Journal of Evidence-Based Psychotherapies,
Vol. 26, No. 1, March 2026, 19-38.

10.24193/jebp.2026.1.2

FEASIBILITY AND USABILITY OF A VR-BASED TRAINING FOR THERAPEUTIC RELATIONSHIP SKILLS IN FUTURE PSYCHOTHERAPISTS: A PILOT STUDY

Dana-Andreea Cristolțan[1, 2]*

[1] The International Institute for the Advanced Studies of Psychotherapy and Applied Mental Health, Babeș-Bolyai University, Cluj- Napoca, România
[2] Kind Strategies SRL, Cluj-Napoca, Romania, 400445

Abstract

Empathy and therapeutic alliance are key components of evidence-based psychotherapy and strong predictors of positive outcomes. Yet future psychologists often lack practical experience, underscoring the need for training that integrates theory with practice. Virtual reality (VR) offers a safe, controlled environment to rehearse relational skills. This pilot single-group pre–post study examined the feasibility and usability of the *Future Psychotherapists* VR application, designed to enhance empathy and alliance-building skills. Fifty-five participants, predominantly students, completed one VR-based training session interacting with a virtual client in a predefined scenario. Simulator sickness was assessed pre- and post-session, whereas usability and perceived cognitive workload were measured post-intervention. As an exploratory aim, state empathy (cognitive, affective, and associative components) was assessed before and after the session. Simulator sickness showed a slight increase from pre to post, but remained low overall. Usability ratings were high and perceived cognitive workload was moderate. For the exploratory aim, significant pre–post improvements emerged across all empathy components. Expert evaluations supported the scenario's relevance for therapeutic skill development. Overall, the Future Psychotherapists VR tool appears to be a feasible and usable method for training empathy and alliance-related skills in psychology students and may help bridge the gap between academic training and clinical practice by offering realistic opportunities to practice core therapeutic competencies.

* Correspondence concerning this article should be addressed to Dana-Andreea Cristolțan, The International Institute for the Advanced Studies of Psychotherapy and Applied Mental Health, Babeș-Bolyai University, Str. Republicii 37, 400015 Cluj-Napoca, Romania.
ORCID: 0009-0003-5402-012X
Email: dana.cristoltan@ubbcluj.ro

Keywords: Virtual reality; Empathy; Therapeutic alliance; VR training; Future psychotherapists; Skills development

A robust body of evidence highlights the therapeutic relationship as the strongest determinant of successful psychotherapy outcomes, with an effectiveness that is both transtheoretical and largely transdiagnostic, extending across therapeutic modalities and a wide range of clinical problems (Norcross & Karpiak, 2025). These findings place the development of relational competencies at the center of psychotherapist training, pointing to the necessity of immersive and experiential approaches that mirror real-world therapeutic interactions. Research on psychotherapy outcomes indicates that client improvement is influenced by extra-therapeutic factors, expectancy effects, specific treatment techniques, and the client–therapist relationship. Among these, the therapeutic relationship—often referred to as a common factor—has been identified as the most significant contributor to positive outcomes from the therapist's side (Lambert & Barley, 2001). The APA's Interdivisional Task Force on Evidence-Based Relationships and Responsiveness highlights that the therapeutic relationship is as critical to successful outcomes as the specific treatment techniques employed (Norcross & Lambert, 2011). In 2018, Norcross and Lambert revisited the 2011 findings and reaffirmed the presence of core relationship factors, including therapeutic alliance collaboration, empathy, goal consensus, positive regard and affirmation, collecting and delivering client feedback and cohesion in group therapy (Norcross & Lambert, 2018). Particular attention is given in this paper to the therapeutic alliance and empathy, because previous studies have shown promising results and because a clearer hierarchical model of therapeutic-relationship elements has been proposed, in which the therapeutic alliance, empathy, and cohesion are superordinate constructs (Horvath et al., 2016, as cited in Norcross & Lambert, 2018).

Empathy is defined by the APA Dictionary of Psychology (2023) as "the ability to understand the thoughts, feelings, motivations, or behaviors of another person from their perspective." Similarly, therapeutic alliance is defined as "the cooperative working relationship between client and therapist, including shared goals, consensus on tasks, and a strong interpersonal bond." A meta-analysis by Elliott et al. (2018) shows that empathy is a moderately strong predictor of psychotherapy outcomes, and the therapeutic alliance exhibits a solid positive association with outcomes (Flückiger et al., 2018). Although empathy and the therapeutic alliance are highlighted as central to the success of the therapeutic process, research shows that developing these core elements in psychologists under supervision requires more than conceptual understanding. A meta-analysis found that didactic instruction alone has a significant effect on empathy development, and that combining didactic components with rehearsal and observation yields even stronger results (Ngo et al., 2022). Similarly, although training programs report

awareness of the alliance's importance, only a few programs include evidence-based methods in their curricula. Best practices would integrate manuals, alliance-focused courses, observation of recorded sessions, and active feedback (Constantino et al., 2017).

