Psicothema (2025) 37(3) 23-32

Psicothema

https://www.psicothema.com/es • ISSN 0214-9915 • eISSN 1886-144X

Colegio Oficial de Psicología del Principado de Asturias

Article

ChatGPT Simulated Patient: Use in Clinical Training in Psychology

Ana Sanz[1], José Luis Tapia[2,3], Eva García-Carpintero[1], J. Francisco Rocabado[2] and Lorena M. Pedrajas[4]

[1]UNIE Universidad (Spain)
[2]Universidad Nebrija (Spain)
[3]Universitat de València (Spain)
[4]Universidada Alfonso X El Sabio (Spain)

ARTICLE INFO

Received: 15/11/2024
Accepted: 17/02/2025

Keywords:
Artificial intelligence
Standardized patients
Psychology education
Clinical simulation
Student cConfidence

ABSTRACT

Background: Incorporating artificial intelligence (AI) as standardized patients (SPs) in psychology education may enhance experiential learning and student confidence. The aim of the study was to analyze the effectiveness of using AI-based simulations to develop communication skills and influence psychology students' affective state. **Method**: A mixed-methods intervention study was conducted with 31 third-year psychology students. Participants engaged in clinical simulations using ChatGPT as an SP. Quantitative data on affective state, communication attitudes, and perceptions of knowledge and skills were collected pre- and post-intervention via questionnaires. Qualitative data were obtained through open-ended questions and a focus group. Data were analyzed using repeated measures ANOVA and thematic analysis. **Results**: Significant reductions in negative affect and increases in perceived knowledge and skills were observed post-intervention. No significant changes were found in communication attitudes. Qualitative findings supported the quantitative results, indicating improved confidence and reduced anxiety during simulated patient interactions. **Conclusions**: Utilizing AI as SPs is an effective pedagogical tool that enhances experiential learning, increases student confidence in professional skills, and positively influences the affective state. This innovative approach offers a valuable supplement to traditional teaching methods in psychology education.

Paciente Simulado con ChatGPT: Utilización en la Formación Clínica en Psicología

Palabras clave:
Inteligencia artificial
Pacientes estandarizados
Educación en psicología
Simulación clínica
Confianza del estudiante

RESUMEN

Antecedentes: La integración de la inteligencia artificial (IA) como pacientes estandarizados (PE) en la educación en psicología puede mejorar el aprendizaje experiencial y la confianza de los estudiantes. Este estudio analizó la efectividad de simulaciones basadas en IA para desarrollar habilidades de comunicación e influir en el estado afectivo de estudiantes de psicología. **Método**: Estudio de intervención de métodos mixtos con 31 estudiantes de tercer año, utilizando ChatGPT como PE. Se recopilaron datos cuantitativos sobre estado afectivo, actitudes hacia la comunicación y percepciones de conocimiento y habilidades antes y después de la simulación. También se obtuvieron datos cualitativos mediante preguntas abiertas y un grupo focal. Los datos se analizaron mediante ANOVA de medidas repetidas y análisis temático. **Resultados**: Los resultados mostraron una disminución significativa en el afecto negativo y un aumento en la percepción de conocimiento y habilidades tras la simulación. Los hallazgos cualitativos respaldaron estos resultados, indicando mayor confianza y menor ansiedad en las interacciones con el paciente simulado. **Conclusiones**: El uso de IA como PE es una herramienta pedagógica eficaz que complementa los métodos tradicionales, mejora el aprendizaje experiencial, refuerza la confianza en habilidades profesionales y tiene un impacto positivo en el estado afectivo de los estudiantes.

Cite as: Sanz, A., Tapia, J. L., García-Carpintero, E., Rocabado, J. F., & Pedrajas, L. M. (2025). ChatGPT simulated patient: Use in clinical training in Psychology. *Psicothema, 37*(3), 23-32. https://doi.org/10.70478/psicothema.2025.37.21

Corresponding author: Ana Sanz, ana.sanz@universidadunie.com

Introduction

Simulation-based education in the health sciences has become a well-established pedagogical approach, offering a dynamic, practice-focused learning experience (Sezgin & Bektas, 2023). This method provides a secure and supervised setting wherein students can develop, practice and hone essential clinical abilities without compromising patient safety or experiencing the stressors associated with direct interaction with actual patients. Additionally, it enables students to cultivate assurance in their capacity to perform effectively in intricate clinical scenarios, equipping them with the skills to navigate the complexities of genuine healthcare settings (Ton et al., 2024).

One of the most promising strategies within this approach is simulation with standardized patients (SPs) (Flanagan & Cummings, 2023). SPs, individuals trained to accurately represent various clinical conditions, have been demonstrated to be effective tools in both formative and summative teaching (Hillier et al., 2023). These SPs are capable of realistically simulating a variety of symptoms, behaviors, and emotions, providing direct and invaluable feedback on the students' performance during simulation sessions (Gerzina & Stovsky, 2023). The use of SPs has been demonstrated to be an efficacious method for the enhancement of clinical skills, decision-making and communication in students of health sciences (Johnson et al., 2020). Recent research indicates that the use of SP-based simulation has the potential to enhance the educational experience and facilitate active learning (Burrell et al., 2023; Dawood et al., 2024; Monahan et al., 2024).

Anxiety and confidence in patient care situations are factors that impact students' clinical performance. Students entering practice frequently encounter deficiencies in their knowledge, clinical skills and competencies in patient communication, which can give rise to feelings of insecurity and anxiety in the clinical setting. Simulation in a controlled environment can facilitate the development of greater confidence and stability, which in turn enhances performance in patient care. Research has identified several factors that influence students' perceptions of confidence and safety after simulation experiences (Basnet et al., 2024; Hawkins & Tredgett, 2016; Yu et al., 2021). In fact, Clinard (2022), all students indicated that these simulations significantly enhanced their confidence in treating patients, particularly in complex scenarios. This practical experience enables them to reinforce their ability and confidence in their own clinical skills. Similarly, Moss (2023) found that students exhibited a notable enhancement in their confidence ratings following the simulation, along with expressing high levels of satisfaction with the educational experience. Such exercises assist students in managing their fears and anxieties prior to encountering authentic care environments. In a focus group session, teachers also discussed how they perceived an increase in students' practical skills and overall satisfaction, which lends further support to the idea that simulation is an effective method to prepare students for clinical care (Carrero-Planells et al., 2021).

Notwithstanding its advantages, clinical simulation practices remain less prevalent within the Bachelor's Degree in Psychology compared to other health sciences fields, such as Nursing (García-Carpintero et al., 2024). A review of the literature revealed that studies on simulation in psychology are relatively scarce in Spain, with only a few investigations by Ruiz-Rodríguez et al. (2016) and Rodríguez et al. (2021). Moreover, recently published findings by López & López-Chicheri (2024) highlight that incorporating this pedagogical approach enhances psychology students' self-efficacy in their competencies and increases their satisfaction with experiential learning. The authors further emphasize the need to extend the duration of simulations with SP within the psychology program. This would provide students with more comprehensive clinical experience, ensuring they are well-prepared before entering formal placements and engaging in direct patient contact. It is interesting to mention the recent publication by Baile (2024), which aims to validate a patient profile in psychology generated with artificial intelligence.

