Delivering Bad News: VR Embodiment of Self Evaluation in Medical Communication Training

1st Tara Collingwoode-Williams
SEE VR Lab, Dept. of Computing
Goldsmiths, University of London
London, United Kingdom
tc.williams@gold.ac.uk

2nd Marco Gillies
SEE VR Lab, Dept. of Computing
Goldsmiths, University of London
London, United Kingdom
m.gillies@gold.ac.uk

3rd Pratheeban Nambyiah
Great Ormond Street Hospital
London, United Kingdom
pratheeban.nambyiah@gosh.nhs.uk

4th Carly Fertleman
The Whittington Hospital
London, United Kingdom
c.fertleman@ucl.ac.uk

5th Xueni Pan
SEE VR Lab, Dept. of Computing
Goldsmiths, University of London
London, United Kingdom
x.pan@gold.ac.uk

***Abstract*—Medical Professionals are often put through situations where they have to communicate something undesirable to patients (or their family members). The emotional reactions from the patient, or, the parent of the patient in a children's hospital, are not always easy to deal with. Here, using Virtual Reality (VR), we developed a new method to train medical professionals to be more prepared for these encounters, through a simulation with virtual parents, and afterwards, a reflection of their performance from the perspective of the virtual parent. This paper presents the technical setups of *Embodied Perspective-Taking* in a medical communication training application. Embodying a Doctor avatar, participants delivered bad news to an angry (virtual) parent that her child's surgery had unfortunately been cancelled due to an emergency. They were then able to review their performance in VR, from either the parent's (1st person) or a dis-embodied 3rd person's perspective on self-efficacy. We conducted a between-group study with 16 medical professionals recruited from a paediatric hospital and found no significant impact of perspective. However, when taking into consideration the participants' roles and experience, we found that Nurses underwent a notable change in self-evaluation compared to Doctors after experiencing their review in VR, and reported more levels of Nervousness speaking to the parent. Results suggest our approach may benefit early-stage or less confident practitioners.**

***Index Terms*—Human-Computer Interaction, Virtual Perspective Taking, Virtual Reality, Medical Communication Training**

I. Introduction

Delivering bad (undesirable) news to patients or parents of patients in healthcare can be difficult. A typical illustrative scenario might be as follows: a healthcare professional is asked to inform a patient or responsible adult that scheduled elective surgery is being cancelled due to an emergency that takes priority. The literature details numerous ways to structure a Breaking Bad News (BBN) consultation [1], but the evidence does not indisputably support one particular model over another, particularly within realistic time and resource constraints. Using standardised scenarios with actors role-playing patients is a successful way of training medical students [2]. However, this method is resource-intensive and time-consuming to facilitate and is not easy to replicate at scale. In contrast, though VR methods have a high upfront cost, the running costs are lower.

In this study, we used Embodied Virtual Perspective Taking (EVPT) to propose a template framework for self-assessment of communication skills for medical practitioners in breaking bad news. Using high-fidelity avatars animated with mocap data from paid professional actors and a realistic virtual hospital environment, we simulated a scenario where healthcare workers can practice their responses against an interactive virtual character whilst providing an assessment of VR for non-technical skills training.

II. Background

A. Virtual VS Standardised Patients in Medical Training

It has been argued that using virtual characters can elicit the same information from medical students as a real human. Lok et al, [3] created an interactive virtual clinical scenario of a virtual patient with acute abdominal pain – depicted by a life-sized projection on the wall of an exam room in a medical centre. The participant was tasked to act as a Doctor and evaluate the patient's condition by asking questions, providing a cost-effective and objective way to practice communication skills. Results suggested that, in comparison with using a standardized patient approach, participants elicited the same information from both virtual and standardized patients and performed equally well overall. O'Rourke et al., [4] investigated the emotional and behavioural impact of delivering bad news to virtual versus real standardized patients amongst a group of medical students. Results suggested that the students had similar emotional and behavioural responses when delivering bad news to a virtual simulated patient compared to

979-8-3503-8438-3/24$31.00 ©2024 Crown

a real simulated patient, with participants in both conditions performing similarly except for in tone of voice.

In this study, we used high-fidelity models and motion-captured animation to create our virtual characters so that they are believable and contribute to the transfer of skills from virtual to real life. See details of the technical pipeline in the Methodology section.

B. Embodiment and Perspective Taking

Embodiment plays an important part in building plausibility for skill transfer. *Embodiment*, in relation to virtual reality applications, can be defined as the ensemble of sensations that arise in conjunction with owning – being inside, having, and controlling – a body [5]. To foster an active construction of knowledge, junior practitioners should be encouraged to take on a self-regulating role in the learning process [3]. This approach is emphasised in many educational mission statements, as exemplified by the assertion that "the self-regulated learner must have a healthy self-concept with a strong understanding that they, alone, are in control of their learning, mastery of tasks, and attainment of goals" [6]. Techniques such as *Embodiment* and EVPT in VR could provide avenues for junior and senior practitioners in training to establish a personal agency for their growth and assessment.

