Academic Psychiatry
https://doi.org/10.1007/s40596-023-01909-z

EMPIRICAL REPORT

# Using Virtual Patients to Assess and Improve Clinicians' Emotional Self-awareness: a Randomized Controlled Study

**Giuseppe Sarli[1] · Megan L. Rogers[2] · Sarah Bloch-Elkouby[1,3] · Olivia C. Lawrence[1] · Alexandre Gomes de Siqueira[4] · Heng Yao[4] · Benjamin Lok[4] · Adriana Foster[5] · Igor Galynker[1]**

Received: 6 March 2023 / Accepted: 18 November 2023


## Abstract

**Objective** Working with suicidal patients can elicit negative emotional responses that can impede clinicians' empathy and affect clinical outcomes. Virtual human interactions represent a promising tool to train clinicians. The present study investigated the impact of virtual human interaction training to enhance clinicians' emotional self-awareness and empathy when working with suicidal patients.

**Methods** Clinicians were randomly assigned into two groups. Both groups interviewed a virtual patient presenting with a suicidal crisis; clinicians in the intervention condition ($n = 31$) received immediate feedback about negative emotional responses and empathic communication, whereas those in the control condition ($n = 33$) did not receive any feedback. All clinicians interviewed a second virtual patient 1 week later. Clinicians' emotional response to the two virtual patients and their empathic communication with each of them were assessed immediately after each interaction. Linear mixed models were used to assess change in clinicians' emotional response and verbal empathy between the two interactions across conditions.

**Results** Clinicians' emotional responses toward the suicidal virtual patients were unchanged in both conditions. Clinicians in the intervention condition presenting low empathy level with the first virtual patient showed higher empathy level with the second virtual patient than with the first ($B = 1.15$, $SE = 0.25$, $p < 0.001$, 95% CI [0.42, 1.89]).

**Conclusions** This work demonstrates the feasibility of using virtual human interactions to improve empathic communication skills in clinicians with poor empathy skills. Further refinement of this methodology is needed to create effective training modules for a broader array of clinicians.

**Keywords** Suicide · Virtual human interaction (VHI) · Virtual patient (VP) · Empathy · Emotional self-awareness

Despite the presence of extensive research and suicide prevention initiatives [1], suicide still represents a leading cause of death worldwide. According to data from the Centers for Disease Control and Prevention, suicide was responsible for over 47,000 deaths in the USA in 2021 [2]. Suicide rates increased by 30% from 1999 to 2018 in the USA [3]. Patients' contact with health care providers prior to suicide is a common event, with recent research showing that 44% of patients who died by suicide were seen by a clinician in the month preceding their suicide [4]. Patients at high risk for suicide oftentimes have difficulty connecting with and trusting their clinician [5], with up to two-thirds of patients who die by suicide not reporting suicidal ideation or intent during their last visit with their clinician [6]. It is possible that suicidal ideation may not be present at the time of the visit, given the fluctuating nature of suicidal thoughts [7]; however, it may also be possible to reduce nondisclosure of suicide risk by enhancing clinician-patient relationships. Clinician-patient encounters, thus, represent a crucial opportunity for clinicians to conduct a comprehensive assessment of suicide risk and engage in life-saving evidence-based interventions [8, 9].

✉ Giuseppe Sarli
giuseppesarli194@gmail.com

1 Icahn School of Medicine at Mount Sinai, New York, NY, USA

2 Texas State University, San Marcos, TX, USA

3 Ferkauf Graduate School of Psychology, Yeshiva University, New York, NY, USA

4 University of Florida, Gainesville, FL, USA

5 Hospital Corporation of America Florida Woodmont Hospital, Tamarac, FL, USA

Published online: 06 December 2023

Clinicians have been documented to experience negative emotional responses toward their patients at high risk of suicide, such as anger, disinterest, anxiety, inadequacy, confusion, rejection, and distress [10–12], which can lead to empathic failures and prevent clinicians from developing a strong therapeutic alliance [13] with their patients [12]. In addition, patients' perception of their clinicians' negative emotions and empathic failures are likely to exacerbate their own anxiety level, leading to negative interpersonal vicious cycles and ruptures in the therapeutic alliance [14]. Herein, we focus on an educational approach aiming to increase clinicians' ability to identify and regulate their negative emotions toward patients at risk of suicide in real time (referred to as emotional self-awareness in this paper) and to enhance clinicians' ability to convey empathy toward their patients.

High empathetic communication with patients at risk of suicide could be particularly difficult to achieve [12] and could be challenging to measure. A method to operationalize empathy in clinician-patient encounters, extensively used by our group in prior research [15, 16], is the Empathic Communication Coding System, an interaction analysis system in which the patients' statements that trigger clinicians' empathy are categorized as emotion, progress, and challenge empathic opportunities. Clinicians' responses to these opportunities can be described in seven categories, from complete denial of the patient's opportunity to confirmation or even sharing of similar experience with that of patients [17]. Latter research on empathy collapsed the clinicians' empathic responses to the patients to two categories, low empathy and high empathy [18].

Although clinical training often includes various combinations of didactic, experiential [19], and deliberate practice tools [20], virtual human interactions provide promising tools to train clinicians in emotional self-awareness and empathetic communication in a controlled environment. This technology consists in a live clinician interacting with a virtual human subject, more specifically, with a virtual patient. From the early use of virtual patients to promote clinical reasoning in health care [21], clinicians' interactions with virtual humans have expanded to teach interprofessional communication in health care, with a focus on empathy [22].