Advancements in artificial intelligence (AI), particularly the emergence of Large Language Models (LLMs), have generated new opportunities across various industries, including clinical simulation. LLMs, such as ChatGPT developed by OpenAI©, are sophisticated AI systems trained on vast amounts of text, allowing them to understand and generate natural language in a way that mimics human communication. By leveraging deep learning techniques, these models engage in conversational and adaptive interactions, responding fluidly to a wide range of inputs. In clinical simulation, LLMs can act as virtual patients, offering realistic, personalized interactions that enrich the educational experience and support the development of essential clinical competencies in psychology students (Isaza-Restrepo et al., 2018; Scherr et al., 2023). Human and AI-based simulated patients each offer dynamic, context-driven interactions but differ in how they adapt, display emotions, and provide feedback. Whereas human patients exhibit genuine emotional responses shaped by cultural and social contexts, AI simulations rely on programmed algorithms. Early chatbots like ELIZA (Weizenbaum, 1966) and AIML-based systems established foundational conversational structures but were limited by rule-based designs and minimal contextual awareness (Gutiérrez-Maldonado et al., 2008; Peñaloza-Salazar et al., 2011; Rizzo et al., 2011; see Gutierrez-Maldonado et al., 2017, for a recent demonstration of an AIML-based system in VR). The emergence of large language models (LLMs) significantly expanded chatbot capabilities, allowing for more natural, flexible dialogue and enhanced contextual depth. This innovation makes simulated patients particularly valuable for training: they offer immediate feedback, reduce costs, and enable large-scale practice environments (Liu et al., 2023). Building on studies such as Scherr et al. (2023), which used ChatGPT for general clinical training, this approach tailors LLM-powered simulated patients to the field of psychology. While Scherr's study focuses on general medical scenarios, this application specifically trains students in psychology, enabling them to practice diagnosis and intervention for psychological disorders through interactive simulations. This approach harnesses AI's text-processing adaptability to provide students with lifelike chat encounters, allowing for more personalized learning and expanding opportunities for objective assessment, ultimately enhancing students' cognitive growth and self-efficacy in simulated scenarios (Morcela, 2022).

The main objective of the study was to analyse the feasibility and effectiveness of integrating artificial intelligence as a common pedagogical tool in the teaching of psychology. This general objective is in turn divided into two specific objectives:

1. To determine the impact of clinical simulation, through the use of AI-driven as SPs, on the improvement of self-perception of knowledge, students' affective state and communication skills.

2. To explore the students' perspective on the appropriateness of AI-driven as SPs and the learning opportunity derived from its use.

Method

A mixed-methods intervention study design was employed, incorporating a qualitative component (Fetters et al., 2013). Following an integrated concurrent design (Curry & Nunez-Smith, 2015), quantitative methodology was used in the first phase of data analysis, with qualitative methodology employed in the second phase. Qualitative data were collected post-intervention to elucidate potential mechanisms and explain the quantitative results. Quantitative and qualitative data were methodologically integrated by embedding one within the other, and jointly interpreted and reported through narrative and combined presentation approaches (Johnson, 2019). The study utilized ChatGPT-3.5 (OpenAI, 2023) as a large language model to simulate patient interactions during the intervention.

The study was conducted in accordance with the principles of the Declaration of Helsinki of the World Medical Association (WMA), and the protocol was approved by the Research Ethics Committee of University Nebrija (approval number UNNE-2024-0020). All participants were thoroughly informed, given the opportunity to ask questions, and provided their consent through signed forms for the focus group recording and for their inclusion in the quantitative and qualitative studies.

Participants

To determine the required sample size, a statistical power analysis was conducted using G*Power software (v3.1.9.7; Faul et al., 2007), employing the 'ANOVA: Repeated measures, within factors' statistical test appropriate for repeated measures. This analysis was performed under the assumption of a medium effect size ($f = 0.3$), in the absence of specific prior data. A significance level of 0.05 and a planned power of 0.85 were set, establishing that a total of 27 participants would be necessary to reliably assess the pre and post-intervention changes.

The study sample consisted of 31 third-year psychology students (74% female) from a private university, recruited through convenience sampling. One participant identified as non-binary. Mean age was 21.03 years ($SD = 1.43$).

For the qualitative phase, a sampling method based on the information power criteria was used. This approach suggests that the more relevant the information provided by the sample is to the study, the fewer participants are needed (Moser & Korstjens, 2018). Therefore, the same participants recruited for the intervention in the quantitative phase who agreed to participate were included in the focus group ($n = 12$). None of them withdrew from the study.

Instruments

The study employed a mixed-methods design, combining quantitative and qualitative approaches to thoroughly evaluate the intervention. Four quantitative instruments were used (see https://osf.io/se7dq/?view_only=891d4fb6d1304f6597496bfe69b29319), two of which were specifically created for this study. These instruments provided precise data on participants' knowledge, professional competencies, social impact, attitudes toward communication, and emotional well-being.

PANAS: Positive and Negative Affect Schedule (Watson et al., 1988) (adaptation to Spanish, López-Gómez et al., 2015). A 20-item questionnaire that measures individuals' positive (PA) and negative (NA) affects. The items are divided into two subscales: one for positive affects (such as joy and enthusiasm) and another for negative affects (such as sadness and irritability). Each item is rated on a scale from 1 to 5, where 1 indicates that the *affect has not been experienced* at all and 5 indicates a *very intense experience*. The PANAS is widely used in both academic research and clinical applications to assess emotional well-being. The direct score ranges from 20 to 100. In a general sample from Spain (López-Gómez et al., 2015) the Pearson's bivariate correlation between the PA and NA scales was -0.19 ($p < 0.001$) and Cronbach´s alpha was 0.92 for Positive Affect Scale and 0.88 for Negative Affect Scale. The item-total correlations of the PA factor ranged from 0.67 to 0.74, while those of the NA factor ranged between 0.52 and 0.69.

HCAS: Healthcare Communication Attitudes Scale (Escribano et al., 2021). An 11-item scale designed to assess healthcare professionals' attitudes towards communication in clinical settings. Each item is rated on a scale from 1 to 5, with higher scores indicating a more positive attitude towards effective communication. The HCAS helps identify professionals' perceptions and predispositions regarding the importance of communication in patient care, facilitating the implementation of training and development programs that enhance these critical skills in clinical practice. The direct score ranges from 11 to 55. In a sample of 255 nursing students with an average age was 22.66 years ($SD = 4.75$) and 82% were female, the internal consistency of the scale was adequate (0.75), and the data fit well with the model (CFI = 0.99; TLI = 0.99; RMSEA = .01 95% *CI* [.00–.05]). The overall instrument score poorly correlated with the self-efficacy in communication skills variable.

PIES: Perception and Impact Evaluation Scale. An ad hoc tool consisting of three items designed to measure students' self-perception of their knowledge, professional competencies, and the social impact of their field of study. Each item is rated on a scale from 1 to 5, where 1 indicates a *low level of perception* or impact and 5 indicates *a high level*. This instrument aims to evaluate the development of key competencies and social awareness among students, providing valuable data to improve educational programs and pedagogical interventions. The direct score ranges from 3 to 15.