EVPT can be viewed as a single or double-tiered embodied experience within the role that is usually observed. This would result in an experience where participants would embody an initial role A for a time, then they would switch to another perspective, role B, to either continue the experience [7] or re-watch the experience [8]. Examining different viewpoints or immersing oneself in another person's perspective has been associated with various positive outcomes. [9] connected it with moral development, while [10] linked it to increased empathy and altruism, as well as enhanced prosocial behaviour [11]. Researchers have also shown that adopting different perspectives can also play a role in reducing biases in social thinking and mitigating intergroup conflicts [12].

Furthermore, recent studies have shown that EVPT could reduce negative stereotyping [13] and increase cognitive empathy in immersive VR [14]. In a study in which researchers compared online and virtual reality perspectives for gender bias in STEM hiring decisions, results depicted EVPT resulted in significant changes to participant behaviour following exposure to a gender-incongruent avatar such that men showed a preference for the female candidate, and women showed a preference for the male candidate [15]. However, can this method also manipulate self-efficacy of performance in a communication skills simulation?

C. Embodiment and Perspective-Taking in Medical Training

EVPT training for communication skills in Medical training is limited. [16] looked at two patient-embodied VR experiences from a first-person patient perspective, deploying both negative or positive communication styles during a pre-operation consultation and induction of anaesthesia. Ten anaesthesiologists experienced both conditions, and a semi-structured interview followed each experience. Interviews revealed acknowledgement of the importance of good communication skills and highlighted that embodied experiences of patients in VR can influence beliefs and values on preoperative anxiety and its reduction. This experiment, however, did not first include an initial embodiment of the anaesthesiologist as it was done using 360 Video VR. We propose that adding this condition first will help to facilitate a robust tool for self-evaluation and assessment.

Other literature suggests that EVPT elicits reflection on the perspectives of others [8]. During the EVPT in this study, participants reflected on their use of empathy and perspective-taking – they took part in a medical interview where medical students would converse with a patient and then relive the conversation as the patient. Results suggested a decrease in participants' self-ratings of perspective-taking and empathy between their first and second exposures to the interview in VR.

As per the findings of Gorisse *et al.*, [17], a first-person perspective is ideally suited for tasks requiring intensive interaction, while a third-person perspective offers better spatial awareness and environmental perception, potentially extending to an understanding of how other individuals are engaging within the environment. This might explain why the conventional approach to self-evaluation post-training tends to favour the third-person perspective, as depicted by Pan et al., [18]. However, we posit a hypothesis that considering the evidence suggesting that a first-person perspective can influence empathy and behaviour, there may be notable disparities in participants' self-efficacy regarding their performance.

D. Assessing Communication Skills in Medical Training in VR

VR is also being used as a way of assessment for medical training in communication skills. This has the advantage of providing replicable conditions for self-reflection and evaluation. In [19], participants were instructed to deliver bad news to a standardized female avatar in a 3D simulated clinic. The trainee then evaluated their self-efficacy via an effective competency score (ASC) before and after the experience. Results showed that the participants' ASC scores increased overall; however, they mentioned the lack of nonverbal behaviour impeded realism. Similarly, more recently, in Ochs et al., [20] experiment, they looked at comparing the impact of virtual environment displays on the sense of presence and evaluating the system as a means of self-report. Their research results favoured Head Mounted Displays (HMD) and Cave Automated Virtual Environments (CAVE), producing a higher presence; however, their results on self-assessment are yet to be published.

Though there isn't much in the literature on the potential impact of EVPT on self-evaluation, there is enough empirical evidence to make some inferences about the possibilities. In a study that looked at an EVPT paradigm for self-counselling, participants were asked to engage in a conversation with Sigmund Freud embodied in an avatar that looks like themselves.

Then, they were asked to respond to themselves from the embodied perspective of Sigmund Freud. Results suggested that "this form of embodied perspective taking can lead to sufficient detachment from habitual ways of thinking about personal problems to improve the outcome"[21]. Participants also recorded that they felt their mood improved overall. These results demonstrate the power of EVPT to effect cognitive changes. Therefore, we believe that participants in our study watching their performance from an embodied perspective will express more of a cognitive impact, i.e. in this case, due to the nature of the task, more criticism of their performance.

Healthcare professionals engage with patients from a wide range of cultural backgrounds, each with unique expectations and coping strategies. They must have a secure environment to thoroughly assess their communication skills without fear of judgement or bias. VR proves highly effective in providing such a platform. Hence, we have selected this domain as the testbed for our research.

III. Methodology

Fig. 1. Figure shows consultation with Emily in Virtual Reality. The (male) participant is Embodied in the avatar of a Male Doctor.

