Immersive Virtual Reality in Child Interview Skills Training: A Comparison of 2D and 3D Environments

Pegah Salehi*
SimulaMet, Norway

Syed Zohaib Hassan†
SimulaMet, Norway

Gunn Astrid Baugerud
OsloMet, Norway

Martine Powell
Griffith University, Australia

M. Cayetana López Cano
OsloMet, Norway

Miriam S. Johnson
OsloMet, Norway

Ragnhild Klingenberg Røed
OsloMet, Norway

Dag Johansen
UiT The Arctic University of Norway

Saeed Shafiee Sabet
SimulaMet, Norway

Michael A. Riegler†
SimulaMet, Norway

Pål Halvorsen†
SimulaMet, Norway

ABSTRACT

The current study aims to evaluate and compare the subjective quality of an AI-based training system developed for conducting child interviews, focusing on the distinction between immersive 3D (using virtual reality) and 2D desktop environments. To this end, a structured user study was conducted, involving 36 participants who were exposed to these two distinct environments. The study evaluated various aspects of user experience, namely *presence*, *usability*, *visual fidelity*, *emotion*, *responsiveness*, *appropriateness*, and *training effectiveness*. The findings reveal significant differences in user experience between the 2D and 3D environments. Notably, the 3D environment enhanced *presence*, *visual fidelity*, *training effectiveness*, and *empathy*. In contrast, the 2D environment was favored for *usability*. The study highlights the potential of immersive VR while also pointing out the need to improve the system response and emotional expressiveness of the avatars.

CCS CONCEPTS

• **Human-centered computing → User studies**; **Virtual reality**;
• **Computing methodologies** → *Machine learning*.

KEYWORDS

Virtual Reality (VR), Virtual Environments (VEs), Immersion, Quality of Experience (QoE), Large language model (LLM)

ACM Reference Format:
Pegah Salehi, Syed Zohaib Hassan, Gunn Astrid Baugerud, Martine Powell, M. Cayetana López Cano, Miriam S. Johnson, Ragnhild Klingenberg Røed, Dag Johansen, Saeed Shafiee Sabet, Michael A. Riegler, and Pål Halvorsen. 2024. Immersive Virtual Reality in Child Interview Skills Training: A Comparison of 2D and 3D Environments. In *Proceedings of the international workshop on immersive mixed and virtual environment systems (MMVE '24).* ACM, New York, NY, USA, 7 pages. https://doi.org/3652212.3652219

1 INTRODUCTION

Child abuse is a critical global issue that negatively affects children's development, mental and physical well-being. Meta-analysis studies have estimated that 22.6% of children experience physical abuse and 11.8% experience sexual abuse before reaching adulthood [36]. In particular, less than 15% of child sexual abuse (CSA) cases are corroborated by physical evidence [42], and in 70% of these cases, the child is the sole witness [11]. Properly conducted investigative interviews are crucial for prosecution, as children can be reliable witnesses when interviewed according to best-practice guidelines [4]. Such guidelines encourage communication with the child through open-ended questions to obtain elaborate and relevant evidence details [7]. Unfortunately, these guidelines are not often followed despite investments in training programs [18]; interviewers tend to use too many suggestive and closed questions, and directive questions generate short, rather than elaborate, responses [1, 2, 8]. This is due to training programs that contain inadequate practice opportunities (with feedback) that shape interviewer performance in the use of open questions [44]. Recent studies, including a decade-long Norwegian study, indicate that there is no significant advancement in interview quality despite training innovations. Mock interviews with trained respondents have been shown to be highly effective in shaping interviewer performance [27].

However, face-to-face training with a trained respondent is expensive and cumbersome to set up. It requires the availability of both parties and considerable investment to ensure that the trained respondent responds in a way that effectively shapes the performance of the trainee interviewer. To potentially improve the update of practice opportunities, we introduced an innovative AI-based training platform to conduct investigative interviews [14, 31]. This platform integrates VR technology with an advanced avatar system. The avatar is designed to emulate the responses of children in high-fidelity abuse scenarios, enhancing the realism of the digital training environment. The core of our system lies in the seamless

*Also affiliated with UiT The Arctic University of Norway
†Also affiliated with OsloMet, Norway

MMVE '24, April 15-18, 2024, Bari, Italy

ACM ISBN 979-8-4007-0618-9/24/04...$15.00
https://doi.org/3652212.3652219

integration of state-of-the-art natural language processing (NLP) and vision technologies. NLP empowers the avatar with the ability to understand and express human language, while vision technologies enable them to present visually accurate and responsive representations of human-trained respondents. This dual-technology approach effectively simulates the complexities involved in interviewing child abuse victims, providing a crucial training tool for professionals in this field.

In our previous work [14], we conducted a user study that evaluated the efficacy of different interactive platforms, including VR, 2D desktop, audio, and text chat; however, it lacks crucial elucidation aspects of realism, such as *emotion* and *visual fidelity*. Also, the study may not have possessed ideal statistical power to discern certain subtle effects. This study focuses on a detailed assessment of qualitative feedback within both 2D and 3D visual environments, using a comprehensive questionnaire to investigate various evaluation aspects. In this paper, we present the following main contributions:

- Fine-tuning of the GPT-3 [6] using interview data from abused children and seamlessly integrating the dialogue model with a Unity3D framework.
- Demonstration of the potential of virtual reality (VR) for professional training in real-world scenarios.
- A comparison of trainee interviewers' perceptions across different evaluative aspects in 2D and 3D environments, including *presence*, *usability*, *visual fidelity*, *emotion*, *responsiveness*, *appropriateness*, *training effectiveness*, and *empathy*.

