Evaluating Virtual Human Role-Players for the Practice and Development of Leadership Skills

Gonzalo Suárez*, Sungchul Jung and Robert W. Lindeman

Human Interface Technology Lab NZ, College of Engineering, University of Canterbury, Christchurch, New Zealand

This article reports on a study to evaluate the effectiveness of virtual human (VH) role-players as leadership training tools within two computer-generated environments, virtual reality (VR) and mixed reality (MR), compared to a traditional training method, real human (RH) role-players in a real-world (RW) environment. We developed an experimental training platform to assess the three conditions: RH role-players in RW (RH-RW), VH role-players in VR (VH-VR), and VH role-players in MR (VH-MR), during two practice-type opportunities, namely pre-session and post-session. We conducted a user study where 30 participants played the role of leaders in interacting with either RHs or VHs before and after receiving a leadership training session. We then investigated (1) if VH role-players were as effective as RH role-players during pre- and post-sessions, and (2) the impact that the human-type (RH, VH) in conjunction with the environment-type (RW, VR, MR) had on the outcomes. We also collected user reactions and learning data from the overall training experience. The results showed a regular increase in performance from pre- to post-sessions in all three conditions. However, we did not find a significant difference between VHs and RHs. Interestingly, the VH-MR condition had a more significant influence on performance and task engagement compared to the VH-VR and RH-RW conditions. Based on our findings, we conclude that VH role-players can be as effective as RH role-players to support the practice of leadership skills, where VH-MR could be the best method due to its effectiveness.

Keywords: virtual humans, role-playing, virtual reality, mixed reality, leadership training

1. INTRODUCTION

Leadership and communication skills are considered among the most valuable professional skills an employee can have (Conrad and Newberry, 2011). Nevertheless, despite significant annual investment in training programs, many organizations are still unsatisfied with the outcomes of their leadership training efforts (Hunt and Baruch, 2003; Kaiser and Curphy, 2013). According to Sogunro (2004), leadership skills are not so easily acquired using only theory, talk, and discussion groups. Leadership skills are commonly learned during practice and interaction with other people (Wexley and Latham, 2002). Similarly, McCauley and Van Velsor (2004) have noted that when leaders have the opportunity to practice leadership competencies, they can reflect on their own experience and engage in the learning process. Along with these statements, Weaver et al. (2010) indicate that practice-based training techniques, such as role-playing, are the most critical and effective for influencing training outcomes. Furthermore, constructivist learning theory (Piaget, 1952) also supports practice-based methods, suggesting that

learning improves when learners develop constructions of the world through direct experience, and when they can reflect on these experiences. However, role-playing also suffers from several drawbacks. First, its implementation can be expensive, so the number of role-playing activities that are frequently organized is low. Second, to be taken seriously, the development of the activities could require both specialist knowledge and professional actors' involvement. Lastly, role-playing can also be threatening to the people who do not feel comfortable talking in the presence of others (Van Ments, 1989; Sogunro, 2004; Bosman et al., 2019). In response, the rapid technological advances with virtual agents in computer-generated environments offer an opportunity to complement and support traditional training delivery methods with a practical and cost-effective approach (Guadagno et al., 2007; Kim K. et al., 2018; Bosman et al., 2019). Based on these observations, we consider that an effective leadership training solution should provide continuous practice while interacting with others, similar to the traditional role-playing, but with improvements able to overcome current limitations.

Consequently, our work aims to address the general dissatisfaction that organizations have concerning their leadership development efforts by overcoming the current role-playing limitations. Thus, given the recent advances in technology that indicate that virtual reality (VR) and mixed reality (MR) are becoming more practical solutions due to the higher levels of immersion offered in contrast to desktop platforms (Schmid Mast et al., 2018), we propose to (1) evaluate the effectiveness of virtual-human (VH) role-players as leadership training tools within VR and MR by comparing them to real-human (RH) role-players in the real world (RW); and (2) explore the impact of the human-type (RH, VH) in conjunction with the proposed environment-type (RW, VR, MR).

Moreover, since our primary focus is to investigate the effectiveness of VH role-players, and not necessarily the well-known and validated approach of RH role-players, we omit some expected human-environmental-type conditions (e.g., RH-VR, RH-MR) to reduce the complexity of our study design. Accordingly, we handled the human-type (RH, VH) and the environment-type (RW, VR, and MR) as one single independent variable called *mediation-type*. We then compared the three mediation-type conditions (RH-RW, VH-VR, and VH-MR) during two *practice-type* opportunities where participants interacted with either RHs or VHs before (pre-session) and after (post-session) receiving leadership instruction. We built both practice-type sessions with structured role-play scenarios designed with content of a time-tested leadership model, the Situational Leadership® II (SLII®)[1] model, developed by The Ken Blanchard Companies®. We illustrate the three mediation-type conditions in **Figure 1**. In the following section, we provide relevant related work. We then describe the design and implementation of our experimental training platform, followed by details of our user study. This is followed by a discussion of the results and the implications of our findings.

---

[1]SLII: https://www.kenblanchard.com/Products-Services/Situational-Leadership-II/.

2. RELATED WORK

2.1. Virtual Humans and Immersive Environments

During the early 1980s, short movies displaying high-quality animations involving realistic VHs (e.g., The Juggler from Adam Powers and Sexy Robot from Robert Abel; Magnenat-Thalmann and Thalmann, 2005) began to appear. Since then, computer graphics has increasingly evolved in computational speed and control methods, allowing the rendering of three-dimensional (3D) VHs much faster, with realistic graphics, and also suitable for real interactive applications (Badler, 1997; Waltemate et al., 2018; Bosman et al., 2019).

Generally conceived and perceived as digital representations of real humans, VHs coexist in virtual environments where they can move around and communicate with participants using social attributes like facial expressions, posture, and voice (Hartholt et al., 2019). According to Fox et al. (2015), VHs can be categorized as agents or as avatars, where *agents* are programmed and controlled by computer algorithms, with the ability to interact with real people using verbal and non-verbal behaviors (Guetterman et al., 2017). On the other hand, *avatars* are fully controlled and animated by RHs, representing a direct extension of ourselves into the virtual world (Slater et al., 1998; Kilteni et al., 2015; Jung et al., 2017; Waltemate et al., 2018).

VHs have shown to be valuable tools in different areas, including cognitive-science studies, training, education, and entertainment (Campbell et al., 2011; Hartholt et al., 2013). In clinical psychology, VHs have seen a range of beneficial uses, treating patients with depression, anxiety, or post-traumatic stress disorder (PTSD; DeVault et al., 2014; Hartholt et al., 2019). Moreover, several research efforts have identified the benefits of using VH role-players for teaching social skills in the areas of communication, leadership, and negotiation (Core et al., 2006; Johnsen et al., 2007; Campbell et al., 2011; Batrinca et al., 2013). Guetterman et al. (2017) evaluated a VH application for competence assessment in breaking bad news to a VH patient. This study provided initial evidence that VH programs can be effective for assessing communication skills in medical education. Hays et al. (2012) implemented a turn-based branching narrative system where a life-size VH, projected on a 2D screen, was compared to a RH in the ability to help Navy Officers organize and understand how to apply counseling skills. The results demonstrated that both VHs and RHs affected trainees' task engagement in similar ways. Another study by Gratch et al. (2016) found that VH role players could help make people practicing social skills feel more comfortable than human role players. Lastly, the rich configurability of VHs can also provide a robust learning experience. VHs are virtually always available for trainees to put their skills into practice, and they can also be easily customized to fit diverse scenarios and social contexts (Storey and Cox, 2015; Schmid Mast et al., 2018).

The virtual environment where VHs are deployed represents a multi-dimensional experience that can be totally or partially computer generated (Bosman et al., 2019). As introduced by Milgram and Kishino (1994), there is a continuum of possible combinations between the real and virtual environments. These

FIGURE 1 | The three conditions of our experimental leadership training platform. In a user study, we evaluated the effectiveness of Virtual Human Role-Players to support the practice of leadership skills within two computer-generated environments: **(B)** Virtual Reality and **(C)** Mixed Reality, in comparison to a traditional gold-standard method: **(A)** Real Human Role-Players.

combinations ranged from overlaying virtual objects into the real world, Augmented Reality (AR), to capturing real objects and superimposing them into the virtual space, Augmented Virtuality (AV). MR combines real and virtual worlds all along the reality-virtuality continuum, encompassing both AR and AV.