SPI-MET: Simulated Patient Interaction Measurement & Evaluation Tool. An 11-item questionnaire designed ad hoc to evaluate healthcare professionals' performance in interactions with simulated patients. The items focus on aspects such as linguistic adequacy and emotional expression, with each item rated on a scale from 0 to 5, where 0 indicates *inadequate performance* and 5 indicates *excellent performance*. In addition, the instrument includes an extra item to assess the overall adequacy of the tool as a simulated patient, rated on a scale from 0 to 10, where 0 indicates poor performance and 10 indicates excellent performance.

The collection of qualitative data was carried out using two main methods: participants' responses to an open-ended question included in the SPI-MET (*"What would you add or improve about the activity?"*) and a focus group consisting of 12 participants conducted after the intervention, led by an observer and a moderator. This focus group encouraged participant interaction, fostering the emergence of diverse opinions and perceptions. Information was gathered using a question guide developed from a prior literature review, focusing on specific topics of interest (Table 1). This facilitated an in-depth exploration of students' perceptions and experiences related to the categories proposed by the aforementioned instruments. The focus group was audio recorded with prior consent from the participants, lasting 57 minutes. Additionally, researchers' field notes were used as a secondary source of information to provide more detailed insights and support the data obtained in the focus group. The qualitative methodology offered a rich and contextualized perspective on their interaction with artificial intelligence in a clinical simulation setting.

Table 1
Categories and Focus Group Questions

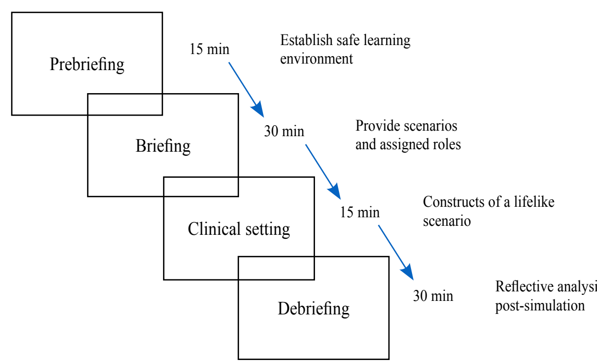
Categories	Focus group questions
Keen	How did you feel during the activity interacting with ChatGPT? How do you think this activity has influenced your safety? How do you think simulation has influenced your anxiety?
Patient Communication	How would you evaluate the way ChatGPT communicated as a patient? Did you find ChatGPT's behavior as a patient realistic and appropriate for the activity? Do you think ChatGPT's answers to your questions were appropriate and consistent with the situation?
Utility of the tool	How would you rate ChatGPT's performance as a simulated patient? How would you describe the degree of difficulty you experienced using ChatGPT during the activity? Were there any technical or interface aspects that made the tool difficult to use? Do you think this activity is useful to improve your skills in the subject?
Overall satisfaction	What level of overall satisfaction did you experience with the activity as a whole? What aspects of the activity would you highlight as positive or negative?

Procedure

Before the experimental session, a clinical case was developed based on a pathology previously studied by students in the Mood Disorders module. Three clinical psychologists reviewed the case to ensure its suitability, after which the AI training process began. This involved two primary tasks: first, equipping the AI with sufficient information to address students' questions accurately, and second, fine-tuning its responses to maintain a consistent patient role. To create a realistic persona, the AI was given a detailed life history and personal profile, enhancing the coherence and authenticity of its responses (Pedrajas et al., 2024). Careful selection of verbs, instructions, and specific prompts shaped the AI's responses to embody the character and communication style needed for the exercise. The AI's identity was further defined by essential sociodemographic traits and communication aspects, supported by a dedicated chat and clear clinical context to prevent inconsistent responses. After establishing a coherent character profile, a pilot test was conducted with 10 subjects outside the experimental group to identify any unusual responses from the large language model (LLM). This preliminary test ensured the model's stability and reliability in delivering consistent, relevant answers. The clinical case is accessible at https://osf.io/se7dq/?view_only=891d4fb6d1304f6597496bfe69b29319.

With all materials prepared, the experiment was conducted in a single 2-hour session as part of the students' curriculum, facilitated by their regular professor and supported by two additional instructors. Each student had access to a computer and interacted with the same case study, in a psychological assessment first interview simulation scenario; however, each interaction was unique due to the LLM's adaptive responses. The clinical simulation process followed structured stages: *Prebriefing*, *Briefing, Clinical Setting*, and *Debriefing* (Duff et al., 2024; Kolbe et al., 2015) (Figure 1). Evaluation instruments, including PIES, PANAS, and HCAS, were administered individually both before the Prebriefing and after the Debriefing to assess changes across the simulation. The SPI-MET and focus group assessments, designed specifically for post-simulation feedback, were conducted only at the end of the session.

Figure 1
Stages of the Clinical Simulation Process



Data Analysis

The quantitative analysis assessed internal consistency using Cronbach's alpha and McDonald's omega for instrument reliability. Sensitivity analysis evaluated sample size adequacy (Lakens, 2022; Perugini et al., 2018). Repeated measures ANOVA compared pre- and post-scores to examine intervention effectiveness. Due to small sample size and limited gender diversity, gender variables were excluded. Bayesian hypothesis testing strengthened evidence for each instrument (Rouder et al., 2009, 2012).

Thematic analysis identified excerpts relevant to the research question (Nowell et al., 2017). Open-ended responses were descriptively coded, triangulated with focus group transcripts, and categorized manually, ensuring study reliability (Moser & Korstjens, 2018). Categories reflected variables measured by instruments to verify consistency. Adhering to COREQ guidelines (Tong et al., 2007), the mixed-methods approach enhanced quantitative reliability and deepened understanding of the phenomenon.

Results

The internal consistency of the positive and negative affect subscales of the PANAS, as well as the HCAS scale, was evaluated using Cronbach's alpha and McDonald's omega coefficients. The results indicated excellent internal consistency for both PANAS subscales in both pre and post assessments (Cronbach's $\alpha > 0.85$

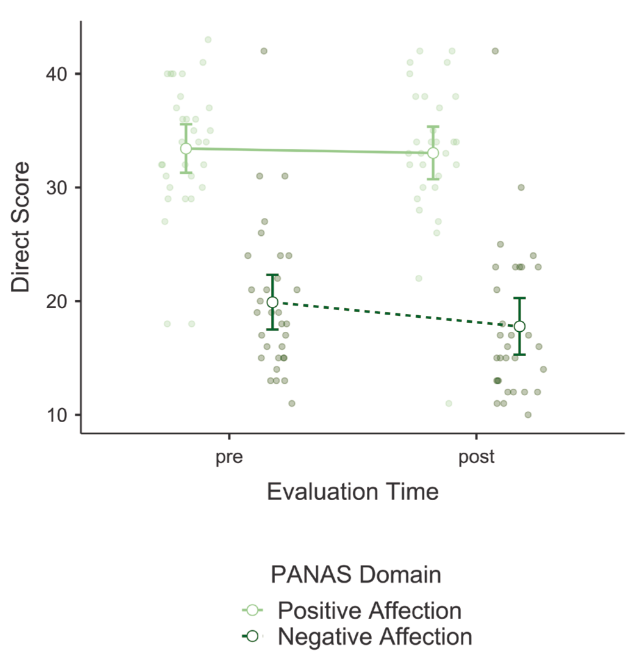
and McDonald's ω > 0.86). In contrast, the internal consistency of the HCAS scale was moderate to low, with Cronbach's alpha values of 0.570 (pre) and 0.500 (post), and McDonald's omega values of 0.695 (pre) and 0.701 (post).