2 RELATED WORK

Child avatar training systems aimed at improving interview quality exist, but none are fully automated. One uses prerecorded child responses manually selected by an operator [9], another [26] employs a probabilistic rule-based algorithm for response selection after a person has categorized the question manually, that still requires human intervention. The existing systems do not utilize advanced large language models such as GPT-3 [6]. Instead, they are constrained and limited in generating responses, relying on predetermined sentences from which to select a response. Also, constrained by their dependency on human operators, these systems can have potential issues with flexibility, operational costs, and error rates.

Parallel to these developments, the emergence of VR as a dynamic educational tool is significant [25]. Several studies have shown that it enhances direct learning experiences and helps in memory retention [29, 30]. It increases the engagement and motivation of the learner [24], and improves the understanding of spatial and visual concepts [17]. VR also facilitates better decision-making in simulations [30], although its effectiveness depends on the application of sound pedagogical methods [12].

Furthermore, the adaptability and acceptability of VR, particularly in specialized training scenarios, are noteworthy [13]. Furthermore, the efficacy of VR in job interview training for individuals with serious mental illness underscores its wide applicability in various educational and training contexts [41].

The comparative analysis between 2D and 3D VR environments further emphasizes these benefits. Research indicates that the 3D environment generates stronger emotional reactions than the 2D interfaces when the same content is presented [19, 34]. This increased participation in VR is correlated with its proven efficacy in enhancing educational outcomes, as evidenced in studies focusing on vocabulary acquisition and memory retention [20, 21]. However, Madden et al. [22] did not identify significant differences in learning about moon phases, indicating that the effectiveness of VR may depend on the specific nature of the task.

Current research trends in this domain include the use of electroencephalogram (EEG) measurements to gain more profound insights. These studies reveal a greater sense of presence in 3D environments and a reduced cognitive load in comparison to 2D tasks [35]. Notably, cognitive load appears higher in 2D tasks, suggesting that 3D technologies might provide cognitive benefits in learning contexts due to their reduced cognitive demands [10]. In support of this notion, Tian et al. [38] experienced that VR's stereoscopic vision can lead to increased emotional arousal, further differentiating 3D experiences from 2D ones.

This study aims to obtain feedback on the effectiveness of our tools and to gain a deeper understanding of the experiences and perceptions of the interviewers associated with child abuse.

3 SYSTEM ARCHITECTURE

The architecture of our interactive child avatar system, which supports both 2D and 3D simulation environments, is shown in Figure 1. The system is structured into three main components: i) the language component, which features large language models (LLMs) based on OpenAI's GPT-3 [6]; ii) the speech synthesis component leveraging IBM Watson services for effective speech-to-text (STT) and text-to-speech (TTS) capabilities; and iii) a Unity-based user interface supporting two distinct interactive modalities: a 3D virtual environment using the Oculus Quest2 Head Mounted Display (HMD), and a 2D virtual environment.

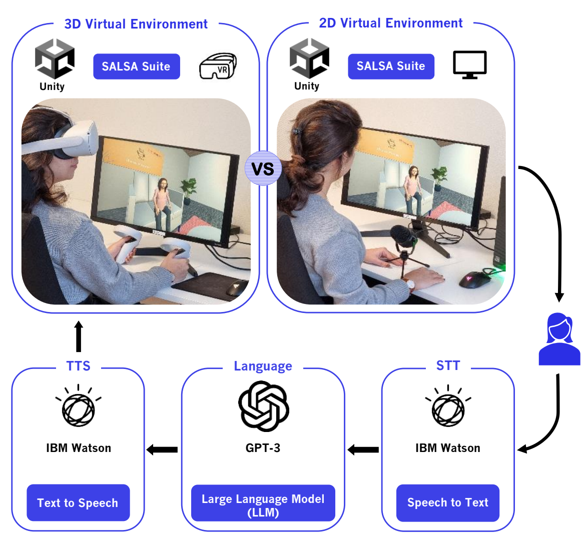

Figure 1: Architecture of the child avatar system for the comparative study.

3.1 Language

The model's language component is designed to process and respond to interviewers' inquiries, emulating a child's conversational patterns. This is based on a dataset of interview transcripts from the Centre for Investigative Interviewing at Griffith University, Australia [28]. The data set includes mock interviews, conducted by trained professionals, that simulate interactions between actors representing children and interviewers from Child Protection Services or law enforcement. In our earlier work [14], our dialogue model was developed using the RASA framework. Now we have fine-tuned the GPT-3 [6] Davinci model for two specific situations: sexual abuse and physical abuse. This was done using 10 simulated forensic interviews with children aged 6 to 8 years who were potential victims. The aim was to produce dynamic and appropriate responses to the questions asked by the interviewers in these contexts. The dialogue model was integrated with a Unity-based user interface using OpenAI API calls.

3.2 Speech Synthesis

IBM Watson's STT and TTS services are used to link the dialogue model with the interactive user interface. The IBM TTS API primarily offers adult voices; however, for a more realistic interaction, we altered the pitch and speed of a female voice to simulate a childlike sound. This modification was based on feedback from a pilot study that indicated that the voices of adults for the children were not well received by the participants [15, 32].