For many years, investigators have shown that the absence of many cues (e.g., posture, facial expression, orientation, dress, and physical appearance) identified in traditional computer-mediated communication affects both the capacity to transmit information and the attention of the user to focus on the presence of others (Albertson, 1980; Daft and Lengel, 1986). However, the arrival of more sophisticated computer-generated technologies, where VHs exist, can help overcome these disadvantages by enhancing how information is transmitted and perceived (Blascovich et al., 2002). In support, immersive computer-generated environments such as VR, can provide a first-person perspective where participants are fully surrounded by a 3D world (Milgram et al., 1995). Moreover, in VR, users can experience a higher level of immersion in contrast to desktop platforms (Schmid Mast et al., 2018). The sensation can be improved using multi-sensory feedback (Jung et al., 2020), and the systems can allow VHs to make users have stronger feelings of social presence (Bailenson et al., 2001).

2.2. Situational Leadership® II

The Situational Leadership® (SL®) model was initially developed by Paul Hersey and Kenneth Blanchard in the late 1960s (Blanchard et al., 2015). During the early 1980s, the SL® model was revised to address areas that needed improvement based on research and observations from clients and colleagues at Blanchard Training and Development, Inc (Blanchard et al., 1993). The result was a new generation of SL® thinking called Situational Leadership® II (SLII®). According to Blanchard et al. (2015), the main idea behind SLII® is that there is no single best style of leadership; the best style will depend on the situation or development level of the follower (a person who supports and is led, guided, or commanded by a leader). Moreover, leadership styles are driven based on the follower's level of competence and commitment regarding a specific task. Hence, the best leaders should be flexible and able to diagnose the situation before selecting the appropriate leadership style, in

order to promote performance and ensure satisfaction from their followers (Zigarmi and Peyton, 2017).

The SLII® framework proposes four leadership styles and four follower development levels (Blanchard et al., 2015). The four *Leadership Styles* consist of a combination of directive and supportive behaviors, namely:

- **Directing (S1)**: high directive and low supportive behaviors
- **Coaching (S2)**: high directive and high supportive behaviors
- **Supporting (S3)**: low directive and high supportive behaviors
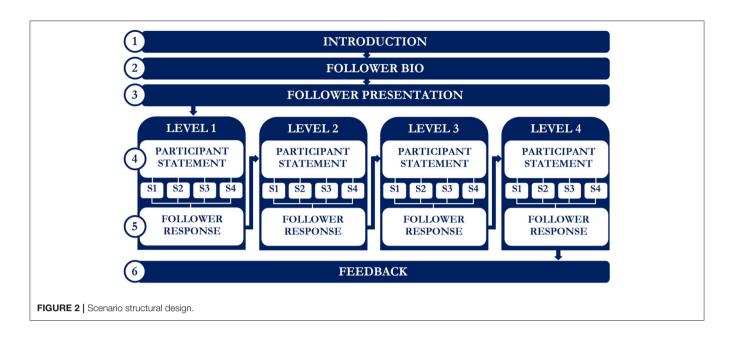- **Delegating (S4)**: low directive and low supportive behaviors

The four *Development Levels* consist of a combination of competence and commitment, namely:

- **Developing (D1)**: low competence and high commitment
- **Developing (D2)**: some to low competence and low commitment
- **Developing (D3)**: moderate to high competence and variable commitment
- **Developed (D4)**: high competence and high commitment

SLII® provides an easy-to-understand and apply approach widely used for more than thirty years for leading and developing people (Blanchard et al., 2015). Given the time-tested nature of SLII®, and the practicality offered by VHs, we explored a way of combining them to create a validated approach that would be as valuable as traditional role-playing, but with the greater configurability provided by VHs. We then tested the idea empirically to assess effectiveness.

3. METHODS

In this study, we developed an experimental training platform to provide an interactive experience with either RHs in RW or VHs in VR and MR, during two practice-type opportunities (pre- and post-session) built with a total of eight role-play scenarios. We developed each scenario with structured narratives using SLII® content with permission from The Ken Blanchard Companies Inc®. We then conducted a user study to investigate: (1) if VH role-players were as effective as RH role-players during pre- and post-sessions, and (2) the impact that the human-type (RH, VH) in conjunction with environment-type (RW, VR, MR) had on the

FIGURE 2 | Scenario structural design.

outcomes. Finally, we also collected user reactions and learning data from the overall training experience.

3.1. System Overview
3.1.1. Role-Playing Scenario Design
To be consistent with the SLII® framework, we designed eight role-playing scenarios based on the development level of the follower(s) regarding a specific task. Thus, we built the D1 scenarios for followers with low competence and high commitment in need of a directive style of leadership (S1). For the D2 scenarios, the followers had low to some competence and low commitment in need of a coaching style of leadership (S2). For the D3 scenarios, the followers had moderate to high competence and variable commitment in need of a supporting style of leadership (S3). Finally, for the D4 scenarios, the followers had high competence and high commitment in need of a delegating style of leadership (S4).

We organized the narratives of each scenario according to the structure shown in **Figure 2**. The sequence had six sections: (1) The **Introduction** described the scenario, providing context, e.g., *This is a conversation between you (Supervisor) and Alice (either RH or VH), a Team Leader who reports directly to you. You have asked Alice to organize an end-of-year party for more than forty employees within the next two months.* (2) The **Follower Bio** provided necessary information about the follower and his/her development level, e.g., *Alice is excited about the task. However, she does not have previous experience organizing parties.* (3) In the **Follower Presentation** section, the follower introduced him/herself, e.g., *Thank you for having this meeting with me. I am feeling really excited about this party.* As shown in **Figure 2**, for sections 4 and 5, we implemented a simple branching structure using a broader range of choices to deliver a more realistic storyline. We arranged the structure in four consecutive levels of progression. This is, the choices given in level 1 led to new choices in level 2, which each led to new choices in level 3, and so on. The

four levels represented a good measure to keep all choices under control and clearly organized. From here, the scenario progressed from level 1 to level 4, where the story concluded based on the choices selected. (4) The **Participant Statement** section provided four possible choices designed based on each leadership style (S1, S2, S3, and S4), from which participants had to select only one per level, e.g., *I appreciate your enthusiasm for organizing our end-of-year party.* (5) The **Follower Response** provided the response of the follower to the specific statement selected by the participant. We designed the responses to be consistent with the followers' development level regarding a specific task, e.g., *Thank you. I am very excited about this task, and I want to start as soon as possible.* Lastly, (6) the **Feedback** section provided information about the development level of the follower and the best leadership style required for each scenario, e.g., *Alice seems excited about the task. Her commitment is high. However, she has no previous experience. Alice is definitely in need of a Directing style of leadership.*

3.1.2. Real-Human Interaction Design
Before the study, we recruited and trained two collaborators to play the followers' roles in the RH-RW condition. Although the impact of the followers' gender (male or female) was not one of the focuses of the investigation, our intention was to replicate a conventional social interaction allowing participants to equally interact with both genders, not giving preference to either. Thus, we included one male and one female collaborator. The two collaborators played four scenarios each and changed their outfits in between to give a diverse impression. To manage all the narratives, we used Articulate Storyline[2], an eLearning-authoring platform for instructional designers. Using Storyline, we designed a computer application with a consistent structure similar to the one used for the VHs. Our application triggered the narratives automatically according to the participant's selection.

[2]Articulate Storyline: https://articulate.com/360/storyline.

FIGURE 3 | VHs developed using Adobe Fuse CC.

Thus, participants and collaborators were continuously and simultaneously informed with the possible statements and responses as scenarios progressed.

3.1.3. Virtual-Human Interaction Design

For the creation of the VHs, we used Adobe Fuse CC to provide natural-looking 3D characters (**Figure 3**), since VH appearance, including realism, can affect the users' emotional reactions and perceptions (Jung and Hughes, 2016; Volonte et al., 2016; Jung et al., 2017, 2018a,b). Moreover, to avoid perceptual confounds from the gap between environments (Jung et al., 2018b), we tried to match the similarities and characteristics of all three conditions. Therefore, to match the gender of the two RH collaborators, we provided male and female characters, and we also adjusted their outfits and facial expressions.