A. Script and Scenario

The script was written with the support of parent feedback and two external medical practitioners. This ensured that the scenario and dialogue were plausible, felt familiar to the participants, and provoked empathy. Since empathy creates an isomorphic response to another person's feelings, an empathetic response to the distress of others can cause overwhelming distress in the observer and can lead to "an egoistic motivation to reduce stress by withdrawing from the stressor" [22] and therefore lead to social avoidance. However, moderate levels of distress may be necessary to drive one to feel empathetic concern, which is the desire for the well-being of others and, therefore, the desire to help. Due to this concern, the dialogue was first tested in a pilot with an actress and voluntary staff at the Great Ormond Hospital to gauge the response provoked by the scenario.

In our scenario, a virtual character, Emily, is the parent of a five years old young boy, Sam, who has been scheduled to have a routine surgery for a "PICC" line insertion. PICC stands for "peripherally inserted central catheter", and a PICC line is a long, thin tube inserted through a vein in the arm and passed through to the larger veins near the patient's heart. Due to primary immunodeficiency, Sam needs long-term immunoglobulins therapy and thus has been scheduled for this procedure. In order to be performed on for this surgery for minors, a general anaesthetic is needed. Thus the background of our scenario is that Emily is waiting with Sam, who has been sedated and ready for this routine procedure. However, due to an emergency (e.g., another child needing an unexpected and complex surgery for something life-threatening), the doctor scheduled for this procedure is, very unfortunately, no longer available. Therefore, someone will need to inform Emily about this and reschedule the appointment. As most hospitals are often understaffed, when there is an emergency, the responsibility to explain this to the parent may fall on some junior members of staff. This is not an uncommon scenario at hospitals and is just one example of many which could have been used for this framework.

Our script is a dialogue between Emily (virtual character) and the medical staff (participant). Although it is not possible to anticipate exactly how the participant would respond towards our virtual character, based on experience, it was decided that Emily should go through three stages of emotions which we used to structure our script development:

- Stage 1: Anticipation for procedure + Break news - uncertainty/denial
- Stage 2: Anger Escalation
- Stage 3: Shut down – Parent accepts rescheduling reluctantly

TABLE I
Dialogue stages with examples from Script

Stage	Example Dialogue from Emily
Stage 1:	"What? No, we were told by Dr Lacey that we would definitely have the procedure today."
Stage 2:	"Don't tell me to calm down. This is your fault!"
Stage 3:	"Fine. Okay. Can he keep the line in overnight?"

The dialogue needed to follow a uniform linear format so that participants experienced Emily in the same way and received the same distress cues (see Table I). This also ensured that the focus of whether the consultation went smoothly was not reliant on the reaction from Emily but on their confidence in their behaviour and training. The feedback from this pilot provided great insight into ways the dialogue can be amended to make Emily seem realistic.

B. Implementation

A testbed motion capture session was first held with researchers in our lab as actors. This was to block the recording segments and synchronisation of media captures, and also test our script with medical Doctors, to ensure that we have included enough responses to create a plausible conversation.

Technically, to animate Emily, we would need to collect Facial Tracking data, Motion Tracking data and Audio data. Motion capture was done using the Opti-Track Motive system with a 12-camera sensor setup. Facial animation was captured using the Live Capture Package and the Apple ARkit XR Plugin simultaneously with body-tracking. A mobile stand was used to hold an iPhone for face-tracking in position during the motion capture session. This worked well as the researcher was seated, as was Emily, during the experience. The audio was collected using a wireless microphone attached to the researcher and recorded with Audacity software from a separate Desktop. These captures were used as placeholders during the development of the testbed.

The consultation environment in Figure.1 was created in Unity3D version 2023.1.2f1. The hospital ward environment was a modified asset purchased on the asset store, and the user was immersed in virtual reality using the Meta Quest Pro. The virtual characters were sourced from Microsoft's Rocket-box Library package of high-fidelity models. We developed a system to record the animation and audio during the consultation, and immediately replay them back inside VR.

To bring Emily to life, we used the Wizard-of-Oz method, where an experimenter selects the dialogue reaction from a UI command window of set responses. A similar method was used in communication training for General Practitioners in dealing with demanding patients [18]. Once we had a working Unity application, several trials were conducted over Zoom where an animated Emily can be seen via a shared screen (rendered from Unity on a 2D display), with the researchers triggering Emily's response in real-time, to converse with a medical Doctor. We gathered feedback and improved our script over these trial sessions. A paid professional actress was recruited for the mocap session for Emily following the same production pipeline as described above. The animation was then integrated into Unity3D to update Emily's animation State Machine.

Finally, two different versions of the experience were created for medical professionals. In both versions, inside VR, the medical professional would go through a consultation session with Emily, with the aim of informing her about the cancellation of the procedure for her son. The medical professional would be given a gender-matched virtual doctor inside the scenario, and they would be driving the body language of the virtual Doctor directly with their hands as they are holding the VR controllers. The animation of the virtual Doctor would be recorded, together with the audio. Immediately after the VR consultation, the medical professional would be able to see a replay of their consultation, either through the perspective of Emily, the virtual parent (version 1, or 1st person perspective), or from a third person perspective (version 2). We think the ability to see the replay from VR would be a beneficial way for medical professionals to reflect on their performance and develop new strategies for future communications, and that this effect is more significant with 1st person perspective version. Link to a demo video: https://youtu.be/JDA4KwJXHE4.