3.3 Visual Environment

This system allows users to interact with a virtual child avatar in two distinct environments. Despite the difference in environments, the system maintains a consistent dialogue model in its back-end, ensuring uniform responses from the avatar in both settings. The system's front-end, developed using the Unity game engine, offered two environments: a 3D environment accessible via an Oculus Quest 2 HMD and a 2D environment displayed on a 24-inch desktop monitor. Both 2D and 3D environments feature the same avatar, created using the Unity Multipurpose Avatar (UMA) open source project [39], which allows the customization of character meshes and textures. For realistic avatar movements, the Salsa Suit asset [40] is employed to synchronize lip, eye, and head movements with a generated voice. To enhance naturalism, prerecorded animations are used to animate the avatar's hands and neck.

4 EXPERIMENT DESIGN

This study involved two groups of participants: experienced investigative interviewers and individuals with backgrounds in psychology, criminology, or related fields but who lacked professional interviewing experience. Recruitment of participants was facilitated through the support of managerial personnel at their respective workplaces, guiding the researchers towards professionals best suited for participation. Eligibility for selection required experience in the criminal justice sector, with a preference for those who had previously worked as investigative interviewers. Out of those approached, all but a few individuals agreed to participate.

The methodology involved simulations in both 2D and 3D settings in random order. Each participant engaged in these environments for a duration ranging from five to ten minutes. Following the simulations, participants were asked to complete a questionnaire. This questionnaire focused on their experiences and observations while interacting with the child avatars in both the 2D and 3D environments. The study was conducted over six separate days without any specific arrangement to distinguish between participants with and without experience in interviewing children.

Our research has received ethical approval from the Norwegian Agency for Shared Services in Education and Research (SIKT) (project number #614272), titled "Interview training of child-welfare and law-enforcement professionals interviewing maltreated children supported via artificial avatars." Additionally, this study was approved by Griffith University, Australia (project number #2023/501), titled "Evaluation of Child Avatar Interview Simulation Learning Activity."

4.1 Participant Demographics

Our study initially included 39 individuals, but with the exclusion of three participants who only interacted in the 2D environment, the number of participants considered for the final analysis was reduced to 36. Within this group, 24 identified as female, 11 as male, and one participant preferred not to disclose gender. The age distribution revealed that most of the participants, 26 participants, fell within the 30-49 age range, 9 were over 50 years old, and one was under 29 years old. Regarding their professional background in child interviewing, 14 participants had no prior experience, 11 had more than ten years of experience, and the remaining 11 had less than ten years of experience. Regarding their exposure to VR, 16 participants had previous VR experience, while 20 had not used VR technology before.

4.2 Questionnaires

The questionnaire comprised 30 items, including 28 questions on a 5-point Likert scale and two open-ended questions. These questions were categorized into eight different evaluation keys: *presence* [16], *usability* [5], *visual Fidelity* [3, 43], *emotion* [3], *responsiveness* [33], *appropriateness* [45], *training effectiveness*, and *empathy*. Each question was designed to provide specific insights relevant to its category, ensuring comprehensive coverage of the user experience. For the full text of these questions, please refer to Figure 2.

5 RESULTS

In this section, we analyze the subjective feedback of participants for both 2D and 3D environments. The results are organized into three segments: Descriptive Analysis, Comparative Analysis, and Qualitative Feedback.

5.1 Descriptive Analysis

We began our analysis by calculating the central tendencies and dispersion of each evaluation aspect. In Figure 3, we present a visual comparison of mean scores and standard deviations between 2D and 3D environments. In the 2D environment, the *usability* scored the highest ($Mean$ = 4.09, SD = 0.67), and in the 3D environment, *presence* scored the highest ($Mean$ = 4.12, SD = 0.72).