We then used Mixamo[3] to give our characters a full human rigged skeleton and facial blend shapes ready to animate. We then completed the animation process using the Unity3D (2019.2.12f1) game engine on an Alienware desktop equipped with an Intel i7-8700K CPU @ 3.70 GHz, 32 GB RAM, and an NVIDIA GTX 1080 Ti. **Figure 4** shows the animation set up. For body animation, we used the Unity plugin final inverse kinematics (Final IK[4]), an HTC Vive Pro HMD with two Vive controllers, and three HTC Vive trackers. For the facial expressions, we used the HTC Vive Pro integrated microphone

and the Unity plugin SALSA[5] to provide lip-sync movement and random eye movement. Finally, we organized the animation and audio recordings into a finite state machine following the same structure shown in **Figure 2**.

3.1.4. Computer-Generated Environments

We rendered VHs within two computer-generated environments (VR and MR) using an HTC Vive Pro HMD with a resolution of 1,440 × 1,600 pixels (per eye), a refresh rate of 90 Hz, and a field of view of 110°. To provide a video see-through based MR experience, we attached a Zed Mini stereo camera in front of the HMD as per Jung et al. (2018a,b). As shown in **Figure 6**, we displayed the narratives with a clear definition in VR and MR using the high resolution of the Vive Pro. For the VH-VR condition, we used a 3D representation of the physical room where we conducted our experiment. As shown in **Figure 5**, we scaled the 3D model to have similar dimensions to the physical space. Moreover, because participants could see their real bodies in the RW and MR conditions, but not so in the VH-VR condition, we provided them with a virtual body (avatar) in VH-VR to compensate for the sense of embodiment experienced. Finally, VH-VR participants were also able to move the avatar's arms using the two Vive controllers.

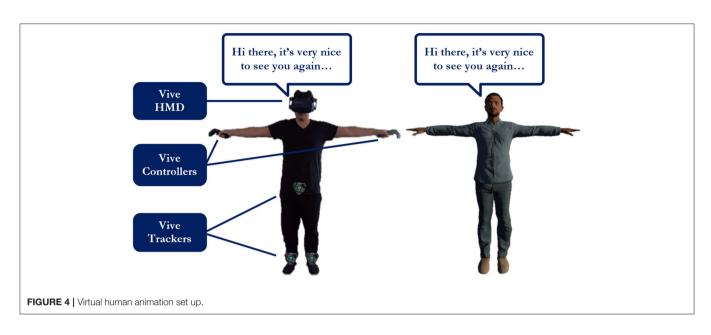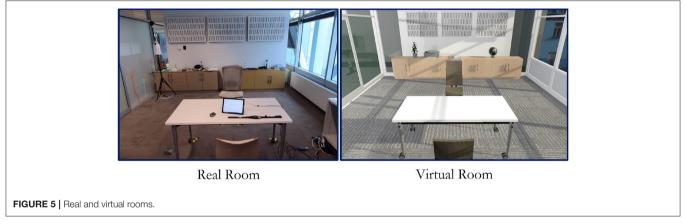### 3.1.5. Implementation

We conducted our study in an isolated experiment room. We prepared three different setups according to each condition

---

[3]Mixamo: https://www.mixamo.com/.
[4]Final IK: https://assetstore.unity.com/packages/tools/animation/final-ik-14290.
[5]SALSA: https://crazyminnowstudio.com/.

FIGURE 4 | Virtual human animation set up.

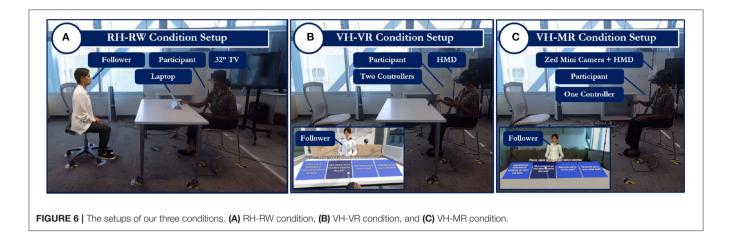**FIGURE 5 |** Real and virtual rooms.

(**Figure 6**). In the RH-RW condition, we ran the Storyline application with a Microsoft Surface (laptop) equipped with an Intel i5-4300U CPU @ 1.90 GHz with 4 GB RAM. We positioned the laptop in front of the participants. Additionally, we used a 32 inch TV and Bluetooth earbuds to reproduce and support RH followers (collaborators) with their scripts, which we set up to change automatically according to the statement selected by participants when touching a button on the laptop screen. We positioned the TV behind the participants (leaders) and visible to the collaborators. Simultaneously, collaborators were able to listen to the recordings of their scripts through the earbuds they wore while playing their roles. We implemented this last extra measure to encourage eye contact with participants, as the TV necessarily required followers to look to the side.

For the VH-VR and VR-MR conditions, we incorporated a virtual pointer as an extension of one of the controllers. Participants only had to point to their statement and then press the trigger button to confirm the selection. As section 3.1.4 described, only participants in the VH-VR condition were given both Vive controllers, used to point and move the arms of their avatar simultaneously. Moreover, for the VH-MR condition, we had to remove the white desk located at the center of the room, as it was affecting the rendering of the VHs due to an occlusion problem. However, we displayed a virtual desk in replacement. We rendered both experiences using a desktop computer equipped with an Intel i7- 8700 CPU @ 3.20 GHz, 32 GB RAM, and an NVIDIA RTX 2080.

Finally, we also used a 52 inch TV and a Zephyr Bioharness heart-rate device in all three conditions. We used the 52 inch TV to deliver a 20 min video session, where an accredited SLII® Certified Trainer provided the key concepts of the model. We asked participants to wear headphones during the video session to reduce the number of distractions. Participants also wore a chest strap with the Zephyr BioHarness[6], used for the recording of their heart rate (HR) and heart rate variability (HRV) during interaction with either RHs or VHs.

---

[6]Zephyr Heart-rate sensor: https://www.zephyranywhere.com/.

FIGURE 6 | The setups of our three conditions. **(A)** RH-RW condition, **(B)** VH-VR condition, and **(C)** VH-MR condition.

3.2. Participants

We recruited 30 participants through our institute's social network pages and advertisements posted on billboards around the campus. Among the 30 participants, 18 (60%) were males, and 12 (40%) were females. Participants' ages varied between 22 and 70 years ($M = 35.0, SD = 11.51$). Before conducting the study, we asked participants about (1) their experience in leadership positions (e.g., Manager, Supervisor, Director, etc.), meaning the function of a person who leads, guides, or commands someone; (2) if they had received training in SLII®; and (3) their familiarity with a head-mounted display (HMD). Eight said they had never been assigned to a leadership position, 12 had been assigned a few times per year, three participants a few times per week, and seven participants said they were assigned daily. Four participants said they had received training in SLII®. Sixteen participants said they had never used an HMD before, 11 had used one a few times per year, and three reported to be more frequent users. All 30 participants completed the experiment and received a $15 voucher for their participation. The experiment was conducted with the approval of our institute's Human Ethics Committee.

3.3. Study Design

We conducted a 3 × 2 mixed factorial design with the two independent variables (IV) *Mediation-type* and *Practice-type*, and the four dependent variables (DV) Learning Performance, Stress Levels, Social Presence, and Overall Training Experience. The IV *Mediation-type* was a between-subjects factor with three levels; Real Human role-player in RW (RH-RW), Virtual Human role-player in VR (VH-VR), and Virtual Human role-player in MR (VH-MR). The IV *Practice-type* was a within-subjects factor with two levels; Pre-session, and Post-session. To control ordering effects, we randomly assigned in a counterbalanced order our 30 participants into the three mediation-type conditions (10 participants per condition) using Research Randomizer[7]. Please note that throughout this article, the naming conventions *mediation-type* and *group* are used interchangeably.

[7]Research Randomizer: https://www.randomizer.org/.

3.3.1. Experimental Tasks

We asked participants to play the role of leaders during structured scenarios before and after receiving an instructional leadership session. The main task of participants was to lead, guide, or command a total of eight followers, one per scenario, using the best matching leadership style. *Follower* means someone who supports and is led, guided, or command by a leader. Participants in the RH-RW group interacted with live humans (collaborators). Participants in the VH-VR and VH-MR groups wore VR or MR HMDs, respectively, and interacted with VHs. The specific tasks of the three groups were organized into stages described below.

3.3.1.1. Pre-session

During this stage, we asked participants to complete four SLII® scenarios presented in the following order: D1, D2, D3, and D4. According to the structural design shown in **Figure 2**, we informed participants to start by reading the introduction section at the beginning of each scenario. The introduction section provided relevant information about both scenarios and followers to put participants into context. Then, the follower (RH or VH), sitting in front of the participant, started his/her acting. Once the follower finished his/her introduction, we asked participants to select only one of the four leadership statements displayed in front of them. We informed participants to select the statement they considered the best according to the follower's narrative. We also asked participants to say aloud their chosen answers before selection. The scenario continued with the follower's response, and so on until the four levels specified in **Figure 2** were completed. Each scenario finalized with the feedback section, which reported the best approach to be taken according to the follower's development level. Participants continued to the next stage (Information-based session) once they completed the four scenarios.