IV. Experimental Study

A. Participants and Procedure

We conducted a user study with a total of 16 participants. There were 9 doctors and 7 nurses recruited from Great Ormond Street Hospital.

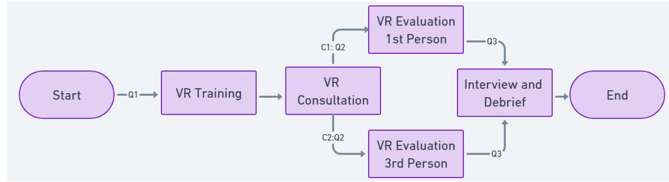

Fig. 2. Table of Experiment Procedure. Note: Q1 = Demographic Questionnaire, Q2 = Post VRConsult Questionnaire, Q3 = Post VREval Questionnaire, VR Eval 1st = VR Evaluation from Emily's Perspective, VR Eval 3rd = VR Evaluation from 3rd Person Perspective.

This is a between-group study with two conditions: after going through the same virtual consultation, the participant either went through the replay from Emily's Perspective (*1st Person*) or a third person's Perspective (*3rd Person*). Half of the participants went through each condition.

The study was approved by the University ethics board. All participants gave written consent. Upon signing the consent form, participants were asked to fill out a Demographic questionnaire and were given an information sheet that gave them context on the consultation scenario. They were then placed in a Meta Quest Pro and familiarized themselves with their gender-matched virtual body in front of a mirror. After two minutes, they were placed in a consultation room to speak to Emily, the virtual mother of a virtual patient, in VR. This consultation with Emily ranged from 4-7 minutes (*VRConsult*), depending on the participants' individual responses. Afterwards, the participants were helped out of the headset and given a Post VR questionnaire to fill out, which included a Modified Self-assessment Checklist (*VRPreCheckList*) taken from [23] and a Social presence questionnaire. Following this, the participants were helped back into the Meta Quest Pro and either placed into the control condition, *3rd Person*, where they did not have a body, or the treatment condition, *1st Person*, where they were embodied with *Agency* in Emily's body. Here, they watched the playback of their consultation. When the consultation was finished, they were prompted that the experiment had ended and they could remove their headset. This was the second exposure to stimuli (*VREval*). They were asked to complete a Post-VR Evaluation Questionnaire (*VREvalQ*). This questionnaire repeated the first set of evaluation questions, such as (*VRPostChecklist*). Upon completion, participants were interviewed, debriefed and thanked for their time.

B. Measurements

Our key measurements are the two Checklists filled by the participants after their VR consultation (*VRPreCheckList*), and after they went through the VR Evaluation replay session (*VRPosCheckList*) (see Figure 2). We took the 21-point checklist extracted from the guidelines for disclosure of adverse events developed by the CPSI and published by the CMPA. This questionnaire has been used in previous studies, [23]. We removed three questions that did not apply to this use case, leaving us with 18 Checklist options. The Checklist included statements such as, ‘Introduce oneself,’ ‘Check for understanding,’ and ‘Determine what they know.’ Participants need to go through each of the 18 items as a self-evaluation. The result will be between 0 (did none of the items on the checklist) to 18 (performed all items on the checklist). The results were then scored as percentages.

Embodiment was a factor measured by two items *Body-Ownership – Own* and *Agency – Move.* The Social Presence Questionnaire was an adaption of the questionnaire found in [24] and consisted of items *Nervous* and *Stressed* – we were interested in how they would be impacted by *Role* during the consultation as there is evidence to suggest that cognitive stress can influence performance [25].

C. Hypotheses

We think being able to see the virtual consultation replay in VR, despite the perspective, would give participants an opportunity to reflect on their performance. We measure their self-evaluation of their performance with the checklist. We think that there would be a drop in their Checklist Scores after they see the replay of their performance (**H1**).

Secondly, we think that being able to see the replay of their performance from the virtual parent’s perspective (1st person perspective) would be particularly beneficial for their self-reflection. Here, we hypothesise that the drop after condition 1 (*1st person*) will be bigger than in condition 2 (*3rd person*) (**H2**).

Medical education for Doctors often focuses on diagnostic and procedural skills, with a growing emphasis on communication skills, and not all medical schools include training on communication or empathy [26]. Research from [27] indicated that while senior hospital doctors recognise the importance of breaking bad news, they often did not pursue courses in this area. In contrast, nursing education places a strong emphasis on patient-centred care and interpersonal communication, as nurses are trained to meet patients’ physical, emotional, and social needs through communication. Nurses play a crucial role in breaking bad news by providing information, preparing patients, and offering support, though often the responsibility of delivering the news falls to physicians [28]. Therefore, our hypothesis (**H3**) posits a significant difference in Checklist Scores between roles after a VR Replay, with Nurses likely experiencing a greater decline in self-evaluation scores compared to Doctors, reflecting more critically on their performance.