Previous work has demonstrated that real and virtual patients elicit comparable responses from clinicians. Namely, in several studies, the verbal empathy displayed by clinicians in training toward virtual patients with cranial nerve injury [22], depression, and bipolar disorder [23, 24] was comparable to that exhibited with real patients. Thus, training clinicians with virtual patients offers significant benefits over traditional training with real patients. First, virtual patients can embody a wide range of demographic characteristics and clinical presentations, thus allowing clinicians to interact with patients that they do not often encounter in their practice [24]. Second, virtual human interactions provide a fully controlled training environment [24] in which clinicians do not take the risk of being liable for their mistakes. In other words, virtual interactions present clinicians with the best of two worlds: virtual patients whose presentation and symptoms are very similar to real patients, together with absolute safety. Third, this approach provides clinicians with ample time to think, reflect, and make clinical decisions. Fourth, virtual human interaction-based trainings can be standardized and thus provide the same training opportunities worldwide at a very low cost, allowing training sites to save much-needed resources.

The current study strived to fill this gap and investigate the impact of a virtual human-based clinical training on clinicians' emotional awareness and empathy with suicidal patients. Specifically, the aim of the present study was to assess the impact of a virtual human-based clinical training on clinicians' emotional self-awareness and verbal empathy during interactions with virtual humans representing patients at risk of suicide. We hypothesized that, compared to controls, clinicians in the intervention condition (who received feedback about their negative emotional responses and verbal empathy immediately after interacting with a virtual patient at time 1) would report fewer negative emotional responses (hypothesis 1) and would display greater verbal empathy (hypothesis 2) when interacting with another virtual patient 1 week later (time 2).

## Methods

The current study is a prospective, double-blinded, randomized trial. The total sample included 64 clinician participants from four psychiatric outpatient recruitment sites, three within Mount Sinai Health System (MSHS) and one from Florida International University, affiliated with Citrus Health Network, Inc. The clinician participants were second- to fourth-year psychiatry residents and psychiatry attendings; psychology interns, postdoctoral trainees, and licensed psychologists; masters-level trainees in social work, mental health counseling, and licensed social workers; and nurses with master of science degrees in nursing. Clinician participants were recruited through email outreach using available hospital directories on behalf of research staff and via announcement at departmental rounds biannually.

Inclusion criteria for the clinician participants were being a staff clinician or trainee in one of the psychiatry outpatient clinics at any participating site and having a caseload of 5–10 patients meeting inclusion criteria for patient participants. This last requirement stemmed from the fact that the current study was embedded in a larger study that assessed real patients' outcomes within clinicians. Patient outcomes were beyond the scope of the present analyses. There were no exclusion criteria for the clinician participants.

All study procedures were approved by the Icahn School of Medicine at Mount Sinai and Florida International University institutional review boards. This paper represents the first stage of a larger study investigating the use of virtual human interaction in improving clinicians' emotional self-awareness and empathic communication toward suicidal patients.

## Measures

### Therapist Response Questionnaire-Suicide Form

Clinicians' emotional self-awareness about their emotional response toward each virtual patient with whom they interacted was measured using the Therapist Response Questionnaire-Suicide Form (TRQ-SF) [11], a 10-item Likert-type self-report scale capturing clinicians' emotional responses to suicidal patients. The TRQ-SF has a three-factor structure: affiliation (5 items), distress (3 items), and hopefulness (2 items). Each individual item score ranges from 0 (not at all) to 4 (extremely), yielding a 0–40 total score for the scale. Affiliation and hope were reverse coded as detachment and hopelessness, respectively [25]. In a previous study [11], clinicians' average TRQ-SF total score toward psychiatric outpatients was 9.12 (SD = 5.2, range 0–33). The TRQ-SF demonstrated good reliability both as a three-factor and a general one-factor scale, indicating general negative emotional responses toward suicidal patients, with a Cronbach's alpha value of 0.88 [11]. In this study, we only used the TRQ-SF total score because it better reflects the negative emotional response; more specifically, higher scores reflect more intense negative emotional responses.

### Empathic Communication Coding System

We measured clinicians' verbal empathic communication during the virtual human interactions using the Empathic Communication Coding System (ECCS), a method extensively used by our group in prior research [15]. The ECCS is an interaction analysis system composed of two steps: identifying patient-generated empathic opportunities and evaluating clinicians' responses to those opportunities [17, 26]. This instrument has been used to characterize empathy in real-patient, standardized-patient, and virtual-patient interactions [15, 17, 18, 22]. Empathic opportunities are operationalized along three categories: patients' verbalized emotions (i.e., the patient explicitly describes feeling an emotion, such as happiness, fear, hate, and sadness, at the time of assessment), patients' statements about progress (i.e., the patient reports current physical and/or psychosocial improvement in quality of life), and patients' expression of challenges (i.e., the patient describes negative consequences of physical and/or psychosocial problems that are impacting quality of life).

Clinicians' responses to these empathic opportunities can be described in seven hierarchical categories from complete denial of the patient's opportunity to confirmation or even sharing of similar experience with that of patients [16]: level 0, denial of patient's emotions and perspective, ignoring or disconfirming patient's empathic opportunity; level 1, perfunctory (automatic or minimal) recognition of the patient's emotion; level 2, implicit recognition (i.e., the clinician does not explicitly recognize the central issue in the empathic opportunity but focuses on a peripheral aspect of the statement, such as a biomedical issue) versus dealing directly with the progress, challenge, or emotion; level 3, the clinician explicitly acknowledges the statement but does not pursue the topic; level 4, explicit recognition of the empathic opportunity presented by the patient and pursuit of the topic by asking questions, giving advice, and/or elaborating on the problem; level 5, the clinician legitimizes and fully confirms the patient's statement; and level 6, the clinician explicitly either shares the patient's emotion or reports having similar experience.

Coding of empathetic responses was performed in two steps. First, empathic opportunities were embedded in the statements and answers that the virtual patients provided to clinician participants during the virtual human interaction. Then, the clinicians' responses to these empathic opportunities were collected from the interaction transcripts and analyzed by human expert coders. ECCS coders were trained extensively according to the procedures previously described by Foster et al. [15] and achieved high interrater reliability of 0.936, measured with the intraclass correlation coefficient [27]. To achieve instant coding of clinician participants' responses to the virtual patients' statements featuring empathic opportunities, we created a machine learning algorithm called "classifier" to replace human raters and to make ECCS fully scalable, as previously described by Yao et al. [16]. Instant coding of ECCS allowed us to provide clinicians in the intervention condition with feedback about their empathy level immediately after their interaction with the virtual patient.