However, despite such well-documented recommendations, educational psychology programs offer limited practical experience, resulting in novice therapists who feel inadequately prepared for real-world clinical challenges (Rosén, 2019). In this context, Rosén (2019) also emphasizes the importance of deliberate practice. Considering the recommendations made by researchers to prioritize active learning and behavioral rehearsal in psychotherapist training (Beidas & Kendall, 2010; Sudak & Brenner, 2024), we developed a VR-based training application for developing empathic skills and strengthening the therapeutic alliance during individual training, in order to reduce the gap that aspiring psychologists often feel between theory and practice. We chose virtual reality because it offers the opportunity to recreate the real world or to create different worlds that allow users to understand concepts and to perform certain tasks repeatedly, all within a safe environment (Chittaro & Ranon, 2007). Studies show that VR environments provide an effective way to build skills in mental-health interviewing and diagnosis through role-play (Lowell & Alshammari, 2019), enhance students' understanding of age-related health challenges and foster greater empathy toward older adults experiencing vision or hearing impairments or Alzheimer's disease (Dyer, Swartzlander, & Gugliucci, 2018), and improve communication confidence post-training (Sapkaroski et al., 2022), among other benefits. Effective VR learning tools should include feedback and reflection opportunities (Huang et al., 2024). We therefore intend to combine summative and formative feedback, which has been shown to improve learner outcomes (Ismail et al., 2022). Virtual reality interventions have been demonstrated in recent meta-analyses to effectively train healthcare professionals in the assessment and treatment of mental health disorders (Steen et al., 2024), outperforming traditional and other digital training methods in improving post-intervention knowledge and skills (Kyaw et al., 2019; Martingano et al., 2021). We identified two meta-analyses examining VR-based empathy training, although neither focuses specifically on psychotherapists or other healthcare professionals. One reported a statistically significant improvement in perspective-taking following VR exposure, with no notable change in overall empathy levels (Ventura et al., 2020), while the other showed that VR enhances emotional but not cognitive empathy—evoking compassionate feelings without encouraging users to adopt others' perspectives (Martingano et al., 2021).

Regarding the therapeutic alliance, we found no studies directly examining its development in VR-based training. However, existing research highlights its importance even within virtual environments (Varšová & Juřík, 2024; Meyerbröker & Emmelkamp, 2008) and demonstrates that therapeutic presence in VR can positively influence the working alliance (Im, Jo, & Lee, 2024). Despite evidence

supporting empathy and therapeutic alliance as key to effective psychotherapy, and the promise of VR for skill development, no studies have yet focused on using VR to train future psychotherapists in these essential relational skills. In this manuscript, the term future psychotherapists refers to psychology trainees, including university students and early-career therapists in supervised practice, whereas the VR application evaluated in this study is titled *Future Psychotherapists*.

Accordingly, the present pilot study had two aims. The first aim was to assess feasibility and usability of the VR application in future psychotherapists. The second aim was exploratory and examined pre–post changes in state empathy after a single VR session. Although the application targets both empathy and alliance-building skills, therapeutic alliance was not assessed with a standardized questionnaire in this pilot sample. Instead, alliance-related components were examined through expert content validation.

In line with the first aim, we expected low cybersickness, high perceived usability, and low-to-moderate perceived workload. For the secondary, exploratory aim, we expected pre–post changes in state empathy.

Methods

Design

This pilot study used a single-group pretest–posttest design to evaluate the virtual reality application *Future Psychotherapists*. The primary outcomes were feasibility and usability. Feasibility was operationalized as cybersickness measured both before and after the VR session using the Simulator Sickness Questionnaire (SSQ). Usability was evaluated post-session using the System Usability Scale (SUS) and the NASA Task Load Index (NASA-TLX) for perceived workload. The secondary outcome was exploratory and examined pre–post change in state empathy measured with the State Empathy Scale following a single VR session. We did not administer a therapeutic alliance measure, but alliance components were addressed during the expert validation phase.

Ethical Consideration

This study was approved by the Research Ethics Subcommittee of the University (Approval No. 470, dated 16.05.2024). Informed consent was obtained from all participants. All procedures involving human participants were conducted in accordance with the APA Ethical Principles of Psychologists and Code of Conduct.

Participants

The sample included 55 participants comprising undergraduate and master's-level psychology students as well as a small number of early-career therapists in supervised practice. Participants were recruited from the Faculty of Psychology based on the following inclusion criteria: (a) aged 18 or older; (b) currently enrolled in undergraduate or master's-level psychology programs or engaged in supervised practice as early-career therapists. Exclusion criteria included: (c) history of epilepsy, photosensitivity, or severe mental disorders; and (d) significant discomfort during VR tasks. The final sample consisted of 55 participants who completed all study stages.

Procedure

Participant recruitment was conducted through in-class announcements and a post on the Faculty's official communication channels, and interested individuals completed a Google Forms link that supported two stages: the scheduling phase and the participant selection phase. After registering for the study by providing informed consent and meeting the eligibility criteria, participants began with a brief familiarization session with the virtual reality application conducted in the research lab. They then viewed a standardized introductory clip within the VR-based training application (i.e. the first segment of the training; replayed at scenario onset) in which a virtual client presented the reason for attending a psychotherapy session. At this stage, participants only listened to the virtual client and then proceeded to the next step. Following this, the participants completed a set of baseline (T1) questionnaires: the Simulator Sickness Questionnaire (Kennedy et al., 1993) and the State Empathy Scale (Shen, 2010). Participants completed these questionnaires outside (desktop computer in the research lab) VR after the introductory clip to ensure all participants received the same client information before the baseline (T1) assessment and before returning to VR for the interactive task. The next step, they engaged in an experimental task in VR. This task consisted of a psychotherapy session in which participants interacted with a virtual client through an interactive scenario. At the end of the session, participants received feedback with both summative and formative components, based on their answers. If an incorrect answer was selected, the correct option was revealed to support learning. Finally, participants completed the post-intervention set of questionnaires: the Simulator Sickness Questionnaire (Kennedy et al., 1993), the State Empathy Scale (Shen, 2010), the NASA Task Load

Index (Hart & Staveland, 1988), and the System Usability Scale (Brooke, 1996). There was a single VR session, and the procedure lasted 30-45 minutes in total.