Changes in Affective States

Sensitivity analysis on the PANAS scores was conducted using G*Power software (Faul et al., 2007) to determine the minimum detectable effect size for this study design. The analysis used four measures F tests, with an alpha level (α) of 0.05, an expected power of 0.85, and a total sample size of 31 participants. The result indicated that the minimum detectable effect corresponded to $F = 0.23$, with the critical F-value set at $F_{critical}(3, 90) = 2.71$. Therefore, F values equal to or larger than this threshold indicate a statistically meaningful result, corresponding to a probability of less than 5% under the null hypothesis.

Prior to conducting the analysis an inspection of distributional assumptions indicated no significant violations. Consequently, we proceeded with the planned analysis. The repeated measures ANOVA revealed significant main effects for both Affect (positive and negative) and evaluation time (pre- and post-simulation) on PANAS scores (Figure 2). A significant main effect of Affect was found ($F(1,30) = 55.17, p < .001, \eta^2 p = 0.648$) indicating that positive affect scores were significantly higher than negative affect scores ($M_{Diff} = 14.4$), with a large effect size demonstrating substantial impact. Additionally, a significant main effect of Evaluation Time was observed $F(1,30) = 15.67, p < .001, \eta^2 p = 0.343$, showing lower score after the intervention ($M_{Diff} = 1.26$). Furthermore, the interaction between affect and evaluation time was also significant ($F(1,30) = 5.06, p < .032, \eta^2 p = 0.144$). Post hoc analysis showed that negative affect scores decreased significantly from pre- to post-simulation ($M_{Diff} = 2.129$, $SE = 0.447$, $p_{Bonf} < .001$), while positive affect scores did not show significant differences from pre- to post-simulation ($M_{Diff} = 0.387$, $SE = 0.550$, $p_{Bonf} = 1.000$).

Figure 2
Mean Scores for PANAS Positive and Negative Domains with Error Bars Representing the 95% Confidence Intervals



The qualitative analysis reinforces the quantitative results, showing a clear prevalence of positive affective responses (19) over negative ones (15). Some participants highlighted difficulties in fully engaging due to the virtual nature of the interaction, noting that *"it's a more superficial situation than having the patient face-to-face" (GF:30)*, suggesting that the lack of direct contact may influence the perception of authenticity in the experience. However, both in the focus group and open-ended responses, the positive impact of this practice on the development of professional skills was confirmed. For example, *"it helps us improve our therapeutic skills and become familiar with some cases, to get some practical preparation before internships, especially for those of us who want to specialize in clinical practice. It really helps us lose that 'fear' of facing a patient, even if it's just a simulation" (P31)*, and in stress management, *"although it also helped ease my nerves not seeing the patient's face" (GF:7)*. A trend towards reduced anxiety among students was also observed, compared to face-to-face interactions with real patients: *"being able to ask directly is really helpful and makes you feel less shy than if they were in front of you" (GF:28)*. These findings suggest that AI can provide a less intimidating and more accessible learning environment, reducing anxiety and facilitating more effective development of clinical skills.

To further explore the effects on PANAS scores, we conducted Bayesian hypothesis testing comparisons to compute the Bayes Factors for the comparison of positive and negative affect scores across pre- and post-evaluation conditions. For negative affect scores, the Bayes Factor for the alternative hypothesis (BF_{10}) was 532.675, indicating extreme evidence in favor of the presence of a significant difference between pre- and post-simulation scores. Conversely, for positive affect scores, the Bayes Factor for the alternative hypothesis (BF_{10}) was 0.241, providing anecdotal evidence against a significant difference. The Bayes Factor for the null hypothesis (BF_{01}) was 4.153 for positive affect scores, supporting the absence of a meaningful difference across conditions. These results suggest a strong effect of evaluation time on negative affect but no substantial changes in positive affect scores.

Attitudes Toward Communication

For the HCAS scores, the sensitivity analysis was conducted using two measures with an alpha level of 0.05, an expected power of 0.85, and a sample size of 31 participants. The analysis determined that the minimum detectable effect size corresponded to $F = 0.28$, with a critical F-value of $F_{critical}(1, 30) = 4.17$, indicating the threshold for statistical significance at the given parameters.

A previous analysis of the normality distributional assumptions showed that normality was compromised. Thus, a complementary no-parametric analysis is presented to support our findings. The repeated measures ANOVA conducted on the HCAS scale revealed no significant main effect of evaluation time ($F(1, 30) = 0.017$, $p = .896$, $\eta^2 p = 0.001$). Similarly, the non-parametric analysis showed no significant differences on evaluation time (W(31) = 127, $p = 1.000$). Both tests indicate that there were no significant changes in attitudes toward healthcare communication from pre- to post-intervention, suggesting that the intervention did not influence these attitudes measurably (Figure 3).

Figure 3

Mean Scores for HCAS with Error Bars Representing the 95% Confidence Intervals

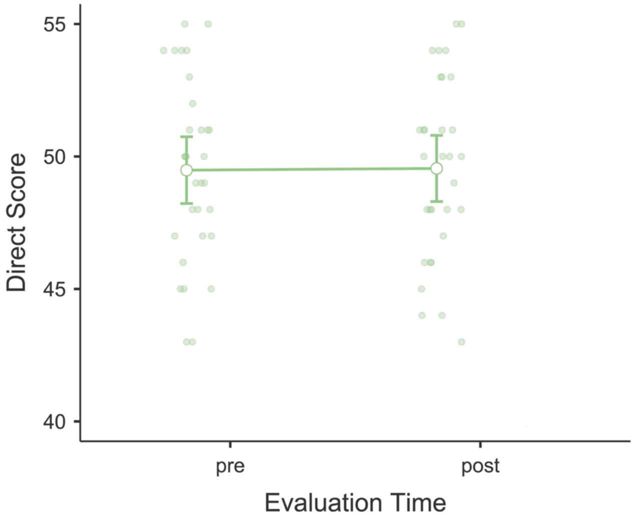
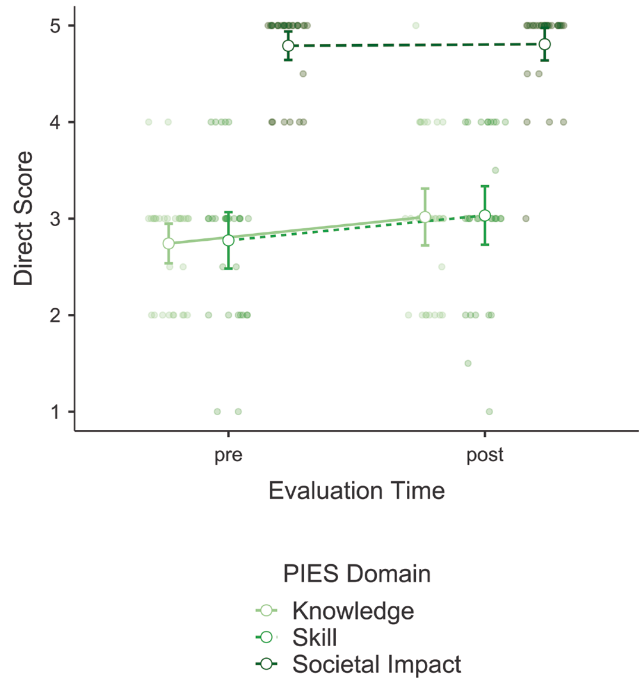