Evaluation Aspects	NUM	Questions	Shorthand	Response Type
Presence	1	I felt engaged during the simulation.	Engagement	Likert scale
	2	I felt immersed in the computer-generated world.	Immersion	Likert scale
	3	I was able to concentrate on the simulation without being distracted by my surroundings.	Concentration	Likert scale
	4	I forgot about the real world during the interaction.	World Forgetfulness	Likert scale
Usability	5	The equipment was comfortable to use.	Equipment Comfort	Likert scale
	6	I felt comfortable interacting with the child Avatar.	Avatar Interaction Comfort	Likert scale
	7	The interface of the tool was easy to understand and use.	Interface Usability	Likert scale
	8	I did not experience technical difficulties while interacting with the child Avatar.	Technical Difficulty	Likert scale
	9	I would feel very comfortable using this tool on my own next time.	Ease of Future Use	Likert scale
Visual Fidelity	10	The appearance of the child Avatar was realistic.	Appearance Fidelity	Likert scale
	11	The virtual environment where the child Avatar was located felt real and contributed to my overall immersive experience.	Environment Fidelity	Likert scale
	12	I perceived hand-movements/gestures from the child Avatar.	Hand Movement Perception	Likert scale
	13	The quality of the child Avatar's movements was satisfactory (naturalness, realism, …).	Movement Quality	Likert scale
	14	The child Avatar's lip movements were well synchronized with the speech.	Lip Sync Accuracy	Likert scale
	15	The child Avatar's face expressions/movements felt realistic and were well synchronized with the speech.	Facial Expression Fidelity	Likert scale
	16	The overall perception was realistic and pleasant.	Overall Realism Perception	Likert scale
Emotion	17	I felt emotionally engaged during the interaction with the child Avatar.	Emotional Engagement	Likert scale
	18	I perceived emotions in the child Avatar's responses.	Emotional Response Perception	Likert scale
	19	The child Avatar's emotional reactions (e.g. body language, facial expressions and behaviour) looked realistic.	Emotional Reaction Realism	Likert scale
	20	The child Avatar's emotional reactions (e.g. body language, facial expressions and behaviour) consistently matched the content of the interview.	Emotion-Content Match	Likert scale
Responsiveness	21	The responsiveness of the system to my inputs felt right, natural and smooth (e.g. the system's reaction time, the consequent responses/actions from the child Avatar).	System Responsiveness	Likert scale
	22	I noticed a delay between my questions and the child Avatar's responses/reactions.	Response Delay Notice	Likert scale
	23	The pace was the usual for a conversation with a child in such circumstances.	Conversation Pace Normalcy	Likert scale
Appropriateness	24	The child Avatar's responses felt age appropriate.	Age-Appropriate Response	Likert scale
	25	The child Avatar's responses were consistent with respect to the general story.	Story Consistency	Likert scale
	26	The child Avatar's responses were appropriate and on-topic with my questions.	Response Relevance	Likert scale
Training Effectiveness	27	From a learning perspective, my interaction with the child Avatar felt as effective as interacting with a human actor/trainer.	Training Comparability	Likert scale
	28	I think this tool should be included in investigative interviewing training programs.	Tool Inclusion Recommendation	Likert scale
Empathy	29	Please provide one or more examples of the aspects of the child Avatar that felt particularly effective in eliciting your empathy and understanding.	Effective Empathy Elicitation Examples	Open-ended
	30	Please provide one or more examples of the aspects of the child Avatar that felt particularly ineffective in eliciting your empathy and understanding.	Ineffective Empathy Elicitation Examples	Open-ended

Figure 2: Categorization of questionnaire items according to Evaluation Aspects.

Conversely, the *responsiveness* scored the lowest in both the 2D and 3D environments, indicating concerns regarding interaction delays. Specifically, the mean score was 2.61 ($SD = 0.61$) in the 2D environment and 2.75 ($SD = 0.71$) in the 3D environment.

5.2 Comparative Analysis

In this section, we investigate the differential impact of the 2D versus 3D environment on QoE. This investigation utilized a paired sample T-test to analyze the variations between the two environments. Additionally, we calculated Cohen's d to assess the effect sizes. The results of these analyses are presented in Table 1.

In terms of *presence*, the results reveal that the 3D environment enhances the sense of presence more significantly than a 2D environment due to its sensory-rich and interactive dynamics. Specifically, participants experienced a higher sense of 'immersion', 'engagement' and 'world forgetfulness' in a 3D environment. However, 'concentration' levels did not show a significant difference, suggesting a complex relationship between immersive features and user focus. It is possible that the use of HMDs captures and diverts the participant's attention, thus impacting their concentration levels.

From the perspective of *usability*, this study claims that due to the absence of physically demanding equipment such as HMD, the 2D environment offers a more user-friendly experience. Specifically, participants experienced higher 'equipment comfort' in the 2D environment, suggesting that the absence of additional gear may contribute to increased comfort. Similarly, 'ease of future use' was

Table 1: Paired Sample T-Test and Cohen's d Effect Size for Learning Experience Variables. The gray areas show significant differences noted at p < .05.

	T-Statistic	P-Value	Cohen's d Effect Size
Engagement	2.798	0.008	0.466
Immersion	5.940	0.000	0.990
Concentration	2.023	0.050	0.337
World Forgetfulness	2.704	0.010	0.450
Equipment Comfort	-4.159	0.000	-0.693
Avatar Interaction Comfort	1.000	0.324	0.166
Interface Usability	-0.867	0.391	-0.144
Technical Difficulty	0.000	1.000	0.000
Ease of Future Use	-2.256	0.030	-0.376
Appearance Fidelity	1.190	0.242	0.198
Environment Fidelity	3.944	0.000	0.657
Hand Movement Perception	2.841	0.007	0.473
Movement Quality	1.847	0.073	0.307
Lip Sync Accuracy	0.849	0.401	0.141
Facial Expression Fidelity	2.142	0.039	0.357
Overall Realism Perception	1.661	0.105	0.276
Emotional Engagement	1.454	0.154	0.242
Emotion Response Perception	1.000	0.324	0.166
Emotional Reaction Realism	1.357	0.183	0.226
Emotion-Content Match	0.452	0.653	0.075
System Responsiveness	0.122	0.903	0.020
Response Delay Notice	1.183	0.244	0.197
Conversation Pace Normalcy	1.540	0.132	0.256
Age-Appropriate Response	1.152	0.257	0.192
Story Consistency	0.273	0.785	0.045
Response Relevance	-1.177	0.246	-0.196
Training Comparability	3.365	0.001	0.560
Tool Inclusion Recommendation	1.715	0.095	0.285