3.3.1.2. SLII® Information-Based Session

During this stage, we asked participants from all three conditions to sit in front of a physical 52 inch TV and watch a 20 min video where an Accredited Certified Trainer in SLII® provided

the fundamental aspects of the SLII® model. Then, participants continued to the final stage of the experiment, Post-session.

3.3.1.3. Post-session

During this stage, we asked participants to repeat the same steps from the Pre-session. However, we introduced them to four new SLII® scenarios with different narratives to prevent the impact of learning effects. We presented the scenarios to the participants in the order D1, D2, D3, and D4.

3.3.2. Hypotheses

To the best of our knowledge and despite the known limitations, traditional role-playing is still considered an effective training technique used to practice and develop leadership skills. Therefore, for this study, we considered traditional role-playing our gold-standard. Moreover, we also investigated that VHs could help make people practicing social skills feel more comfortable than RH role-players, and that immersive virtual environments have the potential to increase the sense of being with another person, allowing VHs to elicit feelings of social presence (Bailenson et al., 2001; Wang et al., 2011; Gratch et al., 2016). Consequently, as proof of the effectiveness, we expected: (1) participants in interaction with VHs to obtain similar learning performance scores to participants in interaction with RH role-players; (2) the stress levels to be higher in the RH-RW condition; (3) VHs would be able to evoke feelings of social presence in immersive environments; (4) leadership training sessions to be able to impact participants consistently, regardless of the mediation type. From our expectations and the related work surveyed, we formulated four hypotheses.

- H_1: There will be no significant difference in learning performance scores and overall training experience among the RH-RW, VH-VR, and VH-MR conditions.
- H_2: Learning performance scores will be significantly higher in the post-sessions, in comparison to the pre-sessions for all three conditions.
- H_3: Stress level indicators will be significantly higher in the RH-RW condition, in comparison to the VH-VR and VH-MR conditions.
- H_4: There will be no significant difference in the sense of social presence between the VH-VR and VH-MR conditions (VR and MR elicit similar levels of social presence).

3.4. Measures

3.4.1. Learning Performance

According to the broadly used and supported SLII® framework (Zigarmi et al., 1997; Blanchard et al., 2015; Zigarmi and Peyton, 2017), leaders should be able to match their leadership style to the development level of the follower. Thus, we assessed the Learning Performance by the number of matches or mismatches obtained by participants in interaction with either RHs or VHs during pre- and post-sessions built with a total of eight structured role-play scenarios. Participant responses were scored: +2, +1, −1, and −2. Positive scores were associated with the answers required to match (+2) or close to match (+1) the development level of the follower. Similarly, negative scores were associated with the answers that represented

the opposite or a mismatch. As an example, leadership styles consist of a combination of directive and supportive behavior, and that combination could vary between S1, S2, S3, and S4. When D1 (low competence and high commitment) represents the development level of a follower, S1 (high directive and low supportive behavior) represents the appropriate leadership style. However, as the D1 follower still requires direction, the leadership style S2 (high directive and high supportive behavior) is not completely incorrect. When that was the case, we evaluated the response as +1, and we used the same logic for the negative scores. Participants were able to select a total of four possible statements per scenario. Therefore, we had a total of 32 chosen statements during pre- and post-sessions, with an overall score ranging from 64 points (highest score) to −64 points (lowest score).

3.4.2. Stress Levels

We used two measures during pre- and post-sessions to provide a robust analysis of the Stress Levels, a Zephyr Bioharness heart-rate sensor, and the DSSQ-3 (Matthews et al., 2005), a short version of the Dundee Stress State Questionnaire (DSSQ; Matthews et al., 1999, 2002). On the one hand, the Zephyr Bioharness has shown good reliability measuring physiological data across multiple contexts (Nazari et al., 2018). We used the Zephyr sensor attached to a chest strap to record participant heart rate (HR) and heart rate variability (HRV). However, while HR focuses on the average beats per minute and is a good indicator of the cardiovascular system, HRV measures the variation in time (ms) between each heartbeat and is a more complex and precise measurement of the body's response to, and recovery from, stress (Thayer et al., 2012; Kim H.G. et al., 2018). Therefore, we decided to evaluate HRV instead of HR as one of the measures to contribute to the analysis of Stress Levels. On the other hand, the DSSQ-3 (see Helton, 2004 for an alternate short DSSQ) has been shown to be a valid alternative to the full version of the scale (Matthews et al., 2005; Matthews and Zeidner, 2012), and it is especially useful in experimental settings where testing time is limited (Matthews, 2016). We also used the DSSQ-3 to contribute to the analysis of Stress Levels by evaluating separately three main subjective state factors: *Task Engagement* (i.e., energy, task motivation, concentration), *Distress* (i.e., tension, unpleasant mood, lack of confidence), and *Worry* (i.e., self-focused attention, low self-esteem, cognitive interference related to task and personal concerns). The DSSQ-3 consisted of 30 items based on a five-point Likert scale.

3.4.3. Social Presence

The Social Presence concept refers to the degree to which the user believes that he or she is interacting with another veritable human being (Blascovich et al., 2002). Based on our findings, the arrival of more sophisticated computer-generated technologies can improve substantially the sense of Social Presence by allowing users to communicate a broader range of signals that were limited in many ways in traditional computer-mediated communication (Albertson, 1980; Daft and Lengel, 1986; Blascovich et al., 2002). Hence, considering our setup with VHs in VR and MR, we were also interested in evaluating Social

TABLE 1 | Descriptive statistics [mean value (SD)] results for Learning Performance and Stress Levels indicators; HRV in milliseconds and DSSQ-3 components; *$p < 0.05$; **Statistically significant decrease ($p = 0.042$) represented by the total sum of the mediation-type scores obtained for Distress during post-sessions.

Practice-type	Mediation-type	Measure				
		Learning performance	HRV	DSSQ-3 Task engagement	DSSQ-3 Distress	DSSQ-3 Worry
Pre-session	RH-RW	11.1 (5.95)	54.0 (17.92)	25.8 (5.59)	8.9 (5.52)	13.6 (5.66)
	VH-VR	12.1 (7.35)	54.7 (10.88)	22.6 (4.85)	9.6 (2.95)	14.5 (4.99)
	VH-MR	11.0 (9.66)	55.1 (12.12)	23.5 (4.50)	7.1 (3.98)	14.6 (7.24)
Post-session	RH-RW	13.0 (6.21)	58.4 (21.46)	26.2 (4.70)	7.5** (4.42)	13.3 (4.05)
	VH-VR	14.7 (6.44)	55.8 (12.55)	23.3 (4.73)	7.9** (3.47)	13.8 (4.41)
	VH-MR	18.4*(8.72)	58.7 (17.14)	26.8*(3.36)	5.4** (4.03)	16.7 (7.25)

Presence as another relevant indicator of VH effectiveness. To evaluate this, we used the Social Presence scale by Bailenson et al. (2001), consisting of five items rated on a seven-point Likert scale. Based on this scale, Social Presence scores increased the more the user believed that a VH represented a real other. On the contrary, negative scores indicate that participants felt the VH was just a computerized image. We measured Social Presence at the end of the experiment, and only the VH-VR and VH-MR groups answered the questionnaire, since RH-RW is ground truth.

3.4.4. Overall Training Experience

To evaluate the Overall Training Experience, we used the Kirkpatrick Model (Kirkpatrick, 1994, 1996; Kirkpatrick and Kirkpatrick, 2016), widely recognized as a standard instrument for analyzing and evaluating the results of formal or informal training (Watkins et al., 1998). The model consists of four levels of evaluation, namely *Reaction* (level 1), *Learning* (level 2), *Behavior* (level 3), and *Results* (level 4). For this study, we only considered the first two levels, as levels 3 and 4 can only be evaluated once a long period of time has elapsed after training. *Reaction* evaluates the extent to which participants find the training beneficial, engaging, and appropriate to their main task or jobs. *Learning* evaluates the extent to which participants obtain the proposed learning, confidence, and commitment. To evaluate *Reaction*, we used a four-point Likert scale (Kirkpatrick, 2008) with six items, *Program Objectives* (e.g., program Goals clearly defined), *Content Relevance* (e.g., appropriate material to fit trainee needs), *facilitator Knowledge* (e.g., facilitator is able to demonstrate a good understanding of the material), *Delivery* (e.g., they were used of a good variety of instructional methods), *Evaluation* (e.g., role-plays or simulations were a fair representation of the program content), and *Facility* (e.g., room set-up). To evaluate *Learning*, we used a 10-point Likert scale (Kirkpatrick and Kirkpatrick, 2009) with two items, *Confidence* (i.e., level of confidence to apply the knowledge learned) and *Commitment* (i.e., level of commitment to apply the knowledge learned). We measured both *Reaction* and *Learning* at the end of the experiment.

