reviews? *IEEE Transactions on Games* 14, 3 (2022), 350–363. https://doi.org/10.1109/TG.2021.3072545

[72] VRChat Inc. 2023. VRChat. https://hello.vrchat.com/. Accessed: 2023-09-13.

[73] Susie Weller. 2017. Using internet video calls in qualitative (longitudinal) interviews: some implications for rapport. *International Journal of Social Research Methodology* 20, 6 (2017), 613–625. https://doi.org/10.1080/13645579.2016.1269505 arXiv:https://doi.org/10.1080/13645579.2016.1269505

[74] Kevin Yu, Gleb Gorbachev, Ulrich Eck, Frieder Pankratz, Nassir Navab, and Daniel Roth. 2021. Avatars for Teleconsultation: Effects of Avatar Embodiment Techniques on User Perception in 3D Asymmetric Telepresence. *IEEE Transactions on Visualization and Computer Graphics* 27 (2021), 4129–4139. Issue 11. https://doi.org/10.1109/TVCG.2021.3106480

A PARTICIPANT DEMOGRAPHICS

Received 14 September 2023; revised 12 December 2023; accepted 19 January 2024

Participant	Condition	Age	Gender	Ethnicity	Virtual Reality Usage	Video Game Usage	Prior Familiarity with Interviewer (1 = none)
P1	Face-To-Face	47	Non-binary	Other	Sub-yearly	Yearly	1
P2	Avatar	40	Male	European	Weekly	Monthly	6
P3	Avatar	30	Female	Asian	Weekly	Weekly	7
P4	Face-To-Face	31	Male	Asian	Weekly	Monthly	6
P5	Immersed	36	Male	Latin American	Yearly	Weekly	1
P6	Face-To-Face	54	Male	European	Yearly	Monthly	7
P7	Immersed	42	Male	Asian	Sub-yearly	Sub-yearly	1
P8	Avatar	23	Male	European	Never	Weekly	1
P9	Avatar	22	Non-binary	Asian	Yearly	Monthly	1
P10	Immersed	27	Male	European	Never	Monthly	1
P11	Face-To-Face	35	Female	European	Yearly	Yearly	1
P12	Avatar	41	Male	European	Sub-yearly	Weekly	1
P13	Face-To-Face	19	Male	Asian	Sub-yearly	Weekly	1
P14	Immersed	23	Male	European	Yearly	Weekly	6
P15	Immersed	21	Male	European	Never	Weekly	1
P16	Avatar	27	Male	European	Sub-yearly	Yearly	1
P17	Face-To-Face	38	Male	Middle Eastern	Never	Yearly	1
P18	Face-To-Face	20	Male	European, Pacific	Monthly	Monthly	1
P19	Immersed	51	Female	European	Sub-yearly	Yearly	6
P20	Face-To-Face	33	Male	Asian	Yearly	Yearly	6
P21	Face-To-Face	26	Male	European,Asian	Yearly	Weekly	1
P22	Immersed	29	Female	Asian	Sub-yearly	Weekly	1
P23	Face-To-Face	30	Male	Prefer not to say	Weekly	Weekly	5
P24	Immersed	45	Male	European	Never	Weekly	2
P25	Avatar	57	Male	European	Sub-yearly	Never	1
P26	Avatar	37	Male	Asian	Sub-yearly	Weekly	1
P27	Immersed	34	Female	Asian	Yearly	Weekly	1
P28	Immersed	30	Male	European	Yearly	Monthly	1
P29	Face-To-Face	31	Non-binary	Asian	Monthly	Yearly	1
P30	Avatar	24	Female	Asian	Never	Monthly	5
P31	Immersed	43	Female	European	Never	Weekly	1
P32	Avatar	39	Female	Prefer not to say, Other	Monthly	Never	7
P33	Immersed	30	Male	Asian	Yearly	Monthly	6
P34	Face-To-Face	27	Male	European	Yearly	Weekly	1
P35	Immersed	24	Male	European	Yearly	Weekly	1
P36	Avatar	25	Female	European	Monthly	Weekly	1
P37	Face-To-Face	23	Female	European	Sub-yearly	Monthly	1
P38	Avatar	22	Male	European	Sub-yearly	Monthly	1
P39	Avatar	27	Male	Other	Weekly	Weekly	5
P40	Immersed	24	Male	Māori, European	Sub-yearly	Yearly	5
P41	Face-To-Face	22	Male	European	Sub-yearly	Yearly	1
P42	Immersed	22	Undisclosed	European	Yearly	Weekly	5
P43	Face-To-Face	26	Male	Asian	Yearly	Sub-yearly	5
P44	Avatar	23	Male	European	Yearly	Weekly	1
P45	Avatar	28	Male	European	Monthly	Monthly	1