## Study Procedures

All clinician participants were invited to interact with two different virtual patients. Clinicians interacted with the first virtual patient at time 1 (T1). One week later, clinicians interacted with the second participant, at time 2 (T2).

Immediately after each of these two interactions with a virtual patient, clinicians completed a self-report measure, the TRQ-SF, about their emotional response to the virtual patient with whom they had just interacted.

### Virtual Human Interactions

The two virtual human interactions used for this study were created in Virtual People Factory [28] and allowed clinicians to interact with the virtual patients. To conduct the clinical interviews, clinicians could ask questions to their virtual patients by talking or typing in a chat window. The virtual patients answered orally, and their answers appeared in writing in the same chat window.

Each virtual patient's scenario and prewritten answers contained predetermined empathic opportunities, as described below. The virtual human interactions were administered in counterbalanced order 7 days apart (i.e., scenario 1 followed by scenario 2 or scenario 2 followed by scenario 1) to minimize order effects.

In scenario 1, Cynthia Young is a 21-year-old college student who presents with symptoms of a major depressive episode. Her functioning declined, and she developed suicidal thoughts following a personal loss. The scenario contains predetermined empathic opportunities (e.g., "My cousin and I were like sisters. I cry every time I think of her"), related to the patient's suicidal ideation [15, 16].

In scenario 2, Bernie Cohen is a 53-year-old gay male who suffered a catastrophic personal loss and presents with suicidal ideation, describing a seemingly visceral feeling of pain and "waves of fear" at the thought of his loss [16, 29]. Statements like "I feel like I am just one walking burning wound" represent empathic opportunities in this scenario.

### Study Procedures in the Intervention (Emotional Self-awareness Feedback) Condition

After completing each virtual human interaction, at T1 and T2, and completing the TRQ-SF measure, participants in the intervention condition received emotional self-awareness feedback which included the three following components: an individual transcript of the interaction with the virtual patient; the scores obtained by the clinician on the TRQ-SF measure, to assess the clinician's self-reported negative emotional response toward the virtual patient [11]; and if the TRQ-SF score indicated high negative emotional response, participants were directed through online live feedback, in a supportive manner, to self-examine possible negative emotional responses at the time of the virtual interaction, such as "anger or coldness triggered by anxiety," and the clinician's score for verbal empathic communication coded by ECCS [15, 26], which reflects the level of verbal empathy, as well as suggestions for alternative empathic responses.

### The Control Condition (No Emotional Self-awareness Feedback)

The clinicians in the control condition also engaged in interactions with the virtual patients at T1 and T2 and completed the TRQ-SF about their emotional responses to the virtual patients. However, they did not have access to their TRQ-SF scores and were not provided any feedback about their verbal empathy level with the virtual patient. Furthermore, the controls did not see their ECCS scores and did not receive the suggestions for alternative empathic responses.

### Randomization and Blinding

Blocked randomization was adopted to promote balanced groups [30]; as clinicians were recruited to the study at each clinic, they were paired by order of recruitment (i.e., the first two clinicians to be recruited at a given clinic formed the first pair, the second and third clinicians to be recruited formed the second pair, and so on). Within each pair, one clinician was randomly assigned to the intervention condition, whereas the other clinician was assigned to the control condition. The randomizeR package for R was used to assign clinicians to specific study conditions [31]. Randomization took place before the first interaction.

The study followed a double-blind procedure to minimize the effects of expectation bias. Prior to the start of the study, the co-investigators assigned different letters (either A or B) to the intervention and control conditions; they did not reveal which letter corresponds to which condition to the coordinators and research assistants until the completion of the study. The co-investigators at each site carried out the randomization procedure and recorded this information in a password-protected file. Prior to the first interaction, the co-investigators informed the coordinators which condition (i.e., A or B) each participant was assigned to. Although the participants in the intervention condition were aware that they were receiving feedback, they did not know that it was the active component of the training. REDCap [32] software was adopted to administer, directly on the virtual interface, the study self-report measure, TRQ-SF.

## Sample Size

An a priori power analysis was conducted to determine the number of clinician participants required to detect a moderate effect size ($f = 0.35$), which was expected based on past research [11, 15]. With power = 0.80, alpha = 0.05, $f = 0.35$, and our planned analyses, we determined that we would need a total of 68 clinician participants to detect this effect size. Although we planned to recruit 80 clinician participants to account for attrition and potentially smaller effect sizes, the study ran into recruitment difficulties due to the COVID-19 pandemic and was thus slightly underpowered for our initial plans.

### Data Analytic Strategy

Descriptive statistics were first computed for the full sample and stratified by time point and study condition (i.e., intervention versus control condition), on each outcome variable (i.e., clinicians' negative emotional responses and clinicians' empathic responses). Linear mixed models [33] were then computed using the *lme4* [34] and *lmerTest* [35] packages to examine changes in clinicians' negative emotional responses and empathic responses over time (i.e., T1 versus T2), as moderated by clinician condition (i.e., intervention versus control condition) and clinician baseline empathy level (i.e., low versus medium/high). Specifically, the main effects and interactions of time, condition, and empathy at T1 were entered as fixed predictors, whereas clinicians' negative emotional responses and empathic responses were entered as primary outcome variables in separate models. Follow-up sensitivity analyses were conducted to compare which type(s) of empathetic responses were most strongly impacted by the intervention condition over time. Random intercepts for clinicians were entered in both models to account for the non-independence of observations. Pairwise comparisons were examined using the *emmeans* package [36]. Missing data were minimal and handled using listwise deletion.