3.5. Procedure

At the beginning of the experiment, we provided participants with a general overview and explanation of the project, followed by an information sheet and the consent form to read and sign. We then gave participants two pre-experiment questionnaires to answer, a demographic questionnaire and the DSSQ-3. Following this, participants were given detailed instructions about the experiment and how to perform the experimental tasks. Then, we asked them to wear a chest strap with the heart-rate sensor. At this point, participants were ready to begin the experiment in the group to which they were randomly assigned. After completing the experience, we asked participants to complete four post-experiment questionnaires, the DSSQ-3, the Social Presence scale, and Kirkpatrick's level 1 and level 2. Lastly, we had a debriefing session where participants received clarification to questions they may have had, and we also asked them for their opinion about the experiment. The total study duration was about 60 min.

4. RESULTS

In this section, we present the results obtained from the experiment. In **Table 1**, we provide an overview of the marginal means and standard deviations (SD). In **Table 2**, we provide an overview of the medians and interquartile ranges (IQR). Once we confirmed the data were parametric, we conducted a mixed-design analysis of variance (split-plot ANOVA) with the DV Learning Performance and Stress Levels. Each variable was analyzed by separately applying a split-plot ANOVA with the within-factors Pre-session and Post-session *(Practice-type)*, and the between-factors RH-RW, VH-VR and VH-MR *(Mediation-type)*. The overall learning performance score obtained during pre-post sessions was studied using a one-way analysis of variance (ANOVA). We also confirmed that the data from the DV Social Presence and the Overall Training Experience were non-parametric. Social Presence was analyzed using Mann–Whitney U. The Overall Training Experience was analyzed using Kruskal–Wallis H. An *a priori* significance level was set at $p < 0.05$, and partial η^2 (η_p^2) is reported as a measure of effect size.

4.1. Learning Performance

The split-plot ANOVA showed there were no statistically significant interaction effects between the groups and pre-post sessions on Learning Performance scores, $F_{(2, 27)} =$

TABLE 2 | Descriptive statistics [median value (IQR)] results for Social Presence (first row); Reaction (second block of rows) and items (third block of rows); Learning (fourth block of rows) and items (Fifth block of rows).

| | Mediation-type | | |
Measure	RH-RW	VH-VR	VH-MR
Social presence	–	2.0 (1.7)	1.0 (6.7)
Reaction	20.0 (4.7)	18.5 (4.0)	18.5 (2.7)
Program objectives	5.0 (1.0)	4.0 (0.7)	4.0 (1.0)
Content relevance	2.0 (1.0)	2.0 (0.1)	2.0 (0.1)
Facilitator knowledge	2.0 (0.7)	2.0 (0.7)	2.0 (0.7)
Delivery	5.0 (2.0)	4.0 (1.7)	4.0 (0.7)
Evaluation	1.5 (1.0)	2.0 (0.1)	2.0 (0.7)
Facility	4.5 (1.7)	4.0 (1.7)	5.0 (1.7)
Learning	16.0 (2.5)	15.0 (2.7)	16.0 (2.0)
Confidence	8.0 (0.7)	7.5 (1.0)	8.0 (1.7)
Commitment	8.0 (2.2)	8.0 (1.0)	8.0 (0.1)

$1.636, p = 0.214, \eta_p^2 = 0.108$. All three groups obtained an increase in performance scores from pre- to post-sessions, and the main effect showed a statistically significant difference, $F_{(1, 27)} = 8.613, p = 0.007, \eta_p^2 = 0.242$. Still, *post-hoc* analysis with Bonferroni adjustment revealed that only the VH-MR group had a statistically significant mean increase of 7.4, 95% CI [3.59, 11.2], $p = 0.002$. We summed participant responses to provide an overall Learning Performance score. Participants in the VH-MR condition obtained better scores ($M = 29.4$, $SD = 17.62$) than participants in the VH-VR condition ($M 26.8$, $SD = 11.24$), and participants in the RH-RW condition ($M = 24.1$, $SD = 8.77$), but we did not find a statistically significant difference in mean scores between groups $F_{(2, 27)} = 0.410, p = 0.668, \eta_p^2 = 0.029$.

4.2. Stress Levels

4.2.1. Heart-Rate Variability
We did not find statistically significant interaction effects between the groups and pre-post sessions on HRV, $F_{(2, 27)} = 0.384, p = 0.685, \eta_p^2 = 0.028$. All three groups had an increase in HRV from pre- to post-sessions, but the difference was not statistically significant between sessions, $F_{(1, 27)} = 3.57, p = 0.069, \eta_p^2 = 0.117$, and between groups, $F_{(2, 27)} = 0.030, p = 0.971, \eta_p^2 = 0.002$.

4.2.2. Dundee Stress State Questionnaire (DSSQ-3)
We found a statistically significant interaction effect between the groups and pre-post sessions on *Task Engagement*, $F_{(2, 27)} = 3.80, p = 0.035, \eta_P^2 = 0.220$. All three groups had an increase in task engagement scores from pre- to post-sessions, but the only statistically significant increase was represented by the VH-MR group ($M = 3.3, SE = 0.578, p = 0.001$). We did not find a statistically significant interaction effect between the groups and pre-post sessions on *Distress*, $F_{(2, 27)} = 0.018, p = 0.982, \eta_P^2 = 0.001$. *Post-hoc* analysis with Bonferroni adjustment revealed

that there was a general decrease in distress from pre-sessions ($M = 8.5, SD = 4.27$), to post-sessions ($M = 6.9, SD = 4.01$), a statistically significant mean decrease of 1.60, 95% CI [.059, 3.14], $p = 0.042$, based on the total score obtained by the three mediation-type conditions. Participants in the VH-MR group obtained a lower marginal mean for *Distress* ($M = 6.2, SE = 1.13$) than participants in the RH-RW condition ($M = 8.2, SE = 1.13$) and participants in the VH-VR condition ($M = 8.7, SE = 1.13$), but the difference was not statistically significant $F_{(2, 27)} = 1.33, p = 0.281, \eta_P^2 = 0.090$. Lastly, we did not find a statistically significant interaction effect between the groups and pre-post sessions on *Worry*, $F_{(2, 27)} = 0.579, p = 0.567, \eta_P^2 = 0.041$. There was no statistically significant difference in mean *Worry* between sessions, $F_{(1, 27)} = 0.102, p = 0.752, \eta_P^2 = 0.004$, and between groups, $F_{(2, 27)} = 0.546, p = 0.585, \eta_P^2 = 0.039$.

4.3. Social Presence
We summed participant responses to provide an overall *Social Presence* score. For the VH-VR group, the average Social Presence score was 1.7 ($SD = 3.16$), the minimum score was -6, and the maximum score was 6. For the VH-MR group, the average score was 0.7 ($SD = 4.24$), the minimum score was -6, and the maximum score was 7. Participants in the VH-VR condition reported a greater sense of Social Presence, obtaining a higher and more-consistent number of positive scores (eight out of 10) in comparison to participants in the VH-MR condition (six out of 10). The Mann–Whitney U showed that Social Presence scores for the VH-VR (mean rank = 11.40) and the VH-MR (mean rank = 9.60) were not statistically different, $U = 41, z = -0.685, p = 0.494$.

4.4. Overall Training Experience
4.4.1. Kirkpatrick's Level 1: Reaction
The results showed that all three groups obtained scores above average in all six *Reaction* items, indicating that participants found the training experience beneficial, engaging, and appropriate to their main tasks/roles. We summed the item scores to provide a single dimension concept. For RH-RW, the average Reaction score was 20.3 ($IQR = 4.7$). For VH-VR, the average score was 18.9 ($IQR = 4.0$), and for VH-MR, the average score was 19.1 ($IQR = 2.7$). Median scores for Reaction items were not statistically significantly different between groups [$\chi^2(2) = 1.123, p = 0.570$].