Figure 4

Mean Scores for PIES Domains with Error Bars Representing the 95% Confidence Intervals



The Bayes Factor for the alternative hypothesis $BF_{10} = 0.193$, indicating substantial evidence against a significant difference between pre- and post-evaluation scores. For the null hypothesis $BF_{01} = 5.178$, providing strong support for the absence of differences in HCAS scores across the two time points. These results suggest that there is no meaningful change in HCAS scores from pre- to post-simulation, providing robust evidence in favor of the null hypothesis. In this case, the qualitative data suggest that students perceive this practice as an opportunity to refine their already acquired skills. Some participants highlighted the ability to steer the direction of the conversation during the simulation, noting that they *"were able to practice changing the direction of the conversation based on the patient's responses" (GF:56)*. Additionally, the usefulness of these practices for applying theoretical knowledge and gaining confidence was emphasized: *"These types of practices help us put into practice all the theoretical knowledge we acquire and help us gain confidence in ourselves" (P13)*. However, there is no reference to the acquisition of new skills.

Perceptions of Knowledge, Skills, and Social Value

Similarly, analyzing the PIES scores with six measures and the same parameters showed that the minimal detectable effect size corresponds to $F = 0.20$ ($F_{critical}(1, 30) = 2.28$).

Prior to conducting the analysis, an inspection of distributional assumptions and Mauchly's test of sphericity indicated no significant violations, allowing us to proceed with the planned analysis. The repeated measures ANOVA for the PIES revealed significant main effects for both domain and evaluation time (Figure 4). A substantial effect of domain on PIES scores ($F(2, 60) = 150.58$, $p < .001$, $\eta^2p = .834$), indicated significant variability across the domains of knowledge, skills, and social value, highlighting a large effect size. Additionally, a significant main effect of evaluation time ($F(1, 30) = 7.52$, $p = .010$, $\eta^2p = 0.200$), showed notable changes in scores from pre- to post-evaluation, suggesting a medium effect size. However, the interaction between domain and evaluation time was not significant ($F(2, 60) = 2.84$, $p = .067$, $\eta^2p = 0.086$), indicating a small effect size and suggesting that changes over time did not differ significantly across the domains. Post hoc comparisons revealed that both knowledge ($M = 2.74$, $SD = 0.561$) and skills scores ($M = 2.77$, $SD = 0.794$) were significantly higher than social value scores ($M = 4.79$, $SD = 0.404$) (all $p < .001$), with large effect sizes evident in these differences. Paired samples t-tests further revealed that knowledge perception scores increased significantly from pre- ($M = 2.74$, $SD = 0.561$) to post-intervention ($M = 3.02$, $SD = 0.801$), ($t(30) = 2.655$, $p = .013$, $d = -0.477$), indicating a medium effect size. Similarly, skills perception scores increased significantly from pre- ($M = 2.77$, $SD = 0.794$) to post-intervention ($M = 3.03$, $SD = 0.826$), ($t(30) = 2.108$, $p = .043$, $d = 0.379$), also reflecting a medium effect size. In contrast, social value impact scores showed no significant difference from pre- ($M = 4.79$, $SD = 0.404$) to post-intervention ($M = 4.81$, $SD = 0.460$), ($t(30) = -0.329$, $p = .745$, $d = 0.059$), indicating a very small effect size. The qualitative analysis suggest that students perceive the simulation as an exercise comparable to a clinical interview with a real patient, which has allowed them to establish smooth communication and guide the interview towards the most relevant topics in the context of the case: *"It gave me the chance to practice not going blank and managing the process of organizing my thoughts" (GF:55)*. Additionally, they reported having applied the basic therapeutic skill of empathy during the exercise, despite it being an AI-based experience. This helped them identify the main clinical manifestations of the case, explore the problem's history, and suggest a potential psychopathological diagnosis *"We share a common fear, and these practices help you understand your tools" (GF:62)*, reflecting how the activity boosted their confidence in managing their clinical skills. Several students emphasized that the activity provided a valuable opportunity to apply the theoretical knowledge they had acquired in a simulated practical setting: *"I think it's a great activity and, overall, a fantastic initiative that, in my view, should be done more often" (P1)*.

Finally, the Bayesian paired samples t-tests for the alternative hypothesis (BF_{10} = 4.389) indicate moderate evidence in favor of a significant difference between pre- and post-evaluation PIES scores. Conversely, the Bayes Factor for the null hypothesis (BF_{01} = 0.228) provides weak evidence against the absence of differences. These results suggest that there is a notable change in PIES scores across the evaluation periods, with evidence supporting the presence of a significant effect.

Evaluation of the AI as a Simulated Patient

The post-simulation evaluation using SPI-MET assessed the students' perceptions of the performance of the large language model (LLM) acting as a simulated patient. Out of the total participants, two did not complete the instrument, resulting in 29 valid responses. Descriptive analysis revealed a mean SPI-MET score of 3.99 (*SD* = 0.597), indicating a generally favorable assessment of the LLM's performance as a simulated patient. Similarly, the mean score given by the students for the additional question, "How would you rate the tool's ability as a simulated patient on a scale of 0 to 10?" was 8.28 (*SD* = 1.13). The frequency distribution showed that 3.4% of the students rated the LLM's performance as 6, 24.1% rated it as 7, 31.0% rated it as 8, 24.1% rated it as 9, and 17.2% rated it as 10. These results indicate that the majority of students rated the LLM's performance highly, with most ratings falling between 7 and 9, suggesting a generally positive perception of the LLM's effectiveness in simulating patient interactions. These data are corroborated in the focus group comments, which highlight how realistic the practice felt: *"It didn't seem impersonal, the responses were long because we weren't face-to-face, but the language used was conversational" (GF:17)*. Another participant added: *"It seemed very realistic to me" (GF:61)*. The ease of maintaining a meaningful conversation was emphasized, and it was noted that the system's ability to provide detailed responses was likely due to the lack of direct visual communication. Several students pointed out that the language used was notably straightforward: *"The language didn't seem unrealistic to me, but it was a bit formal and direct" (GF:13)*. However, some expressed that the IA responses could feel cold and repetitive, with a certain robotic quality, though this did not significantly impact the empathetic nature of the interaction *"The patient repeated the same thing several times, even when asked to elaborate" (P23)*.