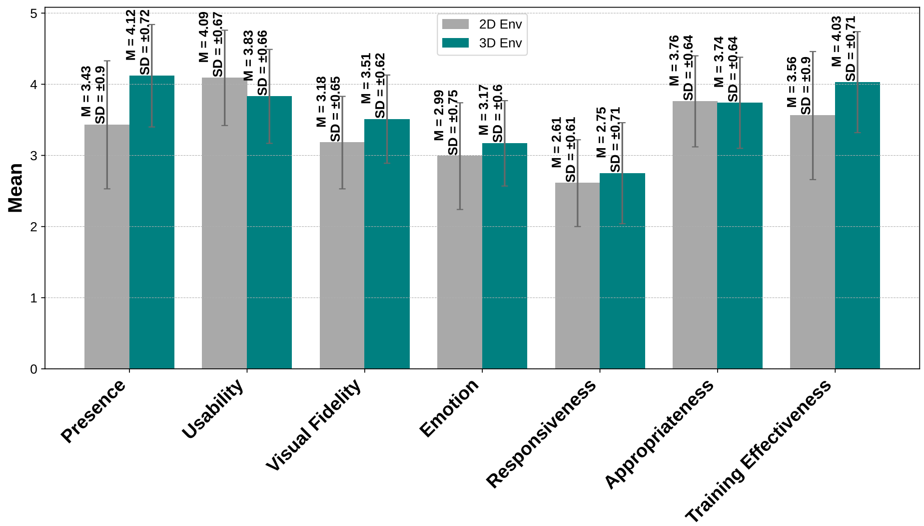

Figure 3: Bar-plot (95% confidence interval) of mean scores across the evaluative aspects in 2D and 3D environments.

significantly better in the 2D environment, which implies that users might prefer the 2D environment for future interactions, possibly due to its less physically exhausting nature. Conversely, 'avatar interaction comfort', 'interface usability', and 'technical difficulty' showed no significant differences, indicating that these aspects of *usability* are perceived similarly across both 2D and 3D environments.

Regarding *visual fidelity*, the study examined several key aspects, which indicated that there is an improved quality of visual experience in the 3D environment. In particular, 'environment fidelity' shows a substantial increase in 3D environment, suggesting that participants perceived a 3D environment as more lifelike. Similarly, 'hand movement perception' is significantly enhanced in 3D, indicating that the fluidity of VR technology provides a more lifelike representation of hand movement. Further, 'facial expression fidelity' is rated significantly higher in the 3D environment, reflecting that the immersive nature of the 3D environment may inherently strengthen the user's perception of facial expression naturalness.

However, 'movement quality', while not reaching statistical significance, shows a trend towards enhanced perception in 3D, implying a possible positive impact of VR technology on the naturalness and consistency of avatar movement. The 'overall realism perception' also trended towards a better experience in a 3D environment, but did not reach conventional significance levels. In addition, according to our expectations, there are no significant differences between 2D and 3D environments regarding 'avatar appearance fidelity' and 'lip sync accuracy'.

In terms of *emotion*, the study investigated how users perceive emotional dynamics with an avatar lacking emotional expressions in 2D and 3D environments. Contrary to the belief that a 3D immersive environment may compensate for the absence of emotional expressions [23, 37], the results indicate that there are no statistically significant differences between the 2D and 3D environments. This finding highlights the importance of explicit emotional signals in improving emotional dynamics during avatar-user interactions. However, a trend is observed to favor 'emotional engagement' and 'emotional reaction realism' in the 3D environment, although this trend is not statistically significant, which implies that immersion provided by the 3D environment leads the user to perceive certain behaviors and facial expressions of the avatar as an emotional reaction.

Regarding two conceptually close aspects, *responsiveness* and *appropriateness*, the results show that there is no significant difference in user experience between the 2D and 3D environments.

Furthermore, from the perspective of *training effectivenes*, we examined participants' perceptions of the training with the child avatar relative to human actors ('training comparability') and their approval of incorporating this tool into investigative interview training programs ('tool inclusion recommendation'). The results demonstrate a statistically significant preference for 3D over 2D in terms of 'training comparability', suggesting that participants found the 3D interaction to be nearly as effective as a real-life training experience. Additionally, there is a trend of preferring 'tool inclusion recommendation' in the 3D environment, although not statistically significant, suggesting that immersion provided by the 3D environment enhances the effectiveness of training by simulating real-world interactions more closely than in the 2D environment.

5.3 Qualitative Feedback

In the open-ended section of the questionnaire survey, participants were asked to clarify the aspects that contributed to the effective and ineffective elicitation of empathy within two simulated environments, as presented in Table 2. This table quantifies the feedback, offering a clear count of positive and negative comments for each evaluated aspect of the avatar's performance.

Notably, the 3D environment was dependent on its *visual fidelity*, with participants expressing mixed reviews, one stating: *'The hand*

Table 2: Distribution of Participant Comments on Effective and Ineffective Aspects of Empathy Elicitation in 2D and 3D Environments.

	Effective Aspects		Ineffective Aspect	
	2D	3D	2D	3D
Presence	0	1	2	0
Visual Fidelity	8	19	9	15
Emotion	2	0	4	5
Voice	6	7	4	5
Responsiveness	16	10	13	9
Response Speed	0	0	7	5

movements and eye contact were great!' In contrast, another participant stated: *'The avatar's body was not that of a child but more of a woman, which was distracting'*, underscoring the importance of appropriate visual cues in engendering empathy.