V. Data Analysis

We conducted analysis in IBM SPSS Statistics 27. Independent t-tests were conducted to check if there was a significant difference between scores. In the instances where a significant difference was found in the data, which were not normally distributed or contained multiple outliners, we ran a non-parametric test (Mann-Whitney U) to validate the result.

VI. Results

A. Checklist Questionnaire

We ran a Paired T-test to check if there was a significant difference between scores between *VRPreChecklist* and *VRPostChecklist*. One outlier was detected that was more than 1.5 box lengths from the edge of the box in a boxplot. Inspection of their values did not reveal them to be extreme and they were kept in the analysis. The difference scores for *VRPreChecklist* and *VRPostChecklist* were normally distributed, as assessed by Shapiro-Wilk’s test ($p = .187$). Data are mean $\pm$ standard deviation unless otherwise stated. Participants rated themselves higher the first time they did the consultation (*VRConsult*) (68.75 ± 17.7) than the second time viewing (*VREval*) (64.59 ± 22.76) suggesting that participants initially felt they performed better before (*VREval*). There, however, was no statistically significant mean difference in results across the two ratings ($4.162(95\%CI, -2.28$ to $10.61), t(15) = 1.376, p = .189, d = 0.34$) (see Figure 3).

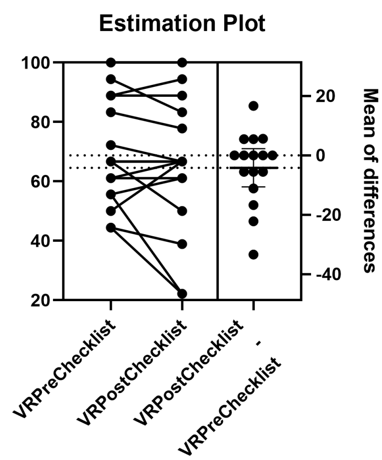

Fig. 3. Estimation plot of the Paired T-test shows the changes between VRPreChecklist and VRPostChecklist.

1) By Condition: We created a new factor *Checklist_Diff* by calculating the difference in the score (percentage) between *VRPreChecklist* and *VRPostChecklist*. Multiple outliers were detected via visual inspection of a Boxplot. Therefore, we ran a Mann-Whitney U test to determine if there were differences in *Checklist_Diff* between *Condition*. Median engagement score for *1st Person* (-2.800) and *3rd Person* (0.0) was not statistically significantly different, ($U = 37, z = .54, p = .592, d = 0.13$).

2) By Role: We checked to see if there was a significant difference using *Role* as a factor. An Independent-sample t-test was run to determine if there were differences in *Checklist_Diff* between *Role*. The difference in score slightly increased with Doctors from baseline (1.2 ± 8.25) than Nurses, which fell from baseline(-11.1 ± 13.21); overall we can see from Figure 4 that there was a significantly heavier drop in ratings from Nurses ($12.4(95\%CI, 0.83$ to $23.99), t(14) = 2.3, p = .037, d = 1.16$).

An Independent-sample t-test was run to determine if there were significant differences in *VRPostChecklist* between *Role*. There were no outliers in the data, as assessed by inspection of a Boxplot. Scores for each level of *Role* were normally distributed, as assessed by Shapiro-Wilk’s test ($p > .05$), and variances were homogeneous, as assessed by Levene's test for equality of variances ($p = .230$). Results showed Doctors rated themselves better (78.4 ± 14) than Nurses(46.8 ± 19.49) post *VREval* experience, ($31.57(95\%CI, 13.66$ to $49.49), t(14) = 3.8, p = .002, d = 1.91$). (See Figure 4 $a - c$).

3) By Experience: Data are mean $\pm$ standard deviation unless otherwise stated. There were 9 participants with over 10 years of experience and 7 with less than 10 years of experience. An independent-sample t-test was run to determine if there were differences in Checklist_Diff between experience levels. One outlier was detected in a boxplot, but their values were not extreme and they were kept in the analysis. Engagement scores for each level of gender were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p = .685$). The Checklist difference mean was more with less experienced participants than those with more ($-5.5510.15$) than those with less ($-3.077813.92$), but this was not statistically significant ($-2.48(95\%CI, -14.93$ to $10.98), t(14) = -0.36, p = .699, d = 0.020$.) (Figure $4f$).

B. Nervous and Stress

An Independent-sample t-test was run to determine if there were differences for *Nervous* between *Role*. Nurses showed higher reported levels of *Nervous* (5.86 ± 1.46) than Doctors (3.44 ± 1.74), and this difference was statistically significant ($-2.41(95\%CI, -4.17$ to $0.65), t(14) = -2.94,$ **p = .011**, $d = -1.48$). There was no significance found for *Stressed* ($-1.29(95\%CI, -3.00$ to $0.43), t(14) = -1.60, p = .131, d = -0.81$). (Figure $4d, e$).