## Results

### Participants Characteristics

The clinicians' sample included 64 subjects, of whom 31 were in the intervention condition and 33 in the control condition. Approximately two-fifths (39.06%) were cisgender women ($n=25$), 59.38% cisgender men ($n=38$), and 1.56% other/non-binary ($n=1$). The enrolled clinicians were 73.44% ($n=47$) medical doctors, 6.25% ($n=4$) had a Master of Arts or Master of Science in psychology, 12.5% ($n=8$) had a Ph.D. or Psy.D. in Psychology, and 6.25% ($n=4$) were social workers. The average age of the total sample was 33.41 years (range = 25–65; SD = 6.98). The average length of reported mental health work experience was 4.44 years (SD = 4.12) (see Table 1). Descriptive statistics stratified by time point and study condition on outcome variables are shown in Table 2.

### Participants' Empathic Responses

A total of 2058 responses to empathic opportunities were recorded, referring to both T1 and T2 and among all participants, both control and intervention conditions. Of those, 327 (15.89%) were classified as level 0, 60 (2.92%) as level 1, 951 (46.21%) as level 2, 392 (19.04%) as level 3, 148 (7.19%) as level 4, 177 (8.60%) as level 5, and 1 as level 6 (0.05%). A minority of responses ($n=2$ [0.09%]) were not coded and then categorized as missing.

The only variable showing statistically significant differences between the control and intervention conditions at T1 was the ECCS score ($t[52]=2.00$, $p=0.047$); the control condition scored slightly higher ($M=2.33$, SD = 0.69) than the intervention condition ($M=1.98$; SD = 0.50), with a low ECCS score at T1 (< 2 on a total of 6 levels) in 10 clinicians in the control condition, versus 7 clinicians in the intervention condition. No significant differences were found in sociodemographic characteristics, as shown in Table 1. The mean virtual human interaction duration in our study was 20.29 min (SD = 9.19). We identified two levels of ECCS average scores at T1, before interacting with the second virtual patient and before any feedback was given: low T1 empathy level, with a T1 ECCS score < 2, and high T1 empathy level, with a T1 ECCS score ≥ 2, an approach utilized by Johnson Shen et al. [18]. Then, an ECCS score < 2 can be translated as low empathy level on the verbal side. Previously published papers reported an ECCS mean score of 2.57 (SD = 0.75), in a range from 2.20 to 2.91, depending on the participants' condition [15]. In that context, the participants were medical students, and the virtual patient was Cynthia Young, the same adopted in scenario 1 of the current study. Our participants' mean ECCS score was 2.20 (SD = 0.62) for both conditions.

### Multilevel Regression Models

Regarding our first hypothesis (i.e., clinicians in the intervention condition would have greater decreases in negative emotional responses from T1 to T2 than those in the control condition), we found no statistically significant main effects or interaction effects for time, condition, or T1 empathy in predicting clinicians' negative emotional responses ($p$s = 0.442 to 0.998, as shown in Table 3).

Regarding our second hypothesis (i.e., clinicians in the intervention condition would exhibit greater increases in empathetic communication from T1 to T2 than those in the control condition), we found that the proposed three-way interaction between time, condition, and T1 empathy was not statistically significant ($p=0.058$) in predicting clinicians' empathic responses (see Table 4). However, there was a significant two-way interaction between time and condition ($p=0.006$), as well as a significant main effect for T1 empathy ($p=0.001$), in predicting clinicians' empathic responses. Simple slopes across time and condition, stratified by the T1 empathy group, are presented in Fig. 1. Specifically, empathic responses increased ($B=1.15$, SE = 0.25, $p<0.001$, 95% CI [0.42, 1.89]) from T1 to T2 among

**Table 1** Sample sociodemographic and clinical characteristics

| | Total sample | Intervention group | Control group | |
|---|---|---|---|---|
| | *N (%)* *N=64* | *N (%)* *N=31* | *N (%)* *N=33* | *p* |
| Gender | | | | 1 |
| Cisgender men | 38 (59.38) | 19 (61.29) | 19 (57.58) | |
| Cisgender women | 25 (39.06) | 12 (38.71) | 13 (39.39) | |
| Other/non-binary | 1 (1.56) | 0 (0) | 1 (3.03) | |
| Age [mean (SD)] | 33.41 (6.98) | 32.71 (6.67) | 34.06 (7.29) | .443 |
| Race | | | | .771 |
| White/Caucasian | 21 (32.81) | 10 (32.26) | 11 (33.33) | |
| Black/African American | 3 (4.69) | 1 (3.23) | 2 (6.06) | |
| Asian | 15 (23.44) | 9 (29.03) | 6 (18.18) | |
| Hispanic | 20 (31.25) | 8 (25) | 12 (36.36) | |
| Other race | 5 (7.81) | 3 (9.68) | 2 (6.06) | |
| Sexual orientation | | | | .604 |
| Heterosexual/straight | 47 (73.44) | 21 (67.74) | 26 (78.79) | |
| Gay/lesbian/homosexual | 10 (15.63) | 7 (22.58) | 3 (9.09) | |
| Bisexual | 3 (4.69) | 1 (3.23) | 2 (6.06) | |
| Other | 2 (3.12) | 1 (3.23) | 1 (3.03) | |
| Decline to state | 2 (3.12) | 1 (3.23) | 1 (3.03) | |
| Marital status | | | | .334 |
| Single/not in a committed relationship | 13 (20.31) | 4 (12.90) | 9 (27.27) | |
| Single, formally married/formally in a committed relationship | 8 (12.5) | 5 (16.13) | 3 (9.09) | |
| Married/in a committed relationship | 43 (67.19) | 22 (70.97) | 21 (63.64) | |
| Type of degree | | | | .203 |
| MD or DO | 47 (73.44) | 26 (83.87) | 21 (63.64) | |
| MA or MS in Psych | 4 (6.25) | 2 (6.45) | 2 (6.06) | |
| PhD or PsyD in Psych | 8 (12.5) | 3 (9.68) | 5 (15.15) | |
| MSW | 4 (6.25) | 0 (0) | 4 (12.12) | |
| MS in nursing | 1 (1.56) | 0 (0) | 1 (3.03) | |
| Years of training [mean (SD)] | 4.44 (4.12) | 4.26 (3.99) | 4.60 (4.30) | .745 |
| ECCS baseline | 2.19 (.62) | 1.98 (.50) | 2.33 (.69) | .047 |