On the contrary, the participant responses indicated that the efficacy of the 2D environment was largely dependent on the avatar's *responsiveness*. One participant highlighted this as a strength, stating, *'When she spoke of harm perpetrated by her father, that tugged the heartstrings.'* Nonetheless, certain comments, like the one regarding the avatar's limited response of *'YEAH'*, underscore the necessity for enhanced emotional expressiveness and sophisticated interaction within virtual training platforms.

Furthermore, the participant feedback in both environments revealed aspects that hindered empathy engagement, indicating potential areas for improvement. A repeated critique related to the delay in the avatar's reactions; as one respondent explicitly observed, *'There was a large delay in her responses that made it hard to engage with the avatar and affected rapport building significantly.'* Emotional reaction was another aspect that received critical attention, with comments such as *'No emotion shown when disclosing sexual abuse'*, implying that more emotional expression could be crucial in augmenting the avatar's realism.

6 DISCUSSION AND FUTURE WORK

We explored the comparative effectiveness of 2D and 3D training environments in the context of CPS interview training. Differing from our earlier research [14], where VR's novelty significantly influenced participants, the current study provided a more comprehensive understanding of VR's practical utility in training scenarios. A noteworthy finding is the pronounced preference for the 3D environment in most of the questionnaire responses, suggesting that immersive experiences are important to further investigate in relation to improving training efficacy. This aligns with previous research suggesting that immersive learning contexts can substantially improve skill acquisition by providing more realistic and engaging experiences, as they closely mimic real-world interactions and scenarios [17, 20, 24]. This is demonstrated by a significant difference in 'facial expression fidelity', a crucial aspect of non-verbal communication, even without explicit emotion settings in the system. Participant feedback such as *'In the 2D version non-verbal behavior of the avatar was more difficult to perceive than in the VR version of the program'* supports this observation. Recognizing the importance of emotional dynamics in virtual environments, future iterations of the system will integrate Unreal Engine technology with NVIDIA Omniverse[1] to enhance emotional reactions. This integration is expected to result in a more sophisticated simulation of human-like emotional responses, thereby enriching the user's immersive experience by providing a nuanced and context-sensitive rendering of non-verbal cues. These advancements will likely contribute to the development of more empathetic and interactive virtual environments.

Another significant outcome is the advantage of visual interaction over text-based methods in developing effective communication skills. User comments like *'making eye contact but then also looking away when talking about negative moments (assault) seemed like a normal response in a human interview'* highlight the system's potential in simulating realistic human interactions. Interestingly, the immersive nature of the visual environment sometimes leads to overlooked responses, highlighting its realism. This was demonstrated in some participant's observations. For instance, one noted, *'When the child paused and then said the age of her brother very quietly. It seemed like a realistic moment'*, suggesting that in the visual environment, users think the delay is part of the child's hesitation. However, there is still a concern raised regarding the delay observed in the avatar's responses, primarily due to the current necessity for manual input to trigger these responses. Recognizing the criticality of fluid interaction in the context of child investigative interview interviews, future iterations will focus on automating the response mechanism, thereby eliminating the need for manual intervention.

Despite these advantages in VR technology, no significant differences were observed between the 2D and 3D environments regarding the 'tool inclusion recommendation'. This lack of distinction could be partly attributed to the discomfort associated with using HMDs, as indicated by participant comments such as *'Using headphones did not cause me pain, (...) with the VR I was in pain.'*

Additionally, it should be noted that the 3D environment requires additional equipment and a special setup, which is intrusive and uncomfortable for some individuals. In contrast, the 2D environment offers greater ease of use, which makes it ideal for 'anytime, anywhere' training sessions, making it a more accessible option for a broader audience.

7 CONCLUSION

This paper has presented a comprehensive analysis contrasting 2D and 3D virtual environments in their application to child interview training for CPS and law enforcement personnel. The findings reveal that the immersive 3D environment significantly enhanced aspects such as *presence*, *visual fidelity*, and *training effectiveness*, providing a more realistic and engaging interaction that greatly benefits training scenarios. Conversely, the 2D environment, while less immersive, was favored for its higher *usability* due to its fewer equipment requirements. However, this research also identifies critical challenges within these virtual training environments, particularly the need for emotional expressiveness and improved system responsiveness. Based on these findings, it is evident that future developments in virtual training tools must adopt a balanced approach. Prioritizing the refinement of the 3D environment is crucial, with a specific focus on integrating emotional expressiveness in avatars and improving avatar response speed. These improvements are essential for increasing the efficacy and accessibility of these virtual training environments for CPS and law enforcement personnel.

[1]https://www.nvidia.com/en-us/omniverse/apps/audio2face/

REFERENCES

[1] Gunn Astrid Baugerud, Miriam S Johnson, Helle BG Hansen, Svein Magnussen, and Michael E Lamb. 2020. Forensic interviews with preschool children: An analysis of extended interviews in Norway (2015–2017). *Applied cognitive psychology* 34, 3 (2020), 654–663.

[2] Gunn Astrid Baugerud, Ragnhild Klingenberg Røed, Helle BG Hansen, Julie Schøning Poulsen, and Miriam S Johnson. 2023. Evaluating child interviews conducted by child protective services workers and police investigators. *The British Journal of Social Work* (2023), bcac245.