Discussion

In psychology education, theoretical knowledge of mental disorders and interventions must be paired with practical experiences to prepare students for real-world challenges. Clinical simulations bridge theory and practice, fostering competencies in a safe environment. However, research on their effectiveness in psychology is limited (Ruiz-Rodríguez et al., 2016; Rodríguez et al., 2021). This study assessed the feasibility and educational value of using AI-based simulations in a first-session patient interview scenario.

This study revealed several notable findings. Quantitatively, significant reductions in negative affect (PANAS) were observed post-intervention, emphasizing the ability of AI-based simulations to mitigate anxiety and stress in a controlled, low-risk environment. Qualitative data reinforced the observed reductions in students' negative affect, indicating that this intervention helped to mitigate stress and anxiety—emotions that can negatively impact clinical performance. These outcomes align with previous research demonstrating the effectiveness of simulation-based training in reducing anxiety and building confidence compared to traditional methods (Abbott et al., 2021; Oliveira et al., 2022). However, no significant changes were detected in positive affect, potentially reflecting a ceiling effect or suggesting that the intervention primarily targeted stress reduction rather than enhancing positive emotional states. This fact is reflected in stress and coping theories (Lazarus & Folkman, 1984), which propose that interventions targeting perceived stress—such as simulated practice—can effectively lower negative affect without necessarily increasing positive emotions.

Similarly, the analysis of PIES scores demonstrated a significant improvement in students' perceptions of their knowledge and clinical skills, further supporting the pedagogical value of AI-simulated patients. Such an approach aligns with existing theoretical models, such as experiential learning theory (Kolb, 2014), which emphasizes the importance of hands-on, reflective practice in skill acquisition. Qualitative findings provided additional depth to these results. Students reported feeling more confident and better prepared to handle clinical scenarios after the intervention, likely due to the controlled environment that simulations offer (Elendu et al., 2024). They appreciated the opportunity to apply theoretical knowledge in a simulated practical setting, particularly valuing the structured feedback and safe environment that allowed them to refine their communication and diagnostic skills. These observations align with previous studies on virtual patients, which have highlighted their value in developing essential health science skills such as clinical interviewing and reasoning (Sezer et al., 2023; Cho & Kim, 2024; Jeon et al., 2024). Conversely, scores from the HCAS indicated no significant changes in communication attitudes, suggesting that this aspect might require more prolonged or varied interventions for measurable improvements.

As per the evaluation of the AI as a simulated patient, students described their interactions with ChatGPT as strikingly similar to real-life conversations. They valued the natural language and conversational flow, although some noted that the lack of visual interaction allowed for more detailed verbal responses. A few students did report that certain responses felt somewhat repetitive or lacked emotional depth, indicating that while ChatGPT performs effectively as a simulated patient, improvements in emotional expressiveness and naturalism are still possible.

The results suggest that incorporating AI-simulated patients can foster an active learning environment where students practice basic clinical and communication skills in controlled, simulated scenarios (Alrashidi et al., 2023; Liu et al., 2023; Farina et al., 2024). These initial findings indicate that such simulations may contribute to a more dynamic and participatory learning process, providing opportunities for students to apply theoretical knowledge in a practical setting (Higgins et al., 2021). While preliminary, the present findings point to the potential of AI-based simulation, specifically using ChatGPT, as a complementary pedagogical tool in psychology education. This approach offers a promising addition to traditional methods, providing opportunities for experiential learning that are scalable and adaptable. The value of integrating new technologies into psychological and educational training has been highlighted by Elosua et al. (2023), suggesting that such tools can support student preparedness and confidence in professional skills.

This study highlights limitations, notably the rapid evolution of AI technologies, which complicates their long-term applicability in clinical education. The research evaluated AI as a simulated patient for basic therapist competencies, such as communication, emotional engagement, and interaction, but did not address its potential in enhancing diagnostic accuracy or advanced therapeutic skills. It also lacked comparisons between AI modalities like audio systems and chatbots, which could enhance realism. Ensuring AI aligns with psychometric standards and mental health frameworks is essential (Elosua et al., 2023).

Future research should examine AI-based simulations' transferability to clinical settings, their impact on diagnostic and decision-making skills, and integration into broader pedagogical frameworks. Advances in NLP and machine learning, such as AIML and LLMs, offer increased flexibility and interaction complexity but face challenges in ensuring safety, explainability, and domain-specific accuracy. Combining rule-based and LLM approaches and integrating conversational models into real-world environments with human-like agents presents both opportunities and challenges (Talbot & Rizzo, 2019).

Author Contributions

Ana Sanz: Conceptualization, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. **Jose Luis Tapia**: Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Eva Garcia-Carpintero**: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. **Francisco Rocabado**: Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Lorena Pedrajas**: Conceptualization, Investigation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing.

Funding

This work was supported by the State Plan for Scientific and Technical Research and Innovation of the Government of Spain (FPU19/02239); and by the Ministry of Science, Innovation and Universities, the State Research Agency, and the European Social Fund Plus (JDC2023-052644-I). This funding source had no role in the design of this study, data collection, management, analysis, and interpretation of data, writing of the manuscript, and the decision to submit the manuscript for publication.

Declaration of Interests

The authors declare that there is no conflict of interest.

Data Availability Statement

The data supporting the findings of this research are accessible at the following link: https://osf.io/se7dq/?view_only=891d4fb6d1304f6597496bfe69b29319

References

Abbott, E. F., Laack, T. A., Licatino, L. K., Wood-Wentz, C. M., Warner, P. A., Torsher, L. C., Newman, J. S., & Rieck, K. M. (2021). Comparison of dyad versus individual simulation-based training on stress, anxiety, cognitive load, and performance: a randomized controlled trial. *BMC Medical Education, 21*(1), Article 367. https://doi.org/10.1186/s12909-021-02786-6

Alrashidi, N., Pasay An, E., Alrashedi, M. S., Alqarni, A. S., Gonzales, F., Bassuni, E. M., Pangket, P., Estadilla, L., Benjamin, L. S., & Ahmed, K. E. (2023). Effects of simulation in improving the self-confidence of student nurses in clinical practice: a systematic review. *BMC Medical Education, 23*(1), Article 815. https://doi.org/10.1186/s12909-023-04793-1

Baile, J. I. (2024). Patient with depression created by freely accessible artificial intelligence for the teaching of Psychology. Preliminary study of its validity. *Tecnología, Ciencia y Educación*, *27*, 7-42. https://doi.org/10.51302/tce.2024.19069

Basnet, S., Shrestha, S. P., Shrestha, R., Shrestha, A. P., Shrestha, A., Sahu, S., Mhatre, B., & Silwal, P. (2024). Effect of simulation-based emergency airway management education on the knowledge, skills and perceived confidence of medical interns. *Annals of Medicine and Surgery, 86*(9), 5191-5198. https://doi.org/10.1097/MS9.0000000000002376

Burrell, S. A., Ross, J. G., D'Annunzio, C., & Heverly, M. (2023). Standardized patient simulation in an oncology symptom management seminar-style course: prelicensure nursing student outcomes. *Journal of Cancer Education: the official journal of the American Association for Cancer Education, 38*(1), 185-192. https://doi.org/10.1007/s13187-021-02096-x