Note: *SD*, standard deviation; *MD*, Medical Doctor; *MA*, Master of Arts; *MS*, Master of Science; *MSW*, Master of Social Work; *ECCS*, Empathic Communication Coding System

**Table 2** Outcome variables at each time point in both study conditions

| Group | Intervention group | Control group | Total |
|---|---|---|---|
| *N* | 31 | 33 | 64 |
| T1 NER (TRQ total) | 15.39 (5.15) | 16.36 (5.56) | 15.81 (5.31) |
| T2 NER (TRQ total) | 13.96 (6.23) | 15.46 (5.22) | 14.60 (5.69) |
| T1 ER (ECCS average) | 1.98 (.50) | 2.33 (.69) | 2.19 (.62) |
| T2 ER (ECCS average) | 2.37 (.62) | 2.40 (.64) | 2.37 (.61) |

Note: *NER*, negative emotional response; *ER*, empathic response; *ECCS*, Empathic Communication Coding System; *TRQ-SF*, Therapist Response Questionnaire-Suicide Form; *T1*, baseline assessment; *T2*, follow-up assessment

clinicians in the intervention condition with low T1 empathy. On the other hand, there were no changes in empathic responses over time among clinicians with high T1 empathy in the intervention condition ($B = 0.04$, SE = 0.17, $p = 1.00$, 95% CI [−0.45, 0.54]) or clinicians with low T1 empathy in the control condition ($B = 0.29$, SE = 0.21, $p = 0.839$, 95% CI [−0.33, 0.90]), or clinicians with high T1 empathy in the control condition ($B = -0.11$, $SE = 0.15$, $p = 1.00$, 95% CI [−0.54, 0.32]).

## Comparison of Challenge-Focused and Emotion-Focused Empathic Responses

Given the nature of the virtual human interaction scenarios, which represented patients at risk of suicide, there were

**Table 3** Negative emotional responses to a virtual patient in relation to time, study condition, and clinicians' baseline empathy levels using multilevel models

| Variable | *B* | SE | *p* | 95% CI | Cohen's $f^2$ |
|---|---|---|---|---|---|
| Fixed effects | | | | | |
| Intercept | 21.27 | 5.24 | < .001 | [10.89, 31.64] | |
| Baseline empathy: moderate/high | 1.77 | 4.23 | .677 | [− 6.68, 10.13] | .002 |
| Study condition | − .01 | 5.30 | .998 | [− 10.59, 10.4] | .0000001 |
| Time | − .19 | 2.13 | .929 | [− 4.46, 4.03] | .0001 |
| Baseline × condition | − 1.20 | 6.43 | .852 | [− 13.94, 11.59] | .0005 |
| Baseline × time | − 1.40 | 2.59 | .589 | [− 6.53, 3.78] | .005 |
| Condition × time | − 2.38 | 3.26 | .468 | [− 8.85, 4.14] | .009 |
| Baseline × condition × time | 3.04 | 3.93 | .442 | [− 4.82, 10.84] | .01 |
| Random effects | | | | | |
| $\tau_{00, \text{subject}}$ | 3.56 | | | | |
| $\sigma^2$ | 21.28 | | | | |
| $N_{\text{subject}}$ | 53 | | | | |
| Observations | 105 | | | | |
| Deviance | 634.2 | | | | |

Note: Covariates were included in models but are not presented in the table to facilitate interpretation of relevant study findings. Baseline clinician empathy: 0 = low, 1 = medium/high. Study condition: 0 = control group; 1 = intervention group. Time: 0 = baseline, 1 = follow-up

**Table 4** Empathic responses to a virtual patient in relation to time, study condition, and clinicians' baseline empathy levels using multilevel models

| Variable | *B* | SE | *p* | 95% CI | Cohen's $f^2$ |
|---|---|---|---|---|---|
| Fixed effects | | | | | |
| Intercept | 2.45 | .48 | < .001 | [1.50; 3.40] | |
| Baseline empathy: moderate/high | 1.29 | .38 | .001 | [.52; 2.05] | .10 |
| Study condition | − .91 | .49 | .064 | [− 1.88; .05] | .03 |
| Time | .29 | .19 | .147 | [− .10; .68] | .02 |
| Baseline × condition | .39 | .59 | .512 | [− .78; 1.57] | .004 |
| Baseline × time | − .40 | .24 | .100 | [− .87; .077] | .03 |
| Condition × time | .87 | .31 | .006 | [.26; 1.48] | .07 |
| Baseline × condition × time | − .71 | .37 | .058 | [− 1.44; .02] | .03 |
| Random effects | | | | | |
| $\tau_{00, \text{subject}}$ | 0 | | | | |
| $\sigma^2$ | .19 | | | | |
| $N_{\text{subject}}$ | 53 | | | | |
| Observations | 107 | | | | |
| Deviance | 127.9 | | | | |

Note: Covariates were included in models but are not presented in the table to facilitate interpretation of relevant study findings. Baseline clinician empathy: 0 = low, 1 = medium/high. Study condition: 0 = control group; 1 = intervention group. Time: 0 = baseline, 1 = follow-up

very few progress-type empathic opportunities. In total, 2058 empathic opportunities were recorded across T1 and T2, specifically, 166 (8.06%) progress-focused opportunities, 1302 (63.27%) challenge-focused opportunities, 589 (28.62%) emotion-focused opportunities, and 1 (0.05%) missing.