[3] Frank Biocca and Mark R Levy. 2013. *Communication in the age of virtual reality.* Routledge.

[4] C.J. Brainerd and V.F. Reyna. 2012. Reliability of children's testimony in the era of developmental reversals. *Developmental review* 32, 3 (2012), 224–267.

[5] John Brooke. 1996. Sus: a "quick and dirty'usability. *Usability evaluation in industry* 189, 3 (1996), 189–194.

[6] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.

[7] Sonja P. Brubacher, Mairi S. Benson, Martine B. Powell, Jane Goodman-Delahunty, and Nina J. Westera. 2020. Chapter Twenty-Two - An overview of best practice investigative interviewing of child witnesses of sexual assault. In *Child Sexual Abuse*, India Bryce and Wayne Petherick (Eds.). Academic Press, 445–466.

[8] Ann-Christin Cederborg, Yael Orbach, Kathleen J Sternberg, and Michael E Lamb. 2000. Investigative interviews of child witnesses in Sweden. *Child abuse & neglect* 24, 10 (2000), 1355–1361.

[9] Kevin Charles Dalli. 2021. Technological Acceptance of an Avatar Based Interview Training Application: The development and technological acceptance study of the AvBIT application. *Digitala Vetenskapliga Arkivet* (2021).

[10] Alex Dan and Miriam Reiner. 2017. EEG-based cognitive load of processing events in 3D virtual worlds is lower than processing events in 2D displays. *International Journal of Psychophysiology* 122 (2017), 75–84.

[11] Mark D. Everson and Jose Miguel Sandoval. 2011. Forensic child sexual abuse evaluations: Assessing subjectivity and bias in professional judgements. *Child Abuse & Neglect* 35, 4 (2011), 287–298.

[12] Chris Fowler. 2015. Virtual reality and learning: Where is the pedagogy? *British journal of educational technology* 46, 2 (2015), 412–422.

[13] Stephanie G Fussell and Dothang Truong. 2023. Accepting virtual reality for dynamic learning: An extension of the technology acceptance model. *Interactive Learning Environments* 31, 9 (2023), 5442–5459.

[14] Syed Zohaib Hassan, Saeed Shafiee Sabet, Michael Alexander Riegler, Gunn Astrid Baugerud, Hayley Ko, Pegah Salehi, Ragnhild Klingenberg Røed, Miriam Johnson, and Pål Halvorsen. 2023. Enhancing investigative interview training using a child avatar system: a comparative study of interactive environments. *Scientific Reports* 13, 1 (2023), 20403.

[15] Syed Zohaib Hassan, Pegah Salehi, Ragnhild Klingenberg Røed, Pål Halvorsen, Gunn Astrid Baugerud, Miriam Sinkerud Johnson, Pierre Lison, Michael Riegler, Michael E. Lamb, Carsten Griwodz, and Saeed Shafiee Sabet. 2022. Towards an AI-Driven Talking Avatar in Virtual Reality for Investigative Interviews of Children. In *Proceedings of the ACM Workshop on Games Systems (GameSys)* (Athlone, Ireland). 9–15.

[16] Igroup Presence Questionnaire (IPQ) Overview. . A Multi-Disciplinary Project Consortium Addressing New Interfaces Between Humans and the Real and Virtual Environment. Available at: http://www.igroup.org/pq/ipq/index.php.

[17] Lasse Jensen and Flemming Konradsen. 2018. A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies* 23 (2018), 1515–1529.

[18] Miriam Johnson, Svein Magnussen, Christian Thoresen, Kyrre Lønnum, Lisa Victoria Burrell, and Annika Melinder. 2015. Best practice recommendations still fail to result in action: A national 10-year follow-up study of investigative interviews in CSA cases. *Applied Cognitive Psychology* 29, 5 (2015), 661–668.

[19] Maria Kandaurova and Seung Hwan Mark Lee. 2019. The effects of Virtual Reality (VR) on charitable giving: The role of empathy, guilt, responsibility, and social exclusion. *Journal of Business Research* 100 (2019), 571–580.

[20] Eric Krokos, Catherine Plaisant, and Amitabh Varshney. 2019. Virtual memory palaces: immersion aids recall. *Virtual reality* 23, 1 (2019), 1–15.

[21] Kuo-Wei Kyle Lai and Hao-Jan Howard Chen. 2021. A comparative study on the effects of a VR and PC visual novel game on vocabulary learning. *Computer Assisted Language Learning* 0, 0 (2021), 1–34.

[22] J Madden, S Pandita, JP Schuldt, B Kim, A S. Won, and NG Holmes. 2020. Ready student one: Exploring the predictors of student learning in virtual reality. *PloS one* 15, 3 (2020), e0229788.

[23] Valentina Mancuso, Francesca Bruni, Chiara Stramba-Badiale, Giuseppe Riva, Pietro Cipresso, and Elisa Pedroli. 2023. How do emotions elicited in virtual reality affect our memory? A systematic review. *Computers in Human Behavior* (2023), 107812.

[24] Thomas K Metzinger. 2018. Why is virtual reality interesting for philosophers? *Frontiers in Robotics and AI* 5 (2018), 101.

[25] Kathy A Mills and Alinta Brown. 2022. Immersive virtual reality (VR) for digital media making: transmediation is key. *Learning, Media and Technology* 47, 2 (2022), 179–200.