C. Embodiment

Results showed There was no significant difference in *Own* between *Role*. ($-1.70(95\%CI, -3.50$ to $0.11), t(14) = -2.01, p = 0.063, d = -1.02$). There was no significant difference in *Move* between *Role*, ($-1.13(95\%CI, -2.71 \text{to} 0.45), t(14) = -1.53, p = 0.145, d = 0.77$).

VII. Semi-Structured Interview

After the VR experiences, we conducted a semi-structured interview with participants. (Not all practitioners could stay to finish it due to work commitments.) Below we list the key findings:

Realism and Immersion: The VR experience was praised for its realism, with participants commenting, "It's incredibly realistic. I mean, I felt like I was in there." (P5) and "It's like you're actually there. You can look around and see everything as if you're really in the hospital room." (P13.)

Self-Reflection and Learning: Participants found the self-reflection component beneficial, "Getting the opportunity to watch yourself back is amazing...it's really fascinating." (P5) and "I was quite surprised to hear how hesitant my voice sounded." (P16.)

Emotional Impact and Stress: The simulations evoked strong emotions, "It was very stressful, very stressful." (P4) and "I could feel myself shaking a bit...I was really nervous."(P7), highlighting the emotional realism of the VR scenarios.

Authenticity of Interactions: Authenticity was a noted strength, "The mom's conversation was very realistic."(P15) Others felt realism could be enhanced by more dynamic interactions, "Sam moving and interrupting would be really good."(P5.)

Utility in Professional Training and Beyond: The practical applications of VR were widely recognised, "This cancellation one is definitely a big one."(P7), "It would be really good to have for staff-to-staff conversation."(P16.)

VIII. Discussion and Conclusion

In this study, we investigated the impact of two different configurations of EVPT on self-evaluation in VR. One was in an embodied first person, and the other in a disembodied 3rd person perspective. It's important to note that the results reflect self-judgement scores that don't correlate with real performance but provide informal observations of self-efficacy. Primarily, results showed that in general there was no significant difference between the Checklist Scores given between (*VRPreChecklist*) and (*VRPostChecklist*) rejecting **H1**. This means that overall participants felt they performed similarly even after evaluation in VR.

The results suggest no difference in evaluation ratings due to perspective, rejecting **H2**, this could be due to the small sample size in each condition, or perhaps a lack of priming for re-embodiment into Emily's perspective - more time could be given to participants to familiarise themselves with the perspective change before beginning the replay. More research needs to be done on exploring configurations of EVPT for Self-Evaluation.

The results were more interesting as we look at them from the perspective of Roles: Doctors were quite confident in their overall performance and less willing to find fault in their conduct. The difference between their self-evaluation score *Checklist_Diff* was mostly close to zero, if not slightly improved. They significantly rated themselves with higher performance which is evident in Figure 4. This may be because Doctors were embodied in an avatar that reflects their role, [29] found that participants embodied in self-avatars performed better in convergent thinking tasks. Nurses, on the other hand, experienced a statistically significant drop in Checklist Scores after *VREval*, suggesting that they believed, on average, that they performed worse than what they initially expected; this rejects the null hypothesis of **H3**. This is an important result as this implies that Nurses could benefit from this format as they could be more inclined to acknowledge errors.

Data revealed that Nurses felt significantly more nervous talking to Emily than Doctors - Figure 4. Other studies have pointed out potential differences in confidence levels between Doctors and Nurses, suggesting that to gain patient confidence in their new roles, Nurses must be both confident and competent in their own abilities [30]. Additionally, results showed that though not significant, participants with