Virtual reality–based intervention training description

Participants experienced the VR scenario through a Meta Quest 2 (Meta Platforms, Inc., 2020) VR headset. The virtual environment simulates an individual therapy session. The client presents mild symptomatology (mild depression and relational conflicts). The clinical scenario was presented as pre-recorded 2D video segments displayed in the head-mounted display against a minimal 360° black background (cinema-like). Response alternatives were presented as 3D interface elements in front of the user at each decision node. A minimalist background was used to standardize exposure and minimize irrelevant contextual information.

The application's script is implemented as an interactive branching narrative via discrete video segments. In each segment, the virtual patient describes their symptomatology, and at each decision node the user selects the response they deem most appropriate for the client from the predefined options on screen (see Figure 1). For an example of a client dialogue and corresponding response options, see Appendix A. After they choose a response, the virtual client will give them a predefined response and move to the next video segment.

The responses were designed to target empathic exploration, empathic understanding, empathic affirmation, and therapeutic alliance–building tasks. In total, there were 16 decision nodes: 9 focused on empathy-building and 7 on therapeutic alliance development. After users completed all decision nodes, they received summative feedback detailing the number of correct responses for each targeted skill (see Figure 2). The application also provided a formative feedback component: users could reopen the corresponding video clip where an incorrect response was selected, listen to it again, and view the expert-informed correct option (see Figure 3). In the review screen, the expert-informed correct option was highlighted in green and the participant's selected incorrect option in red, enabling direct comparison. After reviewing this feedback, users could restart the scenario and repeat it as many times as they wished to support deliberate practice, although in the present study, participants completed only a single VR session to assess initial feasibility. However, the application did not provide an explicit textual rationale explaining why one option was preferable and users were expected to derive the rationale by comparing their choice with the exemplar response. The application was developed using Unity (version 2019) and the clips were filmed in 4K at 30 fps with spatialized audio.

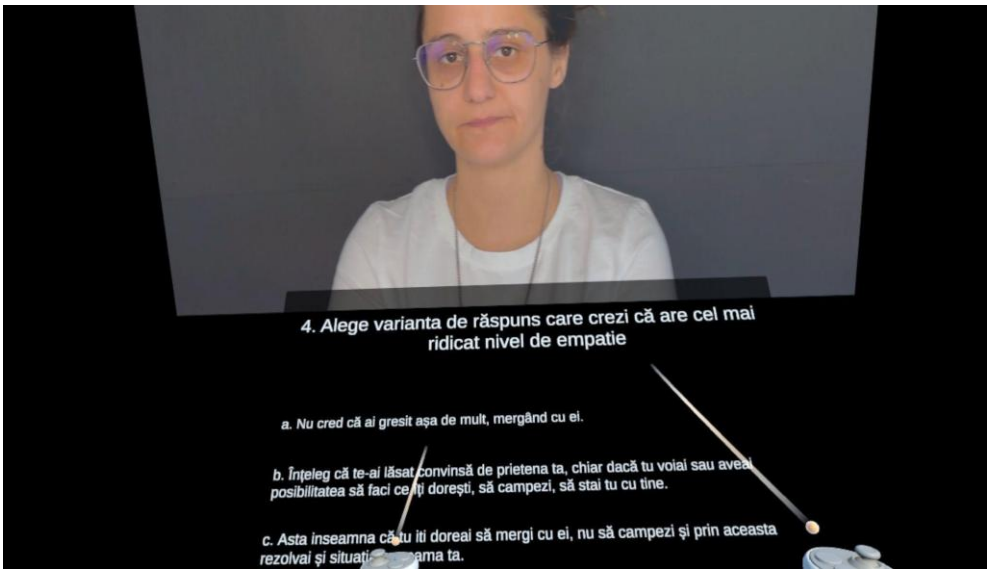

Figure 1. Example of a decision node in the branching narrative.
The virtual patient presents a symptom, and the user selects the most appropriate response from multiple predefined options. The interface is in Romanian.

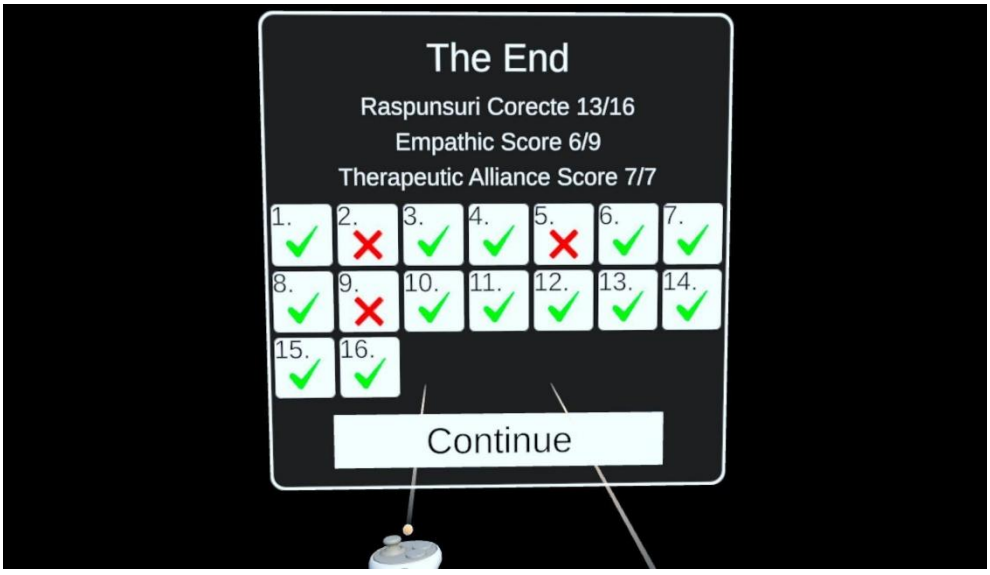

Figure 2. Summative feedback screen displaying the number of correct responses for each targeted skill.