4.4.2. Kirkpatrick's Level 2: Learning
The results showed that all three groups obtained scores above average in both *Learning* items, indicating a high degree of *Confidence* and *Commitment* to apply what was learned. We summed the item scores to provide a single dimension concept. For RH-RW, the average Learning score was 15.7 ($IQR = 2.5$). For VH-VR, the average score was 15.1 ($IQR = 2.7$), and for VH-MR, the average score was 15.7 ($IQR = 2.0$). Median scores for *Confidence* [$\chi^2(2) = 2.580, p = 0.275$] and *Commitment* [$\chi^2(2) = 0.630, p = 0.730$] were not statistically significantly different between groups.

5. DISCUSSION

The results of our study support hypothesis H_1 (There will be no significant difference in learning performance scores and overall training experience among the RH-RW, VH-VR, and VH-MR conditions), and partly, hypothesis H_2 (Learning performance scores will be significantly higher in the post-sessions, in comparison to the pre-sessions for all three conditions). Based on the overall stress-level results, hypothesis H_3 (Stress level indicators will be significantly higher in the RH-RW condition, in comparison to the VH-VR and VH-MR conditions) is rejected. However, we did find support for hypothesis H_4 (There will be no significant difference in the sense of social presence between the VH-VR and VH-MR conditions [VR and MR elicit similar levels of social presence]).

As we expected, both VHs and RHs presented similar outcomes in Learning Performance scores, and the difference among groups was not significant. The increase in participants' performance scores obtained between the three groups during post-sessions might suggest the effective implementation of self-regulation learning strategies (Zimmerman, 1989), offered by the opportunities to practice during the pre-session, and support from both the feedback section and information-based session. Interestingly, the only statistically significant increase from pre- to post-sessions was reported by the VH-MR group. On the other hand, we did not find significant differences among groups in terms of HRV. While a lower HRV suggests a delayed recovery from psychological stressors (Weber et al., 2010), our results showed that HRV measurements were higher during post-sessions in all three groups, which could indicate that participants successfully performed emotion regulation and appropriate body recovery from pre- to post-sessions (Thayer et al., 2012; Kim H.G. et al., 2018). These results are in line with the increase in Learning Performance, corroborating the influence that the pre-session and leadership instruction had in all three groups.

From the analysis of the DSSQ-3, we found an increase in *Task Engagement* scores from pre- to post-sessions in all three groups, but only the VH-MR group reported a significant increase. A study by Matthews (2016) describes that monotonous tasks tend to decrease task engagement. Therefore, we consider the increase in task engagement scores to indicate that participants perceived the training experience as a dynamic and challenging game-like task. Furthermore, increments in task engagement are also comparable to the Learning Performance and HRV trends, suggesting that the opportunities to practice leadership competencies may have allowed participants to reflect on their own experience and firmly engage in the learning process (McCauley and Van Velsor, 2004). For *Distress*, we found a statistically significant decrease in the total scores obtained from pre- to post-sessions. According to Matthews et al. (2002), an increase in distress levels is associated with an overload of the processing capacities. Some tasks have even been shown to reduce distress levels when participants enjoy the task (Matthews, 2016), which is consistent with the results obtained here for task engagement. Lastly, the decrease in *Worry* from pre- to post-sessions in the RH-RW and VH-VR groups suggests a reduction of self-focused attention. In general, high performance tends to

reduce worry, as the attention of the user is focused on external tasks (Matthews, 2016). However, in contrast to this tendency, the VH-MR group reported an increase in worry. We then explored possible explanations and found that worry tends to be maintained when the task provides opportunities for self-reflection and mind-wandering (Neubauer et al., 2012). Thus, we conjecture that the increase in worry obtained by VH-MR was for the same reason.

From the results obtained by the Social Presence responses, the difference between groups was not significant, and 14 out of 20 scores were positive, indicating that more than half of the participants perceived the VHs to be realistic and assigned some degree of consciousness to them. These findings are in line with other investigations where VHs were shown to evoke feelings of presence (Bailenson et al., 2001; Rosenthal-von der Pütten et al., 2009). Moreover, the descriptive statistics indicate that participants in the VH-VR group reported a greater and more consistent sense of Social Presence in comparison to participants in the VH-MR group. This could be because of the technical differences between the devices used to render the VHs. In the MR condition, we displayed VHs with a lower resolution and reduced field of view due to the limited video feed of the stereo camera. Therefore, we think that the behavioral realism of the VHs could have been affected in the MR group. Finally, for the Overall Training Experience, the scales for *Reaction* and *Learning* showed that participants scored consistently and above average in all three groups. These results suggest that participants reacted positively to the training experience and had a high degree of confidence and commitment in applying the knowledge acquired. Still, both *Reaction* and *Learning* outcomes are only descriptive, and their analyses do not allow us to better understand how participants apply what they learned during our training (Level 3: Behavior), or even analyze the degree to which targeted outcomes occur as result (Level 4: Results). Further research should appropriately include these two other Kirkpatrick evaluation levels (Kirkpatrick and Kirkpatrick, 2016) to fully grasp the real impact of the training experience.

5.1. Implications

This study showed that both RH and VH role-players were able to make participants feel comfortable and remain engaged in the task from pre- to post-sessions. Surprisingly, the Learning Performance scores were not only similar but higher with VHs in both computer-generated environments compared to RHs. Although the differences were not statistically significant, we still consider that these interesting outcomes, particularly those obtained for VH-MR, should be further investigated since, during this occasion, we did not set up the necessary conditions (e.g., RH-VR, RH-MR) to make a fair comparison between environments.

In response to the current role-playing limitations, we consider that VHs are still not able to address adequately all the barriers that we have identified in this study. VHs still need to be programmed, so the development of their training experiences may necessarily require the participation of experts, and the computer-generated environments in which VHs operate may also require specific equipment involving additional costs.

Nevertheless, our research has also indicated that with adequate context and instructional support, VHs can become invaluable tools, where their human-like appearance, rich configurability, and ease of replication allow them to play a critical role in social skills training. VHs have been shown to provide (1) several opportunities to practice allowing self-reflective learning and engagement, (2) realistic scenarios of what leaders might encounter on the job, (3) constructive performance feedback, (4) a consistent training experience, and (5) a means for evoking feelings of social presence. As the cost of the new computer-generated technologies decreases and more automated content creation tools arise, we would suggest to developers and instructional designers to embrace the implementation of immersive simulators using VHs, as they can offer numerous applications in diverse fields, where corporate training and medical education could become their key beneficiaries.

5.2. Limitations

In this study, we have identified some limitations that could be helpful for future researchers. First, our ability to interpret the results was restricted by the small number of participants, as we only had 10 participants per mediation-type condition. However, the existence of consistent outcomes found between RHs and VHs is promising. Likely, larger sample sizes will only strengthen our conclusions and allow a more precise analysis of how VHs can impact Learning Performance for the development and practice of leadership skills. Secondly, having a fixed set of narratives was useful for delivering a consistent training experience. However, we also consider this approach to be somewhat restrictive in some ways for participant behavior. Increasing the number of choices rapidly becomes a challenging task for storyline developers, and elaborate narratives may result in unrealistic stories. Not following a proper structure also becomes a problem since users can quickly lose track of the instruction. Perhaps a smarter system could provide a semi-structured design, where the interaction follows a validated framework, but where users can formulate their own statements. Thirdly, the development of the scenarios was a time-consuming task, requiring the involvement of a subject-matter expert. Lastly, the stereo camera that we used in the study had a low field of view, and we set it up using a lower resolution to compensate for the low frame rate at higher resolutions in an attempt to avoid symptoms of cybersickness. These technical limitations may have impacted the sense of Social Presence. As technology develops, it would be interesting to explore the use of more advanced depth-sensing cameras.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we investigated the impact of VH role-players as training tools to complement and support traditional training delivery methods, which have some limitations. To evaluate the effectiveness, we compared VH role-players to RH role-players during pre- and post-sessions built with eight structured scenarios using SLII® content as our framework. With support from the findings in our study, we can conclude that (1) VHs can be as effective as RH role-players to support the practice of leadership skills, and (2) the human-type (RH, VH) in conjunction with the environment-type (RW, VR, and MR) had a positive impact on the training outcomes. Surprisingly, the VH-MR mediation-type condition had a greater influence on Learning Performance and Task Engagement. Finally, we can report that the overall training was a consistent experience with positive reactions and learning.