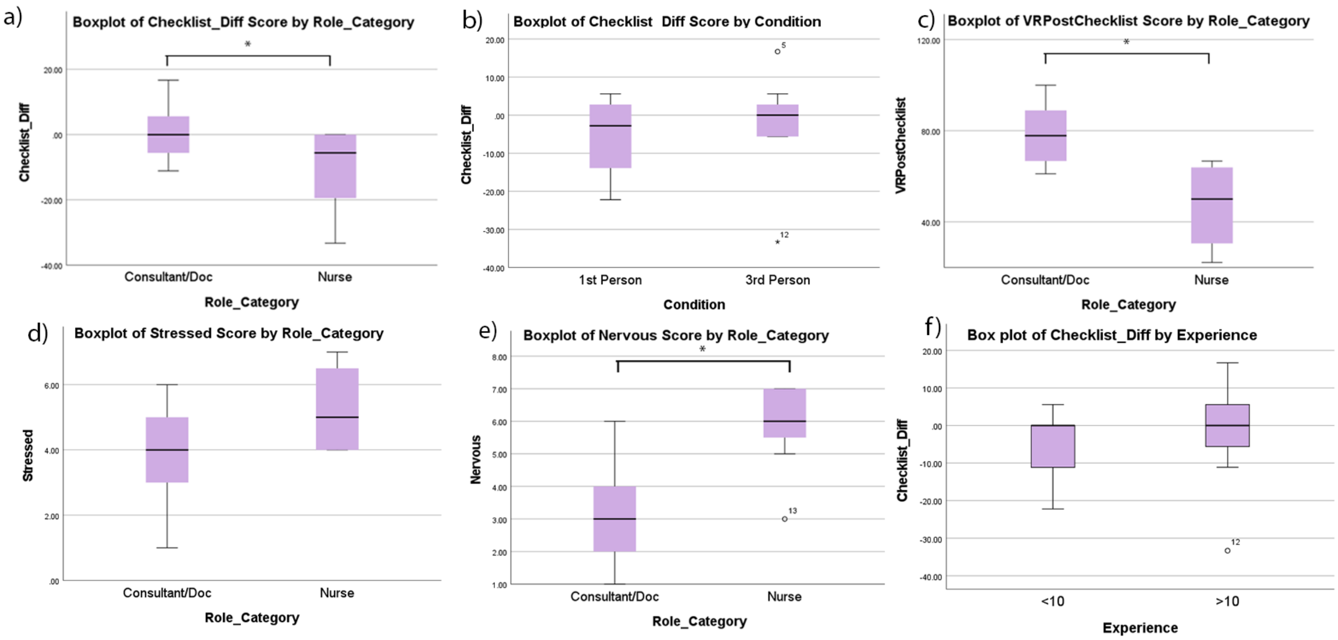

Fig. 4. a) Boxplot of Checklist_Diff Score by Role, b) Boxplot of Checklist_Diff Score by Condition, c) Boxplot of VRPostChecklist Score by Role, d) Boxplot of Stressed Score by Role, e) Boxplot of Nervous Score by Role, f) Boxplot of Checklist_Diff Score by Experience

less experience had a higher decrease in Checklist Scores after *VREval*. Perhaps also indicating interest in the impact within experience levels. More research, therefore, should be done to investigate whether iterative training with feedback can potentially provide the confidence to handle these cases regardless of role and experience.

IX. Limitations and Future Work

This pilot study proved that though there was no effect of EVPT on the evaluation process, there's potential for VR as a tool for evaluating communication skills. In future work, we aim to involve a larger sample size and enhance the accuracy of evaluation ratings by comparing them to expert ratings through collaboration with a panel of practitioners. Additionally, investigating gaze data to determine focus areas during consultations will be explored. To better understand the factors influencing these differences, further research could control for variables like gender, experience, and training within each professional group. This approach will help clarify the complex interactions between professional roles and other influencing factors in this study. More research needs to be done to perfect this framework for consideration for future use in practice.

Acknowledgment

Acknowledgements to Dr Dave Raynard for his support in managing the development of the technical aspect of the project. Fang Ma, Hankun Yu and Buko Studios for their technical support. This work is funded by internal grants from Goldsmiths College and the Great Ormond Street Hospital.

References

[1] J. H. Vermylen, G. J. Wood, E. R. Cohen, J. H. Barsuk, W. C. McGaghie, and D. B. Wayne, "Development of a simulation-based mastery learning curriculum for breaking bad news," *Journal of pain and symptom management*, vol. 57, no. 3, pp. 682–687, 2019.
[2] G. Posner and A. Nakajima, "Assessing residents' communication skills: disclosure of an adverse event to a standardized patient," *JOGC*, vol. 33, no. 3, pp. 262–268, 2011.
[3] B. C. Lok, R. E. Ferdig, A. Raij, K. Johnsen, R. Dickerson, J. Coutt, A. Stevens, and D. S. Lind, "Applying virtual reality in medical communication education: current findings and potential teaching and learning benefits of immersive virtual patients," *Virtual Reality*, vol. 10, p. 185–195, 2006.
[4] T. L. B. Sarah R. O'Rourke, Kelly R. Branford, L. T. Ives, A. Nagendran, and S. N. Compton, "The emotional and behavioral impact of delivering bad news to virtual versus real standardized patients: A pilot study," *Teaching and Learning in Medicine*, vol. 32, pp. 139–149, 2020.
[5] K. Kilteni, R. Groten, and M. Slater, "The sense of embodiment in virtual reality," *Presence: Teleoperators and Virtual Environments*, vol. 21, no. 4, pp. 373–387, 2012.
[6] S. Sandford and K. Richardson, Eds., *Interactive instructional design: old paradigms for new technologies*. Paper presented at the NAU/web.97 (June 14, Flagstaff, AZ), 1997.
[7] F. Ke and X. Xu, "Virtual reality simulation-based learn-

ing of teaching with alternative perspectives taking," *British Journal of Educational Technology*, vol. 51, no. 6, pp. 2544–2557, 2020.

[8] A. Raij, A. Kotranza, D. S. Lind, and B. Loka, "Virtual experiences for social perspective-taking," *2009 IEEE Virtual Reality Conference*, pp. 99–102, 2009.

[9] L. Kohlberg, "Moral stages and moralization: The cognitive-development approach," *Moral Development and Behavior: Theory and Research and Social Issues*, pp. 31–53, 1976.