Figure 3. Formative component feedback screen allowing users to review video segments in which incorrect responses were selected, along with the correct choices. The interface is in Romanian.

Expert contributors

In a first phase, two expert psychologists consulted on the most important elements of the therapeutic relationship that would be suitable for inclusion in a VR-based intervention targeting the development of relational therapeutic skills. A group of five expert psychologists then completed a two-part questionnaire to evaluate the application's textual content (the virtual client's speech and the response options). Following the rationale for expert content validation highlighted by different researchers (Palter et al., 2012; Ceberio et al., 2025), expert evaluations were included to establish a professional "gold standard" for the training scenarios. Since students are novices, expert validation was essential to ensure that the branching narrative options were clinically sound and provided a benchmark for proficiency that students cannot establish alone. Part A focused on content validation of the scenario and response options, whereas Part B captured experts' judgments regarding the application's educational/training potential rather than training efficacy.

Measures

We measured cybersickness using the *Simulator Sickness Questionnaire* (SSQ; Kennedy et al., 1993). It is the most commonly used instrument for measuring sickness induced by virtual reality environments (Bimberg et al., 2020). The SSQ comprises 16 items rated on a 4-point Likert scale (0 = None to 4 = Severe) and is grouped into three subscales—nausea, oculomotor disturbance, and disorientation—and a total score is calculated (Kennedy et al., 1993). We calculated internal consistency with Cronbach's alpha for each subscale. The results show good reliability: $\alpha = .72$ for nausea, $\alpha = .77$ for oculomotor, $\alpha = .83$ for disorientation, and $\alpha = .87$ for total score.

For assessing the usability of the system, the participants were asked to complete the *System Usability Scale* (SUS; Brooke, 1996), which has 10 items that measure the perceived usability on a 5-point Likert scale (1 = strongly disagree; 5 = strongly agree). SUS is a reliable tool for assessing the usability of a wide variety of products and services and generates an overall score on a scale from 0 to 100. We didn't calculate internal consistency because SUS is designed to provide a single overall usability score. The results were interpreted using the cut-off score of 68.

NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988) was used to measure the cognitive load of participants, thus enabling us to see whether the application is easy to use from this point of view. It has 6 subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration, and a very specific method of calculating the score that requires good knowledge in using the instrument. It is one of the most widely used scales for measuring cognitive load (Rutledge et al., 2009). We did not calculate Cronbach's alpha for the NASA-TLX, as the scale is designed to measure multiple distinct aspects of workload, rather than a single underlying construct. Although the NASA-TLX does not include standardized cut-off points, previous research suggests that workload scores can be interpreted as follows: 0–20 = very low; 21–40 = low; 41–60 = moderate; 61–80 = high; and 81–100 = very high (Febiyani et al., 2021).

Participants' empathy was measured with the *State Empathy Scale* (SES; Shen, 2010). The scale includes three subscales—affective empathy, cognitive empathy, and associative empathy—and containing 12 items rated on a 5-point Likert scale (1 = not at all; 5 = completely). The instrument has high psychometric properties (Shen, 2010). Higher scores reflected a stronger level of empathy in participants. Internal consistency was evaluated with Cronbach's alpha for each subscale. The results present good reliability: $\alpha = .84$ for affective empathy, $\alpha = .84$ for cognitive empathy, and $\alpha = .72$ for associative empathy.

For measuring the quality of the application content's text section, we developed an instrument composed of two parts:

Part A: Evaluation of the therapeutic response texts on a 5-point Likert scale (1 = not at all, 5 = very much), measuring the degree to which each response exemplifies empathic skills (exploration, understanding, affirmation) or alliance-building behaviors (validation, collaboration, affective support) (Example item: "Read the text below and indicate to what extent the psychotherapist's response reflects an empathic response or a response that can build the therapeutic alliance.").

Part B: Global assessment of content clarity, relevance, and pedagogical utility through five Likert-scale items (1–5) addressing overall text adequacy, dialogue realism, scenario relevance, and potential for skill development (Example item: "The dialogues in the script are clear and easy to understand.").

We calculated internal consistency using Cronbach's alpha for each subscale. The results show different levels of reliability: $\alpha = .724$ for empathy, $\alpha = .78$ for therapeutic alliance, and $\alpha = .833$ for global assessment from part B.

Data Analysis

All data were processed using SPSS (version 20). Descriptive analyses (i.e., means, SDs) were conducted for all variables. Feasibility (SSQ), measured pre- and post-session, and usability outcomes (SUS and NASA-TLX), measured post-session only, were summarized descriptively. For the exploratory outcome, we used paired-samples t-tests to examine pre–post changes in state empathy. The significance level was set at $p < .05$. The expert evaluation, inter-rater agreement was assessed using Kendall's coefficient of concordance (Kendall's W), and expert ratings were summarized descriptively.

Results

Feasibility (Cybersickness)

Descriptive statistics for cybersickness subscales at pre-intervention (T1) and post-intervention (T2) are presented in Table 1. We can observe only slight increases for cybersickness (nausea, oculomotor disturbance, disorientation, and total SSQ score). To examine whether these changes were statistically significant, we conducted paired-samples *t*-tests. As shown in Table 1, there were no significant changes in cybersickness. Specifically, the scores for nausea ($t(54) = -0.16$, $p = .873$), oculomotor disturbance ($t(54) = -1.62$, $p = .111$), disorientation ($t(54) = -0.96$, $p = .340$), and the total SSQ score ($t(54) = -0.98$, $p = .334$) showed no significant differences.