Follow-up sensitivity analyses examined whether there were differences in empathic responses to challenge-focused and emotion-focused opportunities. Given the small number of the progress opportunities (which was unsurprising due to the nature of the virtual human interaction scenario, in which patients were seeing a clinician for a single session and were at risk for suicide), sensitivity analysis was not performed on progress opportunities. First, among challenge-focused opportunities, the three-way interaction between time, condition, and T1 empathy was significant ($p = 0.042$). Examination of the form of the interaction through the modeling of simple slopes mirrored patterns found in the primary analyses. There

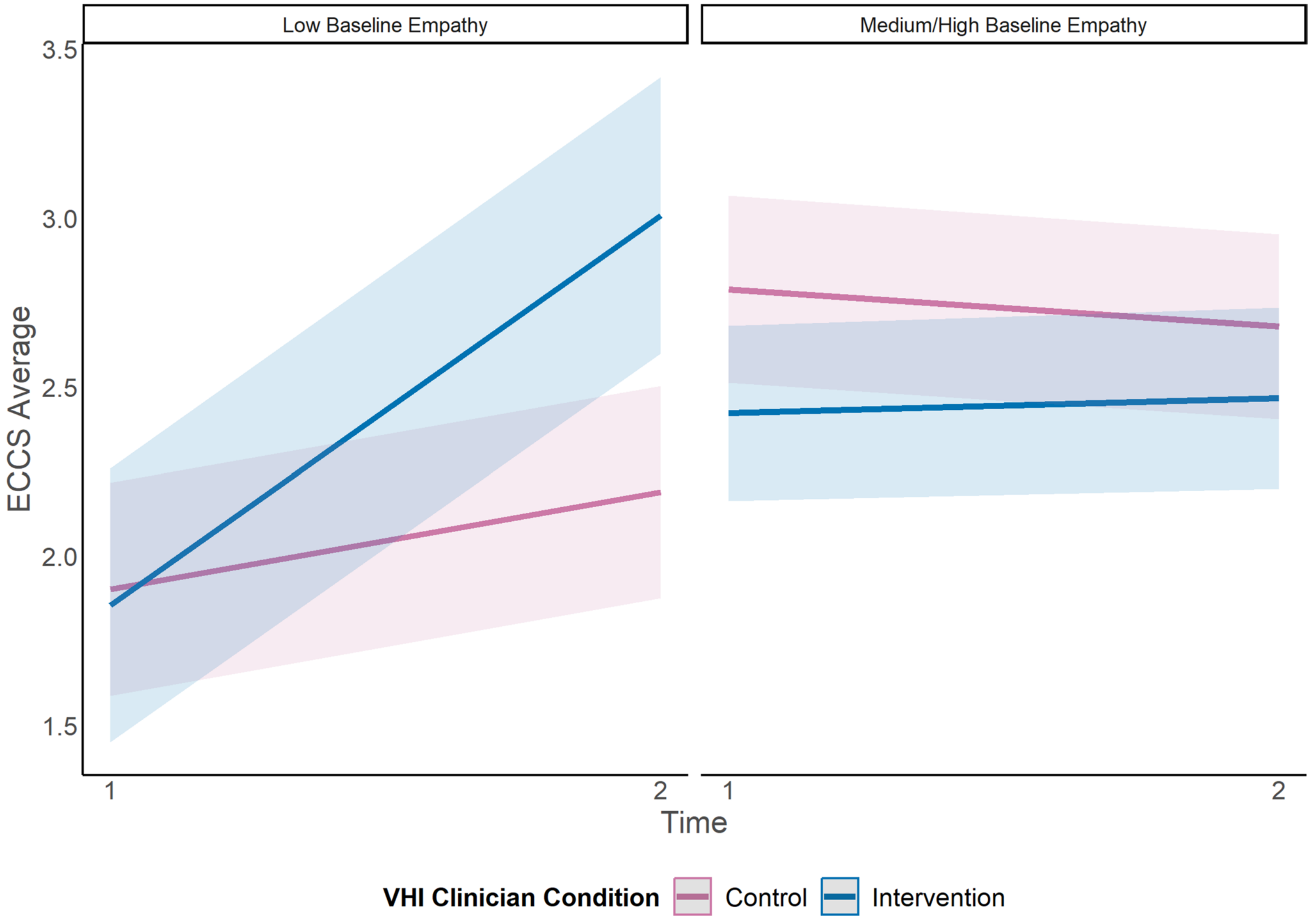


**Fig. 1** Clinician empathetic responses by time, condition, and baseline empathy levels. Note: Time 1 refers to the baseline assessment. Time 2 refers to the follow-up assessment occurring 1 week after the virtual human training. ECCS, Empathic Communication Coding System; VHI, virtual human interaction

was a significant increase in empathic responses to challenge-focused opportunities from T1 to T2 among participants who were in the low T1 empathy group and the intervention condition. There were no changes, however, in empathic responses to challenge-focused opportunities among the other three groups. Second, among emotion-focused opportunities, all interaction effects between time, condition, and T1 empathy were non-significant ($p = 0.116$ to 0.628), which suggests that changes in empathic responses are primarily driven by improvements in responding to challenge-focused, rather than emotion-focused, opportunities.

## Discussion

The current study sought to assess the impact of a virtual human-based clinical training on clinicians' emotional self-awareness and empathic communication toward virtual humans representing patients at risk of suicide. We postulated that clinicians who received emotional self-awareness feedback after the first virtual patient interaction (i.e., the intervention condition) would show lower levels of negative emotional response (hypothesis 1) and enhanced empathic communication (hypothesis 2) with their second virtual patient, respectively resulting in lower TRQ-SF scores and higher ECCS scores compared to the control condition.

Our study findings did not support the first hypothesis. Namely, clinicians' negative emotional response, measured with TRQ-SF, did not change significantly either in the control or the intervention condition. However, the findings supported our second hypothesis; clinicians who scored low in baseline empathy (T1) improved significantly in empathic communication with their second virtual patient (T2) in the intervention condition.

A possible explanation for the absence of significant findings for our first hypothesis is that virtual human interaction is a relatively new technology that has great technical complexities and requires computational expertise to be developed [37]. The great sophistication of this tool, combined with its novelty, makes it vulnerable to instability and occasional malfunctions. Virtual human interactions proved to be effective in various clinical contexts [22–24]; however, further developments are needed to make this tool more reliable and applicable to an increasingly broad range of clinical settings. Such glitches may have interfered with clinicians' ability to experience emotional responses that are comparable to those they typically experience with patients at high risk for suicide.