[26] Francesco Pompedda, Angelo Zappalà, and Pekka Santtila. 2015. Simulations of child sexual abuse interviews using avatars paired with feedback improves interview quality. *Psychology, Crime & Law* 21, 1 (2015), 28–52.

[27] Martine B Powell, Ronald P Fisher, and Carolyn H Hughes-Scholes. 2008. The effect of intra-versus post-interview feedback during simulated practice interviews about child abuse. *Child Abuse & Neglect* 32, 2 (2008), 213–227.

[28] Martine B Powell, Belinda Guadagno, and Mairi Benson. 2016. Improving child investigative interviewer performance through computer-based learning activities. *Policing and Society* 26, 4 (2016), 365–374.

[29] Jaziar Radianti, Tim A Majchrzak, Jennifer Fromm, and Isabell Wohlgenannt. 2020. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education* 147 (2020), 103778.

[30] Fabin Rasheed, Prasad Onkar, and Marisha Narula. 2015. Immersive virtual reality to enhance the spatial awareness of students. In *Proceedings of the 7th Indian Conference on Human-Computer Interaction.* 154–160.

[31] Pegah Salehi, Syed Zohaib Hassan, Myrthe Lammerse, Saeed Shafiee Sabet, Ingvild Riiser, Ragnhild Klingenberg Røed, Miriam S Johnson, Vajira Thambawita, Steven A Hicks, Martine Powell, et al. 2022. Synthesizing a talking child avatar to train interviewers working with maltreated children. *Big Data and Cognitive Computing* 6, 2 (2022), 62.

[32] Pegah Salehi, Syed Zohaib Hassan, Saeed Shafiee Sabet, Gunn Astrid Baugerud, Miriam Sinkerud Johnson, Pål Halvorsen, and Michael A. Riegler. 2022. Is More Realistic Better? A Comparison of Game Engine and GAN-Based Avatars for Investigative Interviews of Children. In *Proceedings of the ACM Workshop on Intelligent Cross-Data Analysis and Retrieval (ICDAR)* (Newark, NJ, USA). 41–49.

[33] Steven Schmidt. 2022. *Assessing the quality of experience of cloud gaming services.* Springer Nature.

[34] Donghee Shin. 2018. Empathy and embodied experience in virtual environment: To what extent can virtual reality stimulate empathy and embodied experience? *Computers in human behavior* 78 (2018), 64–73.

[35] Semyon M Slobounov, William Ray, Brian Johnson, Elena Slobounov, and Karl M Newell. 2015. Modulation of cortical activity in 2D versus 3D virtual reality environments: an EEG study. *International Journal of Psychophysiology* 95, 3 (2015), 254–260.

[36] Marije Stoltenborgh, Marian J. Bakermans-Kranenburg, Marinus H. van IJzendoorn, and Lenneke R. Alink. 2013. Cultural–geographical differences in the occurrence of child physical abuse? A meta-analysis of global prevalence. *International Journal of Psychology* 48, 2 (2013), 81–94.

[37] Feng Tian, Minlei Hua, Wenrui Zhang, Yingjie Li, and Xiaoli Yang. 2021. Emotional arousal in 2D versus 3D virtual reality environments. *PloS one* 16, 9 (2021), e0256211.

[38] Feng Tian, Xuefei Wang, Wanqiu Cheng, Mingxuan Lee, and Yuanyuan Jin. 2022. A Comparative Study on the Temporal Effects of 2D and VR Emotional Arousal. *Sensors* 22, 21 (2022), 8491.

[39] UMA Steering Group. 2022. UMA Repository. https://github.com/umasteeringgroup/UMA. Online repository.

[40] Unity Asset Store. 2022. SALSA LipSync Suite. https://assetstore.unity.com/packages/tools/animation/salsa-lipsync-suite-148442. Software package.

[41] Pınar Üstel, Matthew J Smith, Shannon Blajeski, Jeffery M Johnson, Valerie G Butler, Johanna Nicolia-Adkins, Monica J Ortquist, Lisa A Razzano, and Adrienne Lapidos. 2021. Acceptability and feasibility of peer specialist-delivered virtual reality job interview training for individuals with serious mental illness: A qualitative study. *Journal of technology in human services* 39, 3 (2021), 219–231.

[42] Wendy A Walsh, Lisa M Jones, Theodore P Cross, and Tonya Lippert. 2010. Prosecuting child sexual abuse: The importance of evidence type. *Crime & Delinquency* 56, 3 (2010), 436–454.

[43] Erin Wilson, David G Hewett, Brian C Jolly, Sarah Janssens, and Michael M Beckmann. 2018. Is that realistic? The development of a realism assessment questionnaire and its application in appraising three simulators for a gynaecology procedure. *Advances in Simulation* 3, 1 (2018), 1–7.

[44] Misun Yi, Eunkyung Jo, and Michael E Lamb. 2016. Effects of the NICHD protocol training on child investigative interview quality in Korean police officers. *Journal of police and criminal psychology* 31, 2 (2016), 155–163.

[45] Qingfu Zhu, Lei Cui, Weinan Zhang, Furu Wei, and Ting Liu. 2018. Retrieval-enhanced adversarial training for neural response generation. *arXiv preprint arXiv:1809.04276* (2018).