Carrero-Planells, A., Pol-Castañeda, S., Alamillos-Guardiola, M. C., Prieto-Alomar, A., Tomás-Sánchez, M., & Moreno-Mulet, C. (2021). Students and teachers' satisfaction and perspectives on high-fidelity simulation for learning fundamental nursing procedures: A mixed-method study. *Nurse Education Today, 104*, Article 104981. https://doi.org/10.1016/j.nedt.2021.104981

Clinard, E. S. (2022). Increasing student confidence with medically complex infants through simulation: a mixed methods investigation. *American Journal of Speech-language Pathology, 31*(2), 942-958. https://doi.org/10.1044/2021_AJSLP-21-00234

Cho, M. K., & Kim, M. Y. (2024). The effect of virtual reality simulation on nursing students' communication skills: a systematic review and meta-analysis. *Frontiers in Psychiatry, 15*, Article 1351123. https://doi.org/10.3389/fpsyt.2024.1351123

Curry, L., & Nunez-Smith, M. (2015). *Mixed methods in health sciences research. A practical primer*. SAGE Publications, Inc.

Dawood, E., Alshutwi, S. S., Alshareif, S., & Shereda, H. A. (2024). Evaluation of the effectiveness of standardized patient simulation as a teaching method in psychiatric and mental health nursing. *Nursing Reports, 14*(2), 1424-1438. https://doi.org/10.3390/nursrep14020107

Duff, J. P., Morse, K. J., Seelandt, J., Gross, I. T., Lydston, M., Sargeant, J., Dieckmann, P., Allen, J. A., Rudolph, J. W., & Kolbe, M. (2024). Debriefing methods for simulation in healthcare: a systematic review. *Simulation in Healthcare: Journal of the Society for Simulation in Healthcare*, *19*(1S), S112-S121. https://doi.org/10.1097/SIH.0000000000000765

Escribano, S., Juliá-Sanchis, R., García-Sanjuán, S., Congost-Maestre, N., & Cabañero-Martínez, M. J. (2021). Psychometric properties of the attitudes towards medical communication scale in nursing students. *PeerJ*, *9*, Article e11034. https://doi.org/10.7717/peerj.11034

Elendu, C., Amaechi, D. C., Okatta, A. U., Amaechi, E. C., Elendu, T. C., Ezeh, C. P., & Elendu, I. D. (2024). The impact of simulation-based training in medical education: A review. *Medicine, 103*(27), Article e38813. https://doi.org/10.1097/MD.0000000000038813

Elosua, P., Aguado, D., Fonseca-Pedrero, E., Abad, F. J., & Santamaría, P. (2023). New trends in digital technology-based psychological and educational assessment. *Psicothema*, *35*(1), 50-57. https://doi.org/10.7334/psicothema2022.241

Farina, C. L., Moreno, J., & Schneidereith, T. (2024). Using simulation to improve communication skills. *The Nursing clinics of North America, 59*(3), 437-448. https://doi.org/10.1016/j.cnur.2024.02.007

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. https://doi.org/10.3758/BF03193146

Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs-principles and practices. *Health Services Research, 48*(6 Pt 2), 2134-2156. https://doi.org/10.1111/1475-6773.12117

Flanagan, O. L., & Cummings, K. M. (2023). Standardized patients in medical education: a review of the literature. *Cureus, 15*(7), Article e42027. https://doi.org/10.7759/cureus.42027

García-Carpintero Blas, E., Vélez-Vélez, E., Gómez-Moreno, C., Martínez-Arce, A., Tovar-Reinoso, A., Rodriguez-Gómez, P., Vaquero Velerdas, L., & López-Martín, I. (2024). Simulation with a standardised patient to reduce stigma towards people with schizophrenia spectrum disorder among nursing students: A quasi-experimental study. *Archives of Psychiatric Nursing, 52*, 24-30. https://doi.org/10.1016/j.apnu.2024.07.015

Gerzina, H. A., & Stovsky, E. (2023). Standardized patient assessment of learners in medical simulation. In *StatPearls [Internet]*. StatPearls Publishing.

Gutiérrez-Maldonado, J., Alsina-Jurnet, I., Rangel-Gómez, M. V., Aguilar-Alonso, A., Jarne-Esparcia, A. J., Andrés-Pueyo, A., & Talarn-Caparrós, A. (2008). Virtual intelligent agents to train abilities of diagnosis in Psychology and Psychiatry. In G. A. Tsihrintzis, M. Virvou, R. J. Howlett, & L. C. Jain (Eds.), *New directions in intelligent interactive multimedia* (pp. 497-505). Springer. https://doi.org/10.1007/978-3-540-68127-4_51

Gutierrez-Maldonado, J., Andres-Pueyo, A., Jarne, A., Talarn, A., Ferrer, M., & Achotegui, J. (2017). Virtual reality for training diagnostic skills in Anorexia Nervosa: A usability assessment. In S. Lackey, & J. Chen (Eds.), *Virtual, augmented and mixed reality* (pp. 239-247). Springer International Publishing. https://doi.org/10.1007/978-3-319-57987-0_19

Hawkins, A., & Tredgett, K. (2016). Use of high-fidelity simulation to improve communication skills regarding death and dying: a qualitative study. *BMJ Supportive & Palliative Care, 6*(4), 474-478. https://doi.org/10.1136/bmjspcare-2015-001081

Higgins, M., Madan, C., & Patel, R. (2021). Development and decay of procedural skills in surgery: A systematic review of the effectiveness of simulation-based medical education interventions. *The Surgeon: Journal of the Royal Colleges of Surgeons of Edinburgh and Ireland, 19*(4), e67-e77. https://doi.org/10.1016/j.surge.2020.07.013

Hillier, M., Williams, T. L., & Chidume, T. (2023). Standardization of standardized patient training in medical simulation. In *StatPearls [Internet]*. StatPearls Publishing.

Isaza-Restrepo, A., Gómez, M. T., Cifuentes, G., & Argüello, A. (2018). The virtual patient as a learning tool: a mixed quantitative qualitative study. *BMC Medical Education, 18*(1), Article 297. https://doi.org/10.1186/s12909-018-1395-8

Jeon, Y., Choi, H., Lee, U., & Kim, H. (2024). Technology-based interactive communication simulation addressing challenging communication situations for nursing students. *Journal of Professional Nursing: Official Journal of the American Association of Colleges of Nursing, 53*, 71-79. https://doi.org/10.1016/j.profnurs.2024.05.002

Johnson S. L. (2019). Impact, growth, capacity-building of mixed methods research in the Health Sciences. *American Journal of Pharmaceutical Education, 83*(2), Article 7403. https://doi.org/10.5688/ajpe7403

Johnson, K. V., Scott, A. L., & Franks, L. (2020). Impact of standardized patients on first semester nursing students self-confidence, satisfaction, and communication in a simulated clinical case. *SAGE Open Nursing, 6*, Article 2377960820930153. https://doi.org/10.1177/2377960820930153

Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development*. FT press.