Table 1. Means and Standard Deviations at T1 and T2 and Paired-Samples t-Tests (SSQ)

Subscale	T1 *M* (*SD*)	T2 *M* (*SD*)	*t*(*df*)	*p*
Nausea (SSQ)	10.23 (17.40)	10.93 (28.29)	−0.16 (54)	.873
Oculomotor Disturbance (SSQ)	15.02 (18.13)	20.53 (25.04)	−1.62 (54)	.111
Disorientation (SSQ)	20.25 (35.82)	26.32 (46.61)	−0.96 (54)	.340
Total Cybersickness Score (SSQ)	16.86 (23.23)	21.49 (34.17)	−0.98 (54)	.334

Note. T1 = Pre-intervention; T2 = Post-intervention; *SD* = Standard Deviation; SSQ = Simulator Sickness Questionnaire (Kennedy et al., 1993). Higher scores indicate higher levels on each construct. Negative *t* values indicate higher post-intervention scores.

Usability and Workload (Post-intervention)

Furthermore, descriptive statistics for the post-intervention measures (SUS and NASA-TLX) are presented in Table 2. Usability (SUS) was rated very well, with a mean score of $M = 85.09$ ($SD = 9.54$), with scores above 68 (cut-off) indicating that the application is perceived as well-designed and easy to use. Regarding participants' perceived workload, we used the NASA-TLX. The results show that most participants reported low to moderate workload levels ($M = 32.50$, $SD = 17.13$).

Table 2. Descriptive Statistics for Post-Intervention Measures (T2; SUS, NASA-TLX)

Variable	*N*	Minimum	Maximum	*M*	*SD*
System Usability Scale (SUS)	55	57.50	100.00	85.09	9.54
NASA Task Load Index (NASA-TLX)	55	4.00	65.00	32.50	17.13

Note. SUS = System Usability Scale (Brooke, 1996); NASA-TLX = NASA Task Load Index (Hart & Staveland, 1988); *SD* = Standard Deviation. Higher SUS scores indicate better perceived usability. Lower NASA-TLX scores indicate lower perceived workload.

Expert Content Evaluation (Content validity)

The content validity of the application was assessed by five independent experts. Our purpose was expert consensus, not statistical inference. They evaluated empathy-related content, therapeutic alliance content, and general content quality. Table 3 presents descriptive statistics. The empathy-related content received a mean score of $M = 4.40$ ($SD = 0.30$). These results indicate high perceived relevance and clarity of the script used to foster state empathy. Moreover, the standard deviation indicates a high level of agreement between experts. The evaluation of the therapeutic alliance content also received a high mean score ($M = 4.00$, $SD = 0.17$), indicating that the VR application may facilitate the development of a therapeutic alliance. The content quality (Part B) received the highest ratings with a mean of *M*

= 4.76 (*SD* = 0.33). This suggests that experts considered the application content to be appropriate, coherent, and clinically relevant. Given the small variability across raters and consistently high scores across all categories, these results support the strong content validity of the application from a professional standpoint. This conclusion is reinforced by the high inter-rater agreement, Kendall's $W = .96$, $\chi^2(2) = 9.58$, $p = .008$, indicating consistent expert ratings across dimensions.

Table 3. Descriptive Statistics for Empathy, Therapeutic Alliance, and Content Evaluation

Variable	*N*	Minimum	Maximum	*M*	*SD*
Total Empathy	5	4.00	4.78	4.40	0.30
Total Therapeutic Alliance	5	3.86	4.29	4.00	0.17
Content Score	5	4.20	5.00	4.76	0.33

Note. SD = Standard Deviation. Scores are based on a 1–5 Likert scale where higher values indicate more favorable evaluations.

Empathy Outcomes (Exploratory)

Table 4 summarizes empathy scores at pre-intervention (T1) and post-intervention (T2). Mean scores increased from T1 to T2 for all three empathy components. To examine whether these changes were statistically significant, we conducted paired-samples *t*-tests. As shown in Table 4, data indicated significant improvements in all three dimensions of empathy: affective empathy ($t(54) = -4.11$, $p < .001$), cognitive empathy ($t(54) = -4.71$, $p < .001$), and associative empathy ($t(54) = -3.56$, $p = .001$).

Table 4. Means and Standard Deviations at T1 and T2 and Paired-Samples t-Tests (SES)

Subscale	T1 *M* (*SD*)	T2 *M* (*SD*)	*t*(*df*)	*p*
Affective Empathy (SES)	14.20 (3.12)	16.22 (3.10)	−4.11 (54)	< .001
Cognitive Empathy (SES)	17.56 (2.35)	19.13 (1.28)	−4.71 (54)	< .001
Associative Empathy (SES)	15.02 (2.86)	16.69 (2.48)	−3.56 (54)	.001

Note. T1 = Pre-intervention; T2 = Post-intervention; *SD* = Standard Deviation; SES = State Empathy Scale (Shen, 2010). Higher scores indicate higher levels on each construct. Negative *t* values indicate higher post-intervention scores.

Discussion

The present study investigated the feasibility and usability of the Future Psychotherapists application, and secondarily examined changes in state empathy

from pre-intervention (T1) to post-intervention (T2). The VR-based training was developed to support future psychotherapists in cultivating therapeutic relationship skills.

The low level of cybersickness observed in our application aligns with findings from the literature indicating that increasing cybersickness can compromise in-VR performance and, by extension, the feasibility of VR-based training. Symptom severity has been associated with poorer visuospatial working memory and degraded psychomotor performance (e.g., slower reaction times) during immersion (Kourtesis et al., 2024). Moreover, evidence suggests that VR sickness can reduce attention and impair task performance (Mimnaugh et al., 2023). In addition, VR sickness is associated with non-completion of exposure: a systematic review and meta-analysis reported a mean dropout rate of 15.6% attributable to VR sickness (Saredakis et al., 2020), which may reduce effective exposure to training sessions. Taken together, the low cybersickness observed in our sample suggests that simulator sickness is unlikely to have meaningfully interfered with task performance or engagement, supporting the feasibility of delivering this VR training in a single-session format.