After reviewing the outcomes obtained with VHs within VR and MR environments, we consider future work opportunities. (1) Investigate the effects that different computer-generated environments have when both VHs and RHs coexist within these environments. (2) In this study, we relied upon the support of a RH trainer to deliver the leadership instruction we used in our experiment. Further research could also investigate the impact of a VH trainer or vice versa, a captured RH superimposed onto the virtual environment. Lastly, considering the valuable configurability provided by VHs, it would be interesting to investigate the impact that characters with a rich cultural diversity have on learning performance.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Human Ethics Committee—University of Canterbury. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

The study design was planned by GS, SJ, and RL. The experiment implementation and data collection were completed by GS. GS and SJ performed the data analysis and wrote the paper. RL supervised the process and extensively commented and revised the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

REFERENCES

Albertson, L. A. (1980). Review essay: trying to eat an elephant the social psychology of telecommunications, by John Short, Ederyn Williams, and Bruce Christie. London: John Wiley, 1976. *Commun. Res.* 7, 387–400. doi: 10.1177/009365028000700307

Badler, N. (1997). "Virtual humans for animation, ergonomics, and simulation," in *Proceedings of the 1997 IEEE Workshop on Motion of Non-Rigid and Articulated Objects, NAM '97* (San Juan, PR: IEEE Computer Society), 28–36. doi: 10.1109/NAMW.1997.609848

Bailenson, J. N., Blascovich, J., Beall, A. C., and Loomis, J. M. (2001). Equilibrium theory revisited: mutual gaze and personal space in virtual environments. *Presence* 10, 583–598. doi: 10.1162/105474601753272844

Batrinca, L., Stratou, G., Shapiro, A., Morency, L.-P., and Scherer, S. (2013). "Cicero - towards a multimodal virtual audience platform for public speaking training," in *Intelligent Virtual Agents*, eds R. Aylett, B. Krenn, C. Pelachaud, and H. Shimodaira (Berlin; Heidelberg: Springer Berlin Heidelberg). doi: 10.1007/978-3-642-40415-3_10

Blanchard, K. H., Zigarmi, D., and Nelson, R. B. (1993). Situational leadership® after 25 years: a retrospective. *J. Leadersh. Stud.* 1, 21–36. doi: 10.1177/107179199300100104

Blanchard, K. H., Zigarmi, P., and Zigarmi, D. (2015). *Leadership & The One Minute Manager: Increase Effectiveness Through Situational Leadership II®*. London: HarperCollins.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., and Bailenson, J. N. (2002). Target article: Immersive virtual environment technology as a methodological tool for social psychology. *Psychol. Inq.* 13, 103–124. doi: 10.1207/S15327965PLI1302_01

Bosman, K., Bosse, T., and Formolo, D. (2019). "Virtual agents for professional social skills training: an overview of the state-of-the-art," in *Intelligent Technologies for Interactive Entertainment*, eds P. Cortez, L. Magalhães, P. Branco, C. F. Portela, and T. Adão (Cham: Springer International Publishing), 75–84. doi: 10.1007/978-3-030-16447-8_8

Campbell, J., Hays, M., Core, M., Birch, M., Bosack, M., and Clark, R. E. (2011). "Interpersonal and leadership skills: using virtual humans to teach new officers," in *Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)* (Orlando, FL).

Conrad, D., and Newberry, R. (2011). 24 business communication skills: attitudes of human resource managers versus business educators. *Am. Commun. J.* 13, 4–23. Retrieved from: http://ac-journal.org/journal/pubs/2011/spring/ConradNewberry.pdf

Core, M., Traum, D., Lane, H. C., Swartout, W., Gratch, J., van Lent, M., et al. (2006). Teaching negotiation skills through practice and reflection with virtual humans. *Simulation* 82, 685–701. doi: 10.1177/0037549706075542

Daft, R. L., and Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Manage. Sci.* 32, 554–571. doi: 10.1287/mnsc.32.5.554

DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., et al. (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, AAMAS '14* (Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems), 1061–1068.

Fox, J., Ahn, S. J. G., Janssen, J. H., Yeykelis, L., Segovia, K. Y., and Bailenson, J. N. (2015). Avatars versus agents: a meta-analysis quantifying the effect of agency on social influence. *Hum.-Comput. Interact.* 30, 401–432. doi: 10.1080/07370024.2014.921494

Gratch, J., DeVault, D., and Lucas, G. (2016). "The benefits of virtual humans for teaching negotiation," in *Intelligent Virtual Agents*, eds D. Traum, W. Swartout, P. Khooshabeh, S. Kopp, S. Scherer, and A. Leuski (Cham: Springer International Publishing), 283–294. doi: 10.1007/978-3-319-47665-0_25

Guadagno, R. E., Blascovich, J., Bailenson, J. N., and McCall, C. (2007). Virtual humans and persuasion: the effects of agency and behavioral realism. *Media Psychol.* 10, 1–22. doi: 10.108/15213260701300865

Guetterman, T., Kron, F., Campbell, T., Scerbo, M., Zelenski, A., Cleary, J., et al. (2017). Initial construct validity evidence of a virtual human application for competency assessment in breaking bad news to a cancer patient. *Adv. Med. Educ. Pract.* 8, 505–512. doi: 10.2147/AMEP.S138380

Hartholt, A., Mozgai, S., and Rizzo, A. (2019). "Virtual job interviewing practice for high-anxiety populations," in *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents, IVA '19* (New York, NY: Association for Computing Machinery), 238–240. doi: 10.1145/3308532.3329417

Hartholt, A., Traum, D., Marsella, S. C., Shapiro, A., Stratou, G., Leuski, A., et al. (2013). "All together now," in *Intelligent Virtual Agents*, eds R. Aylett, B. Krenn, C. Pelachaud, and H. Shimodaira (Berlin; Heidelberg: Springer), 368–381. doi: 10.1007/978-3-642-40415-3_33

Hays, M., Campbell, J., Trimmer, M., Poore, J., Webb, A., Stark, C., et al. (2012). "Can role-play with virtual humans teach interpersonal skills?," in *Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)* (Orlando, FL).

Helton, W. S. (2004). Validation of a short stress state questionnaire. *Proc. Hum. Fact. Ergon. Soc. Annu. Meet.* 48, 1238–1242. doi: 10.1177/154193120404801107

Hunt, J. W., and Baruch, Y. (2003). Developing top managers: the impact of interpersonal skills training. *J. Manage. Dev.* 22, 729–752. doi: 10.1108/02621710310487882

Johnsen, K., Raij, A., Stevens, A., Lind, D. S., and Lok, B. (2007). "The validity of a virtual human experience for interpersonal skills education," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '07* (New York, NY: Association for Computing Machinery), 1049–1058. doi: 10.1145/1240624.1240784

Jung, S., Bruder, G., Wisniewski, P. J., Sandor, C., and Hughes, C. E. (2018a). "Over my hand: using a personalized hand in VR to improve object size estimation, body ownership, and presence," in *Proceedings of the Symposium on Spatial User Interaction, SUI '18* (New York, NY: Association for Computing Machinery), 60–68. doi: 10.1145/3267782.3267920

Jung, S., and Hughes, C. E. (2016). "The effects of indirect real body cues of irrelevant parts on virtual body ownership and presence," in *ICAT-EGVE 2016 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments.* (New York, NY), 107–114. doi: 10.1145/2993369.2996346

Jung, S., Sandor, C., Wisniewski, P. J., and Hughes, C. E. (2017). "Realme: the influence of body and hand representations on body ownership and presence," in *Proceedings of the 5th Symposium on Spatial User Interaction, ACM SUI 2017.* (New York, NY), 3–11. doi: 10.1145/3131277.3132186

Jung, S., Wisniewski, P. J., and Hughes, C. E. (2018b). "In limbo: the effect of gradual visual transition between real and virtual on virtual body ownership illusion and presence," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, (Tuebingen; Reutlingen), 267–272. doi: 10.1109/VR.2018.8447562

Jung, S., Wood, A. L., Hoermann, S., Abhayawardhana, P. L., and Lindeman, R. W. (2020). "The impact of multi-sensory stimuli on confidence levels for perceptual-cognitive tasks in VR," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, (Atlanta, GA), 463–472. doi: 10.1109/VR46266.2020.00067

Kaiser, R. B., and Curphy, G. (2013). Leadership development: the failure of an industry and the opportunity for consulting psychologists. *Consult. Psychol. J.* 65, 294–302. doi: 10.1037/a0035460

Kilteni, K., Maselli, A., Kording, K. P., and Slater, M. (2015). Over my fake body: body ownership illusions for studying the multisensory basis of own-body perception. *Front. Hum. Neurosci.* 9:141. doi: 10.3389/fnhum.2015.00141

Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H., and Koo, B.-H. (2018). Stress and heart rate variability: a meta-analysis and review of the literature. *Psychiatry Invest.* 15, 235–245. doi: 10.30773/pi.2017.08.17

Kim, K., Boelling, L., Haesler, S., Bailenson, J. N., Bruder, G., and Welch, G. (2018). "Does a digital assistant need a body? The influence of visual embodiment and social behavior on the perception of intelligent virtual agents in AR," in *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (Munich), 105–114. doi: 10.1109/ISMAR.2018.00039

Kirkpatrick, D. L. (1994). *Evaluating Training Programs: The Four Levels, 1st Edn.* San Francisco, CA: Berrett-Koehler.