[10] R. B. Cialdini, S. L. Brown, B. P. Lewis, C. Luce, and S. L. Neuberg, "Reinterpreting the empathy–altruism relationship: When one into one equals oneness," *Journal of Personality and Social Psychology*, vol. 73, no. 3, p. 481–494, 1997.

[11] D. Batson, "Prosocial motivation: Why do we help others?" *A. Tesser (Ed.), Advanced social psychology*, pp. 333–381, 1995.

[12] A. D. Galinsky and G. B. Moskowitz, "Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism," *Journal of Personality and Social Psychology*, vol. 78, no. 4, p. 708–724, 2000.

[13] N. Yee and J. Bailenson, "Walk a mile in digital shoes: The impact of embodied perspective-taking on the reduction of negative stereotyping in immersive virtual environments," *Proceedings of PRESENCE 2006: The 9th Annual International Workshop on Presence.*, 2006.

[14] J. Z. Austin van Loon, Jeremy Bailenson, J. Bostick, and R. Willer, "Virtual reality perspective-taking increases cognitive empathy for specific others," *PloS One*, vol. 13, no. 8, p. e0202442, 2018.

[15] C. L. Crone and R. W. Kallen, "Interview with an avatar: Comparing online and virtual reality perspective taking for gender bias in stem hiring decisions," *PLoS ON*, vol. 17, no. 6, 2022.

[16] Hoek, Krista, Velzen, Monique, and E. Sarton, "Patient-embodied virtual reality as a learning tool for therapeutic communication skills among anaesthesiologists: a phenomenological study," *Patient Education and Counseling*, p. 107789, 2023.

[17] G. Gorisse, O. Christmann, E. A. Amato, and S. Richir1, "First- and third-person perspectives in immersive virtual environments: Presence and performance analysis of embodied users," *Front. Robot. AI*, vol. 4, p. 270954, 2017.

[18] A. B. X. N. Xueni Pan, Mel Slater, A. I. B. Rivas, D. Swapp, J. Hale, P. A. G. Forbes, C. Denvir, A. F. de C. Hamilton, and S. Delacroix, "The responses of medical general practitioners to unreasonable patient demand for antibiotics - a study of medical ethics using immersive virtual reality," *PLOS ONE*, vol. 11, no. 2, 2016.

[19] K. Z. Allen D. Andrade, Anita Bagri, B. A. Roos, and J. G. Ruiz, "Avatar-mediated training in the delivery of bad news in a virtual world," *Journal of Palliative Medicine*, vol. 13, no. 12, pp. 1415–1419, 2010.

[20] G. d. M. Magalie Ochs, Daniel Mestre, J.-M. P. J. Saubesty, E. Lombardo, D. Francon, and P. Blache, "Training doctors' social skills to break bad news: Evaluation of the impact of virtual environment displays on the sense of presence," *Journal on Multimodal User Interfaces*, vol. 13, no. 1, pp. 41–51, 2019.

[21] S. A. Osimo, R. Pizarro, B. Spanlang, and M. Slater, "Conversations between self and self as sigmund freud-a virtual body ownership paradigm for self counselling," *Sci. Rep.*, vol. 5, no. 13899, 2015.

[22] ean Decety, "The Neurodevelopment of Empathy in Humans," *Developmental Neuroscience*, vol. 32, no. 4, pp. 257–267, 2010.

[23] G. Posner and A. Nakajima, "Assessing residents' communication skills: disclosure of an adverse event to a standardized patient," *Journal of Obstetrics and Gynaecology Canada*, vol. 33, no. 3, pp. 262–268, 2011.

[24] F. Biocca, C. Harms, and J. K. Burgoon, "Toward a more robust theory and measure of social presence: Review and suggested criteria," *PRESENCE: Virtual and Augmented Reality*, vol. 12, no. 5, pp. 456–480, 2003.

[25] A. A. Xueni Pan, Tara Collingwoode-Williams, H. Brenton, B. Congdon, O. Drewett, M. F. P. Gillies, D. Swapp, P. Pleasence, C. Fertleman, and S. Delacroix, "A study of professional awareness using immersive virtual reality: The responses of general practitioners to child safeguarding concerns," *Front. Robot. AI*, vol. 5, no. 80, 2018.

[26] "Teaching how to deliver bad news: a systematic review," vol. 27.

[27] M. M. Barnett, J. D. Fisher, H. Cooke, P. R. James, and J. Dale, "Breaking bad news: consultants' experience, previous education and views on educational format and timing," *Medical Education*, vol. 41, no. 10, pp. 947–956, 2007.

[28] S. A. Piironen, "Nurse's role in breaking bad news : literature review," 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:55957490

[29] G. Gorisse, S. Wellenreiter, S. Fleury, A. Lécuyer, S. Richir, and O. Christmann, "I am a genius! influence of virtually embodying leonardo da vinci on creative performance," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–11, 2023.

[30] C. Rashid, "Benefits and limitations of nurses taking on aspects of the clinical role of doctors in primary care: integrative literature review," *Journal of advanced nursing*, vol. 66, no. 8, pp. 1658–1670, 2010.