Regarding the usability of our application, participants reported high usability and low-to-moderate cognitive workload. High scores on the System Usability Scale (>68) suggest that participants found the interface easy to use, which is essential for user engagement in virtual learning environments. Moreover, the fact that participants reported low to moderate workload aligns with the research of Kirschner (2002), which affirms that an appropriate workload when using VR is preferable because it helps participants optimize cognitive load for learning.

Moreover, expert evaluations confirmed the application's potential to train empathy and therapeutic alliance. From a content perspective, expert validation highlighted the application's clarity, relevance, and clinical usefulness. Kendall's W = .96 indicates excellent agreement among raters.

In relation to our exploratory objective regarding empathy, a single VR training session significantly improved state empathy (cognitive, affective, associative) without increasing cybersickness. Regarding cognitive and affective empathy, our results align with previous research showing that VR can enhance perspective-taking, and at the same time, with the findings reported in the meta-analysis by Martingano et al. (2021). As for associative empathy, this dimension has been less studied. However, from a functional perspective, it plays an important role in fostering social bonding and the development of relationships (Davis, 1994), acting as a bridge between perception and action in the emergence of state empathy (Shen, 2010). Looking at all these results, we can conclude that the current VR-based training application can facilitate multiple layers of empathic response, which is likely to be effective in developing core relational skills.

The VR environment was intentionally designed in a minimalist manner to reduce extraneous processing and keep attention on the instructional objective (i.e.,

to choose the best response option in terms of empathy and therapeutic alliance), in line with multimedia learning principles such as the coherence principle (Fiorella & Mayer, 2021). We therefore prioritized a clear and well-timed presentation of task-relevant information over contextual environmental realism (Fiorella & Mayer, 2021). Recent evidence from educational VR, including technical and vocational contexts, suggests that within immersive 3D VR environments, increasing environmental realism does not necessarily improve learning performance relative to more minimalist designs and may even support motivation by reducing distractions (Larmuseau et al., 2025). Moreover, CAMIL suggests that learning outcomes depend on interactions among presence/agency, affective responses, and cognitive load, rather than on visual fidelity alone (Makransky & Petersen, 2021). Although our implementation relied primarily on video presented in an HMD, this minimalist approach was expected to influence the training effects by reducing competing stimuli and helping participants focus on identifying and selecting the most appropriate therapeutic response.

Considering the results of this study regarding the feasibility and usability of the application, as well as the experts' evaluation of the application's content and the observed change from T1 to T2, these findings contribute to the expanding body of literature supporting the use of VR-based interventions in mental health education (Steen et al., 2024; Kyaw et al., 2019) for training key elements of the psychotherapeutic relationship, such as empathy and therapeutic alliance, among students or novice practitioners. Both formative and summative feedback components reinforce the idea of an application capable of fostering empathy and the therapeutic alliance among future psychologists and early-career practitioners, while also enabling deliberate practice through repeated scenario completion. For example, Sapkaroski et al. (2022) found that immersive VR training improved students' confidence and communication abilities, while Dyer et al. (2018) showed increased empathy in medical students after exposure to aging simulations. Practically, this method can serve as a complementary tool in the training of psychotherapists, responding to current gaps in experiential learning and supporting the development of relational competencies through active engagement in simulated clinical scenarios.

Limitations and future directions

The study has several limitations that should be acknowledged. One limitation is the fact that the pretest-posttest design does not include a control group. This was a deliberate decision, as the main objective at this exploratory phase was to assess the VR application's usability—according to ISO 9241-11 (International Organization for Standardization [ISO], 2018) and its overall feasibility for single-session VR training, rather than to test its efficacy compared to alternative interventions. In line with this definition, our evaluation prioritized feasibility and

user experience, assessing whether users could successfully engage with and complete the training tasks, rather than providing evidence of superior training outcomes. Hence, causal inferences should be made with caution. The second limitation refers to the fact that the intervention was conducted through a single VR session, and therefore, the durability of the effects over time remains unknown. Another limitation is that, although deliberately minimalist, the video-based VR format—consisting primarily of a 2D video presented in an HMD against a black 360° background—may yield lower immersion/presence and fewer contextual cues than a shared, fully modeled virtual therapy room. This should be considered when interpreting the extent to which the observed effects generalize to spontaneous, real-world therapist–client interaction, and should be examined in future research. An additional limitation is that the feedback is primarily evaluative (correct/incorrect), accompanied by an expert-informed exemplar, without a detailed rationale explaining why one option is preferable. Although this structure supports guided practice through repeated scenario completion and provides both formative and summative feedback on empathy- and alliance-consistent responding, deeper reflective learning and transfer to open-ended clinical communication were not established in this pilot. Although the application was designed to foster both empathy and therapeutic alliance, we did not include a standardized measure for alliance. This decision was primarily based on the fact that most participants were psychology students without direct client experience. Although a small number of early-career therapists in supervised practice were included, this subgroup was too small to support separate analyses. Therefore, the findings should be interpreted as primarily reflecting psychology students. Since existing alliance questionnaires typically require respondents to reflect on prior therapeutic interactions, they would not have been appropriate for this population. Finally, future studies should incorporate more behaviorally grounded outcome measures (e.g., observer-rated performance, behavioral coding of responses in role-plays or standardized patient interactions, and performance-based tasks requiring open-ended responding) to evaluate training efficacy beyond self-report measures. These limitations point toward the need for future research in which VR-based training is used in more complex studies, such as randomized controlled trials, and potentially in a longitudinal design as well.