Kirkpatrick, D. L. (1996). "Evaluation," in *The ASTD Training and Development Handbook: A Guide to Human Resource Development*, eds R. Craig and L. Bittel (New York, NY: McGraw-Hill), 59–67.

Kirkpatrick, J. (2008). *The New World Level 1 Reaction Sheets.* Available online at: https://www.kirkpatrickpartners.com/Resources (Retrieved May 08, 2020).

Kirkpatrick, J., and Kirkpatrick, W. K. (2009). *New World Level 2: The Importance of Learner Confidence and Commitment*. Available online at: https://www.kirkpatrickpartners.com/Resources (Retrieved May 08, 2020).

Kirkpatrick, J. D., and Kirkpatrick, W. K. (2016). *Kirkpatrick's Four Levels of Training Evaluation*. Alexandria, VA: ATD Press.

Magnenat-Thalmann, N., and Thalmann, D. (2005). Virtual humans: thirty years of research, what next? *Vis. Comput.* 21, 997–1015. doi: 10.1007/s00371-005-0363-6

Matthews, G. (2016). Multidimensional profiling of task stress states for human factors: a brief review. *Hum. Fact.* 58, 801–813. doi: 10.1177/0018720816653688

Matthews, G., Campbell, S. E., Falconer, S., Joyner, L. A., Huggins, J., Gilliland, K., et al. (2002). Fundamental dimensions of subjective state in performance settings: task engagement, distress, and worry. *Emotion* 2, 315–340. doi: 10.1037/1528-3542.2.4.315

Matthews, G., Emo, A. K., and Funke, G. (2005). "A short version of the Dundee Stress State Questionnaire," in *The 12th Meeting of the International Society for the Study of Individual Differences* (Adelaide, SA).

Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., and Huggins, J. (1999). Validation of a comprehensive stress state questionnaire: towards a state big three. *Pers. Psychol. Europe* 7, 335–350. doi: 10.1037/t27031-000

Matthews, G., and Zeidner, M. (2012). Individual differences in attentional networks: trait and state correlates of the ant. *Pers. Individ. Differ.* 53, 574–579. doi: 10.1016/j.paid.2012.04.034

McCauley, C. D., and Van Velsor, E. (2004). *The Center for Creative Leadership Handbook of Leadership Development, Vol. 29, 2nd Edn*. San Francisco, CA: John Wiley & Sons.

Milgram, P., and Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Trans. Inform. Syst.* 77, 1321–1329.

Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. (1995). "Augmented reality: a class of displays on the reality-virtuality continuum," in *Telemanipulator and Telepresence Technologies, Vol. 2351*, ed H. Das (Boston, MA: International Society for Optics and Photonics, SPIE), 282–292. doi: 10.1117/12.197321

Nazari, G., Bobos, P., MacDermid, J. C., Sinden, K. E., Richardson, J., and Tang, A. (2018). Psychometric properties of the Zephyr bioharness device: a systematic review. *BMC Sports Sci. Med. Rehabil.* 10:6. doi: 10.1186/s13102-018-0094-4

Neubauer, C., Matthews, G., and Saxby, D. (2012). The effects of cell phone use and automation on driver performance and subjective state in simulated driving. *Proc. Hum. Fact. Ergon. Soc. Annu. Meet.* 56, 1987–1991. doi: 10.1177/1071181312561415

Piaget, J. (1952). *The Origins of Intelligence in Children*. New York, NY: W W Norton & Co. doi: 10.1037/11494-000

Rosenthal-von der Pütten, A. M., Krämer, N., and Gratch, J. (2009). "Who is there? Can a virtual agent really elicit social presence?," in *The 12th Annual International Workshop on Presence* (Los Angeles, CA).

Schmid Mast, M., Kleinlogel, E. P., Tur, B., and Bachmann, M. (2018). The future of interpersonal skills development: immersive virtual reality training with virtual humans. *Hum. Resour. Dev. Quart.* 29, 125–141. doi: 10.1002/hrdq.21307

Slater, M., McCarthy, J., and Maringelli, F. (1998). The influence of body movement on subjective presence in virtual environments. *Human Fact.* 40, 469–477. doi: 10.1518/001872098779591368

Sogunro, O. (2004). Efficacy of role-playing pedagogy in training leaders: some reflections. *J. Manage. Dev.* 23, 355–371. doi: 10.1108/02621710410529802

Storey, V. J., and Cox, T. D. (2015). Utilizing TeachLivE^{TM} (TLE) to build educational leadership capacity: the development and application of

virtual simulations. *J. Educ. Hum. Dev.* 4, 41–49. doi: 10.15640/jehd.v4 n2a5

Thayer, J. F., Ahs, F., Fredrikson, M., Sollers, J. J., and Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. *Neurosci. Biobehav. Rev.* 36, 747–756. doi: 10.1016/j.neubiorev.2011.11.009

Van Ments, M. (1989). *The Effective Use of Role-Play: A Handbook for Teachers and Trainers*. London; New York, NY: Kogan.

Volonte, M., Babu, S., Chaturvedi, H., Newsome, N., Ebrahimi, E., Roy, T., et al. (2016). Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations. *IEEE Trans. Visual. Comput. Graph.* 22, 1326–1335. doi: 10.1109/TVCG.2016.2518158

Waltemate, T., Gall, D., Roth, D., Botsch, M., and Latoschik, M. E. (2018). The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Trans. Visual. Comput. Graph.* 24, 1643–1652. doi: 10.1109/TVCG.2018.2794629

Wang, M., Lawless-Reljic, S., Davies, M., and Callaghan, V. (2011). "Social presence in immersive 3D virtual learning environments," in *Ambient Intelligence - Software and Applications*, eds Novais, P., Preuveneers, D., and Corchado, J. M., editors (Berlin; Heidelberg: Springer), 59–67. doi: 10.1007/978-3-642-19937-0_8

Watkins, R., Leigh, D., Foshay, R., and Kaufman, R. (1998). Kirkpatrick plus: evaluation and continuous improvement with a community focus. *Educ. Technol. Res. Dev.* 46, 90–96. doi: 10.1007/BF02299676

Weaver, S. J., Rosen, M. A., Salas, E., Baum, K. D., and King, H. B. (2010). Integrating the science of team training: guidelines for continuing education. *J. Contin. Educ. Health Profess.* 30, 208–220. doi: 10.1002/chp. 20085

Weber, C. S., Thayer, J. F., Rudat, M., Wirtz, P. H., Zimmermann-Viehoff, F., Thomas, A., et al. (2010). Low vagal tone is associated with impaired post stress recovery of cardiovascular, endocrine, and immune markers. *Eur. J. Appl. Physiol.* 109, 201–211. doi: 10.1007/s00421-009-1341-x

Wexley, K. N., and Latham, G. P. (2002). *Developing and Training Human Resources in Organizations, 3rd Edn*. Upper Saddle River, NJ: Prentice-Hall.

Zigarmi, D., Edeburn, C., and Blanchard, K. H. (1997). *Getting to Know the LBAII: Research, Validity and Reliability of the Self and Other Forms, 4th Edn*. Escondido, CA: Blanchard Training and Development.

Zigarmi, D., and Peyton, T. (2017). A test of three basic assumptions of Situational Leadership® II Model and their implications for HRD practitioners. *Eur. J. Train. Dev.* 41, 241–260. doi: 10.1108/EJTD-05-201 6-0035

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *J. Educ. Psychol.* 81, 329–339. doi: 10.1037/0022-0663.8 1.3.329