Another explanation to the absence of improvement in clinicians' negative emotional response lies in the subtle difference between awareness of the emotion and management of that emotion. The TRQ-SF score does not measure emotional managing capabilities, so our result does not mean that training with virtual human interaction does not improve clinicians' skills in managing negative emotional response, which remains to be established in future research. Alternatively, it is possible that methodological challenges accounted for this null finding. Specifically, our study was not powered to detect relatively smaller effects; thus, it is possible that smaller changes in negative emotional response may have occurred that were not detectable via null hypothesis significance testing. These possibilities should each be explored through replication and extension of this work.

Although the study failed to produce significant improvement within clinicians' emotional response to virtual patients, it did demonstrate that clinicians who were not skilled in empathetic communication at T1 showed improved empathic communication with their second virtual patient in the intervention condition, as postulated in hypothesis 2. Namely, clinician participants in the intervention condition who demonstrated low ECCS scores at T1 significantly improved in their ECCS scores by T2. This latter finding replicated results from previous work showing that clinicians are able to display empathy during interactions with virtual patients [22], specifically, that students trained with virtual patients in ways that include immediate, post-interaction feedback scored higher in empathic communication, measured with the ECCS, than those who only interacted with virtual patients and did not receive feedback [15]. These findings suggest that clinicians with less developed baseline empathic communication skills would benefit most from such a training tool.

Moreover, as shown by our post hoc sensitivity analysis, improvement in verbal empathic communication in the intervention condition was driven by challenge opportunities, which differs from previous findings, where challenge and emotion opportunities were equivalent [15]. In challenge opportunities, the patient describes the negative consequence of a physical and/or psychosocial problem on the patient's quality of life, whereas emotion opportunities are arguably easier to recognize even by less-experienced clinicians. The nature of these statements and their intense emotional features might explain why challenge opportunities are more likely to elicit empathic responses from the clinicians and, consequently, lead to a significantly higher ECCS average score. This response replicates the real clinician-patient interactions, coded via ECCS [18, 26, 38], underlining the likely correspondence with real-world clinical practice.

The findings of this study need to be considered in view of its limitations. First, the number of subjects was relatively small, limiting our findings to moderate or larger effects only. Second, the clinicians were diverse in their background and level of training. This limitation can be mitigated in the future by inviting only clinicians in one profession and at relatively similar level of training to participate in this type of study. Third, we only used two virtual scenarios, which increased the internal validity yet limited the generalizability of our findings. Fourth, the virtual patient system used a 3D virtual character that has limitations in mimicking a real patient. These limitations include visual realism differences, occasionally being unable to properly respond to a user's conversation, and having to communicate using typing or speech-to-text interfaces. Fifth, our sample size was too small to conduct secondary analyses to examine if there were different results among clinicians with varying trainings and years of experience, as well as varying sociodemographic characteristics, which will be an important avenue of future research. Sixth, we have only demonstrated the effectiveness of virtual human interaction training in improving verbal empathic communication, which is only one aspect of the complex relationship between a patient and clinician. Furthermore, we only evaluated the short-term impact of this training; assessment of longer-term effectiveness and retention of acquired skills is needed. The TRQ-SF was completed after the virtual human interaction and thus might be affected by recall bias, whereas the ECCS score represents the average of multiple measurements during the encounter. One possible future direction would be to implement multiple TRQ measurements to overcome this limitation. Furthermore, clinicians self-selected for this study, and thus, our results may not be generalizable to clinicians who did not want to participate. Last, virtual human interaction may represent a reliable and feasible training tool to be adopted in educational programs. The innovative nature of this technology, however, makes it a sophisticated tool currently available mostly in academic or research sites that have access to funding, expertise, and specialized professional figures. A further development of such technology will make its use and adoption easier and more reliable and then suitable for all training centers, even for those with less resources.

In conclusion, our study reports the first successful use of virtual human interaction for training clinicians in emotional self-awareness and empathic communication when working with virtual patients representing suicidal patients. We demonstrated that feedback-enhanced virtual human interaction training increases verbal empathic communication in clinicians with low baseline empathic communication skills. This training is a potentially scalable method that should be used in graduate and postgraduate mental health training programs for training clinicians interacting with patients at risk of suicide. It might have an important impact in the emergency setting. Training clinicians with virtual human interaction may represent an important opportunity to simulate severe and emergency situations, such as high-risk suicide scenarios, in a fully controlled setting, without any risk. It might help clinicians to handle these clinical challenges, optimize hospital admissions, and even save lives. To do so, new high-risk suicidal virtual patients must be developed. Further research is needed to refine this training methodology so that it can be applicable to a broader array of clinician specialists and patient populations and become actually scalable. Once enhanced, virtual human interaction will represent a great opportunity to tackle suicide, improving detection and managing risk.