Conclusions

This study demonstrates the feasibility and usability of the VR-based training application for relational skills in future psychotherapists (i.e., psychology students). The application was rated as usable, engaging, and with a low to moderate level of participants' workload. Moreover, the VR intervention produced significant

changes in empathy across all three dimensions (cognitive, affective, and associative). All these aspects suggest that the current VR-based training application can complement traditional methods, so that future psychotherapists have more learning opportunities for the skills that currently represent a significant gap between theory and practice. Given that the results are preliminary, as this is a pilot study, future research should include larger and more diverse samples, control groups, and longitudinal follow-up to evaluate the long-term impact of VR-based relational skills training. To the best of our knowledge, this study introduces a novel educational tool that can serve as a model for future innovations in psychotherapy education and professional development.

Author's Notes

Competing Interests: The authors declare no conflicts of interest. Dana Cristolțan is the administrator of Kind Strategies SRL. The company had no roles in the design of the study, in the collection, analysis or interpretation of data; in the writing of the manuscript, or in the decision to publish the articles. The paper reflects the views of the scientists and not the company.

Funding: The application used in this research was supported by Kind Strategies SRL.

References

American Psychological Association. (2023). Empathy. In *APA Dictionary of Psychology*. https://dictionary.apa.org/empathy

American Psychological Association. (2023). Therapeutic alliance. In *APA Dictionary of Psychology*. https://dictionary.apa.org/therapeutic-alliance

Beidas, R. S., & Kendall, P. C. (2010). Training therapists in evidence-based practice: A critical review of studies from a systems-contextual perspective. *Clinical Psychology: Science and Practice, 17*(1), 1–30. https://doi.org/10.1111/j.1468-2850.2009.01187.x

Bimberg, P., Weissker, T., & Kulik, A. (2020). On the usage of the Simulator Sickness Questionnaire for virtual reality research. In *Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 464–467). IEEE. https://doi.org/10.1109/VRW50115.2020.00098

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, I. L. McClelland, & B. Weerdmeester (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis.

Ceberio, I., Al-Rashaida, M., García, M., Lopez Paz, J. F., Salgueiro, M., Passi, N., Pavel, H., & Amayra, I. (2025). Content and face validity in virtual reality with children: A validation in five steps+1 of a wheelchair basketball game. *Frontiers in Virtual Reality, 5*, 1505630. https://doi.org/10.3389/frvir.2024.1505630

Chittaro, L., & Ranon, R. (2007). Web3D technologies in learning, education and training: Motivations, issues, opportunities. *Computers & Education, 49*(1), 3–18. https://doi.org/10.1016/j.compedu.2005.06.002

Constantino, M. J., Morrison, N. R., Coyne, A. E., & Howard, T. (2017). Exploring therapeutic alliance training in clinical and counseling psychology graduate programs. *Training and Education in Professional Psychology, 11*(4), 219–226. https://doi.org/10.1037/tep0000157

Davis, D. (1994). *Empathy: A social psychological approach*. Brown & Benchmark Publishers.

Dyer, E., Swartzlander, B. J., & Gugliucci, M. R. (2018). Using virtual reality in medical education to teach empathy. *Journal of the Medical Library Association, 106*(4), 498–500. https://doi.org/10.5195/jmla.2018.518

Elliott, R., Bohart, A. C., Watson, J. C., & Murphy, D. (2018). Therapist empathy and client outcome: An updated meta-analysis. *Psychotherapy, 55*(4), 399–410. https://doi.org/10.1037/pst0000175

Febiyani, A., Febriani, A., & Ma'Sum, J. (2021). Calculation of mental load from e-learning student with NASA TLX and SOFI method. *Jurnal Sistem dan Manajemen Industri, 5*(1), 35–42. https://doi.org/10.30656/jsmi.v5i1.2789

Fiorella, L., & Mayer, R. E. (2021). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 185–198). Cambridge University Press. https://doi.org/10.1017/9781108894333.019

Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy, 55*(4), 316–340. https://doi.org/10.1037/pst0000172

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). North-Holland.

Huang, Y.-M., Wang, W.-S., Lee, H.-Y., Lin, C.-J., & Wu, T.-T. (2024). Empowering virtual reality with feedback and reflection in hands-on learning: Effect of learning engagement and higher-order thinking. *Journal of Computer Assisted Learning, 40*, 1413–1427. https://doi.org/10.1111/jcal.12959

Im, S., Jo, D., & Lee, S. M. (2024). Exploring the impact of therapeutic presence on working alliance in Metaverse counseling. *The Counseling Psychologist, 52*(5), 698–724. https://doi.org/10.1177/00110000241234630

International Organization for Standardization. (2018). *Ergonomics of human-system interaction—Part 11: Usability: Definitions and concepts (ISO 9241-11:2018).* https://www.iso.org/standard/63500.html

Ismail, S. M., Rahul, D. R., Patra, I., & Rezvani, E. (2022). Formative vs. summative assessment: Impacts on academic motivation, attitude toward learning, test anxiety, and self-regulation skill. *Language Testing in Asia, 12*(1), Article 40. https://doi.org/10.1186/s40468-022-00191-4

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3*(3), 203–220.