**Funding** This work was supported by the R34 MH119294-01 grant from the National Institute of Mental Health (NIMH).

**Data Availability** The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Declarations

**Disclosures** On behalf of all authors, the corresponding author states that there is no conflict of interest.

## References

1. Platt S, Niederkrotenthaler T. Suicide prevention programs. Crisis. 2020;41(Suppl 1):S99–124.
2. Curtin SC, Garnett MF, Ahmad FB. Provisional numbers and rates of suicide by month and demographic characteristics: United States, 2021. Centers for Disease Control and Prevention. Report No. 24. September 2022. https://doi.org/10.15620/cdc:120830.
3. Hedegaard H, Curtin SC, Warner M. Increase in suicide mortality in the United States, 1999–2018. NCHS Data Brief. 2020;362:1–8.
4. Stene-Larsen K, Reneflot A. Contact with primary and mental health care prior to suicide: a systematic review of the literature from 2000 to 2017. Scand J Public Health. 2019;47(1):9–17.
5. Schechter M, Goldblatt M, Maltsberger JT. The therapeutic alliance and suicide. Br J Psychother. 2013;29:315–28.
6. Berman AL. Risk factors proximate to suicide and suicide risk assessment in the context of denied suicide ideation. Suicide Life Threat Behav. 2018;48(3):340–52.
7. Kleiman EM, Turner BJ, Fedor S, Beale EE, Huffman JC, Nock MK. Examination of real-time fluctuations in suicidal ideation and its risk factors: results from two ecological momentary assessment studies. J Abnorm Psychol. 2017;126(6):726–38.
8. Stanley B, Brown GK, Currier GW, Lyons C, Chesin M, Knox KL. Brief intervention and follow-up for suicidal patients with repeat emergency department visits enhances treatment engagement. Am J Public Health. 2015;105(8):1570–2.
9. Jobes DA, Gregorian MJ, Colborn VA. A stepped care approach to clinical suicide prevention. Psychol Serv. 2018;15(3):243–50.
10. Michaud L, Greenway KT, Corbeil S, Bourquin C, Richard-Devantoy S. Countertransference towards suicidal patients: a systematic review. Curr Psychol. 2023;42:416–30.
11. Barzilay S, Yaseen ZS, Hawes M, Gorman B, Altman R, Foster A, et al. Emotional responses to suicidal patients: factor structure, construct, and predictive validity of the Therapist Response Questionnaire-Suicide Form. Front Psychiatry. 2018;9:104.
12. Hendin H, Lipschitz A, Maltsberger JT, Haas AP, Wynecoop S. Therapists' reactions to patients' suicides. Am J Psychiatry. 2000;157(12):2022–7.
13 Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. Psychother: Theory Res Pract. 1979;16(3):252–60.
14. Bloch-Elkouby S, Barzilay S. Alliance-focused safety planning and suicide risk management. Psychotherapy. 2022;59(2):157–62.
15. Foster A, Chaudhary N, Kim T, Waller JL, Wong J, Borish M, et al. Using virtual patients to teach empathy: a randomized controlled study to enhance medical students' empathic communication. Simul Healthc. 2016;11(3):181–9.
16. Yao H, Gomes de Siqueira A, Foster A, Galynker I, Lok B. Toward automated evaluation of empathetic responses in virtual human interaction systems for mental health scenarios. IVA '20: Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents. 2020;57:1–8. https://doi.org/10.1145/3383652.3423916.
17. Bylund CL, Makoul G. Examining empathy in medical encounters: an observational study using the empathic communication coding system. Health Commun. 2005;18(2):123–40.
18. Johnson Shen M, Ostroff JS, Hamann HA, Haque N, Banerjee SC, McFarland DC, et al. Structured analysis of empathic opportunities and physician responses during lung cancer patient-physician consultations. J Health Commun. 2019;24(9):711–8.
19. Christopher Muran J, Eubanks CF. Therapist performance under pressure: negotiating emotion, difference, and rupture. Washington, DC: American Psychological Association; 2020.
20. Rousmaniere T. Deliberate practice for psychotherapists: a guide to improving clinical effectiveness. New York: Routledge; 2016.
21. Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. Med Educ. 2009;43(4):303–11.
22. Kleinsmith A, Rivera-Gutierrez D, Finney G, Cendan J, Lok B. Understanding empathy training with virtual patients. Comput Human Behav. 2015;52:151–8.
23. Shah H, Rossen B, Lok B, Londino D, Lind SD, Foster A. Interactive virtual-patient scenarios: an evolving tool in psychiatric education. Acad Psychiatry. 2012;36(2):146–50.
24. Foster A, Chaudhary N, Murphy J, Lok B, Waller J, Buckley PF. The use of simulation to teach suicide risk assessment to health profession trainees—rationale, methodology, and a proof of concept demonstration with a virtual patient. Acad Psychiatry. 2015;39(6):620–9.
25. Rogers ML, Vespa A, Bloch-Elkouby S, Galynker I. Validity of the modular assessment of risk for imminent suicide in predicting short-term suicidality. Acta Psychiatr Scand. 2021;144(6):563–77.
26. Bylund CL, Makoul G. Empathic communication and gender in the physician-patient encounter. Patient Educ Couns. 2002;48(3):207–16.
27. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996;1(1):30–46.
28. Rossen B, Lind S, Lok B. Human-centered distributed conversational modeling: efficient modeling of robust virtual human

conversations. In: Ruttkay Z, Kipp M, Nijholt A, Vilhjálmsson HH, editors. Intelligent virtual agents. Berlin: Springer; 2009. p. 474–81.
29. Galynker I. The suicidal crisis: clinical guide to the assessment of imminent suicide risk. Oxford University Press; 2017.
30. Crespi CM. Improved designs for cluster randomized trials. Annu Rev Public Health. 2016;37:1–16.
31. Uschner D, Schindler D, Hilgers RD, Heussen N. randomizeR: an *R* package for the assessment and implementation of randomization in clinical trials. J Stat Softw. 2018;85(8):1–22.
32. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform. 2019;95: 103208.
33. Bryk AS, Raudenbush SW. Hierarchical linear models: applications and data analysis methods. Newbury Park: SAGE Publications, Incorporated; 1992.
34. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1):1–48.
35. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. J Stat Softw. 2017;82(13):1–26.
36. Lenth RV. Estimated marginal means, aka least-squares means [R package emmeans version 1.8.3]. 2022. https://CRAN.R-project.org/package=emmeans. Accessed 6 Dec 2022.
37. Lindner P, Miloff A, Hamilton W, Reuterskiöld L, Andersson G, Powers MB, et al. Creating state of the art, next-generation virtual reality exposure therapies for anxiety disorders using consumer hardware platforms: design considerations and future directions. Cogn Behav Ther. 2017;46(5):404–20.
38 Goodchild CE, Skinner TC, Parkin T. The value of empathy in dietetic consultations. A pilot study to investigate its effect on satisfaction, autonomy and agreement. J Hum Nutr Diet. 2005;18(3):181–5.