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction, 12*(1), 1–10. https://doi.org/10.1016/S0959-4752(01)00014-7

Kourtesis, P., Papadopoulou, A., & Roussos, P. (2024). Cybersickness in virtual reality: The role of individual differences, its effects on cognitive functions and motor skills, and intensity differences during and after immersion. *Virtual Worlds, 3*(1), 62–93. https://doi.org/10.3390/virtualworlds3010004

Kyaw, B., Saxena, N., Posadzki, P., Vseteckova, J., Nikolaou, C., George, P., Divakar, U., Masiello, I., Kononowicz, A., Zary, N., & Tudor Car, L. (2019). Virtual reality for health professions education: Systematic review and meta-analysis by the Digital Health Education Collaboration. *Journal of Medical Internet Research, 21*(1), e12959. https://doi.org/10.2196/12959

Lambert, M. J., & Barley, D. E. (2001). Research summary on the therapeutic relationship and psychotherapy outcome. *Psychotherapy: Theory, Research, Practice, Training, 38*(4), 357–361. https://doi.org/10.1037/0033-3204.38.4.357

Larmuseau, C., De Leersnijder, L., Rotsaert, T., Boel, C., Demanet, J., & Schellens, T. (2025). Beyond realism: Rethinking VR design for optimal learning in technical and vocational secondary education. *Computers & Education: X Reality, 6*, 100098. https://doi.org/10.1016/j.cexr.2025.100098

Lowell, V. L., & Alshammari, A. (2019). Experiential learning experiences in an online 3D virtual environment for mental health interviewing and diagnosis role-playing: A comparison of perceived learning across learning activities. *Educational Technology Research and Development, 67*(4), 825–854. https://doi.org/10.1007/s11423-018-9632-8

Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review, 33*, 937–958. https://doi.org/10.1007/s10648-020-09586-2

Martingano, A. J., Hererra, F., & Konrath, S. (2021). Virtual reality improves emotional but not cognitive empathy: A meta-analysis. *Technology, Mind, and Behavior, 2*(1). https://doi.org/10.1037/tmb0000034

Meta Platforms, Inc. (2020). *Meta Quest 2* [Virtual reality headset]. https://www.meta.com/quest/

Meyerbröker, K., & Emmelkamp, P. M. (2008). Therapeutic processes in virtual reality exposure therapy: The role of cognitions and the therapeutic alliance. *CyberTherapy & Rehabilitation, 1*(3), 247–257.

Mimnaugh, K. J., Center, E. G., Suomalainen, M., Becerra, I., Lozano, E., Murrieta-Cid, R., Ojala, T., LaValle, S. M., & Federmeier, K. D. (2023). Virtual reality sickness reduces attention during immersive experiences. *IEEE Transactions on Visualization and Computer Graphics, 29*(11), 4394–4404. https://doi.org/10.1109/TVCG.2023.3320222

Ngo, H. L., Sokolović, N., & Jenkins, J. M. (2022). Teaching empathy to mental health practitioners and trainees: Pairwise and network meta-analyses. *Journal of Consulting and Clinical Psychology, 90*(11), 851–860. https://doi.org/10.1037/ccp0000773

Norcross, J. C., & Karpiak, C. P. (2025). Psychotherapy relationships. In S. Hupp & D. F. Tolin (Eds.), *Science-based therapy: Raising the bar for empirically supported treatments* (pp. 381–398). Cambridge University Press.

Norcross, J. C., & Lambert, M. J. (2011). Evidence-based therapy relationships. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed., pp. 3–22). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199737208.003.0001

Norcross, J. C., & Lambert, M. J. (2018). Psychotherapy relationships that work III. *Psychotherapy, 55*(4), 303–315. https://doi.org/10.1037/pst0000193

Palter, V. N., Graafland, M., Schijven, M. P., & Grantcharov, T. P. (2012). Designing a proficiency-based, content validated virtual reality curriculum for laparoscopic colorectal surgery: A Delphi approach. *Surgery, 151*(3), 391–397. https://doi.org/10.1016/j.surg.2011.07.084

Rosén, E. (2019). *Learning how to learn: Psychology trainees' self-compassion while implementing deliberate practice, with FIT at a psychotherapy training clinic* [Master's thesis, Umeå University].

Rutledge, T., Stucky, E., Dollarhide, A., Shively, M., Jain, S., Wolfson, T., Weinger, M. B., & Dresselhaus, T. (2009). A real-time assessment of work stress in physicians and nurses. *Health Psychology, 28*(2), 194–200. https://doi.org/10.1037/a0013145

Sapkaroski, D., Mundy, M., & Dimmock, M. R. (2022). Immersive virtual reality simulated learning environment versus role-play for empathic clinical communication training. *Journal of Medical Radiation Sciences, 69*(1), 56–65. https://doi.org/10.1002/jmrs.555

Saredakis, D., Szpak, A., Birckhead, B., Keage, H. A. D., Rizzo, A., & Loetscher, T. (2020). Factors associated with virtual reality sickness in head-mounted displays: A systematic review and meta-analysis. *Frontiers in Human Neuroscience, 14*, 96. https://doi.org/10.3389/fnhum.2020.00096

Shen, L. (2010). On a scale of state empathy during message processing. *Western Journal of Communication, 74*(5), 504–524. https://doi.org/10.1080/10570314.2010.512278

Steen, C. W., Söderström, K., Stensrud, B., Nylund, I. B., & Siqveland, J. (2024). The effectiveness of virtual reality training on knowledge, skills and attitudes of health care professionals and students in assessing and treating mental health

disorders: A systematic review. *BMC Medical Education, 24*(1), 480. https://doi.org/10.1186/s12909-024-05423-0

Sudak, D. M., & Brenner, A. M. (2024). The future of psychotherapy training. *Academic Psychiatry, 48*(5), 511–515. https://doi.org/10.1007/s40596-024-02022-5

Varšová, K., & Juřík, V. (2024). Fostering therapeutic alliance and long-term benefits through virtual collaboration in VRET. *International Psychological Applications Conference and Trends (InPACT2024).*

Ventura, S., Badenes-Ribera, L., Herrero, R., Cebolla, A., Galiana, L., & Baños, R. (2020). Virtual reality as a medium to elicit empathy: A meta-analysis. *Cyberpsychology, Behavior, and Social Networking, 23*(10), 667–676. https://doi.org/10.1089/cyber.2019.0681