Kolbe, M., Grande, B., & Spahn, D. R. (2015). Briefing and debriefing during simulation-based training and beyond: Content, structure, attitude and setting. *Best Practice & Research. Clinical Anaesthesiology*, *29*(1), 87-96. https://doi.org/10.1016/j.bpa.2015.01.002

Lakens, D. (2022). Sample size justification. *Collabra: Psychology, 8*(1), Article 33267. https://doi.org/10.1525/collabra.33267

Lazarus, R. S. & Folkman, S. (1984). *Stress, appraisal, and coping* (Vol. 464). Springer.

Liu, X., Wu, C., Lai, R., Lin, H., Xu, Y., Lin, Y., & Zhang, W. (2023). ChatGPT: when the artificial intelligence meets standardized patients in clinical training. *Journal of Translational Medicine, 21*, Article 447. https://doi.org/10.1186/s12967-023-04314-0

López, A.I., & López-Chicheri, I. (July 2024). *Entrenamiento en habilidades de comunicación orientadas a la intervención mediante simulación clínica en psicología (SCP) [Intervention-oriented communication skills training through clinical simulation in Psychology (SCP)]*. I Congreso Interprofesional de Simulación-OneHealth. https://simulacion-onehealth.org/ponencia/entrenamiento-en-habilidades-de-comunicacion-orientadas-a-la-intervencion-mediante-simulacion-clinica-en-psicologia-scp/

López-Gómez, I., Hervás, G., & Vázquez, C. (2015). Adaptación de las" Escalas de afecto positivo y negativo"(PANAS) en una muestra general española. [Adaptation of the Positive and Negative Affect Scales (PANAS) in a general Spanish sample]. *Psicología Conductual*, *23*(3), 529-548.

Monahan, L., Eaves, C. L., Watson, J. C., Friese, J., McKenna, L., & Estrada-Ibarra, E. (2024). Improving adolescent psychosocial assessment through standardized patient simulation: an interdisciplinary quality improvement initiative. *International Journal of Environmental Research and Public Health, 21*(3), Article 283. https://doi.org/10.3390/ijerph21030283

Morcela, O. A. (2022). ChatGPT: la IA está aquí y nos desafía [ChatGPT: AI is here and challenging us]. *AACINI-Revista Internacional de Ingeniería Industrial*, *6*. https://riii.fi.mdp.edu.ar/index.php/AACINI-RIII/article/view/67

Moser, A., & Korstjens, I. (2018). Series: Practical guidance to qualitative research. Part 3: Sampling, data collection and analysis. *The European Journal of General Practice, 24*(1), 9-18. https://doi.org/10.1080/13814788.2017.1375091

Moss, C. R. (2023). Neonatal fragile skin: novel use of simulation to improve knowledge and confidence for neonatal nurse practitioner students. *Nurse Educator, 48*(4), E122-E125. https://doi.org/10.1097/NNE.0000000000001354

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, *16*(1). https://doi.org/10.1177/1609406917733847

Oliveira, G., Oliveira, F. S. E., Coelho, A. S. G., Cavalcante, A. M. R. Z., Vieira, F. V. M., Fonseca, L. M. M., Campbell, S. H., & Aredes, N. D. A. (2022). Effect of simulation on stress, anxiety, and self-confidence in nursing students: Systematic review with meta-analysis and metaregression. *International Journal of Nursing Studies, 133*, Article 104282. https://doi.org/10.1016/j.ijnurstu.2022.104282

OpenAI. (2023). *ChatGPT (March 14 version) [Large language model]*. https://chat.openai.com

Pedrajas, M. L., Sanz, A., García-Carpintero, E., Martínez, E., Uceda, S. (2024). Training GPT as a standardized patient. In M. D. Díaz-Noguera, C. Hervás-Gómez, & F. Sánchez-Vera (Coords.), *Artificial Intelligence and Education* (pp.189-204). Octaedro. https://doi.org/10.36006/09643-1-12

Peñaloza-Salazar, C., Gutierrez-Maldonado, J., Ferrer-Garcia, M., Garcia-Palacios, A., Andres-Pueyo, A., & Aguilar-Alonso, A. (2011). Simulated interviews 3.0: virtual humans to train abilities of diagnosis –usability assessment. *Studies in Health Technology and Informatics*, *167*, 165-169. https://doi.org/10.3233/978-1-60750-766-6-165

Perugini, M., Gallucci, M., & Costantini, G. (2018). A practical primer to power analysis for simple experimental designs. *Revue Internationale de Psychologie Sociale, 31*(1), 1-23. https://dx.doi.org/10.5334/IRSP.181

Rizzo, A., Lange, B., Buckwalter, J. G., Forbell, E., Kim, J., Sagae, K., Williams, J., Difede, J., Rothbaum, B. O., Reger, G., Parsons, T., & Kenny, P. (2011). SimCoach: An intelligent virtual human system for providing healthcare information and support. *International Journal on Disability and Human Development, 10*(4), 277-281. https://doi.org/10.1515/IJDHD.2011.046

Rodríguez, S., Condés, E., & Arriaga, A. (2021). Irrupción de la simulación clínica online en tiempos de COVID-19. Una experiencia ilustrativa de asignatura en el Grado de Psicología. *Revista de la Fundación de Educación Médica, 24*(2), 101-104. https://dx.doi.org/10.33588/fem.242.1118

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*(5), 356-374. https://doi.org/10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225-237. https://doi.org/10.3758/PBR.16.2.225

Ruiz-Rodríguez, J., Bados López, A., Fusté Escolano, A., García-Grau, E., Saldaña García, C., & Lluch Canut, T. (2016). Aprendizaje experiencial de habilidades terapéuticas y análisis de su utilidad en función de la personalidad. *Behavioral Psychology*, *24,* 405-422. http://hdl.handle.net/2445/155366

Scherr, R., Halaseh, F. F., Spina, A., Andalib, S., & Rivera, R. (2023). ChatGPT interactive medical simulations for early clinical education: case study. *JMIR Medical Education, 9*, Article e49877. https://doi.org/10.2196/49877

Sezer, B., Sezer, T. A., Teker, G. T., & Elcin, M. (2023). Developing a virtual patient: design, usability, and learning effect in communication skills training. *BMC Medical Education, 23*(1), Article 891. https://doi.org/10.1186/s12909-023-04860-7

Sezgin, M. G., & Bektas, H. (2023). Effectiveness of interprofessional simulation-based education programs to improve teamwork and communication for students in the healthcare profession: A systematic review and meta-analysis of randomized controlled trials*. Nurse Education Today, 120*, Article 105619. https://doi.org/10.1016/j.nedt.2022.105619

Talbot, T., & Rizzo, A. “Skip”. (2019). Virtual human standardized patients for clinical training. In A. “Skip” Rizzo & S. Bouchard (Eds.), *Virtual reality for psychological and neurocognitive interventions* (pp. 387-405). Springer New York. https://doi.org/10.1007/978-1-4939-9482-3_17

Ton, D. N. M., Duong, T. T. K., Tran, H. T., Nguyen, T. T. T., Mai, H. B., Nguyen, P. T. A., Ho, B. D., & Ho, T. T. T. (2024). Effects of standardized patient simulation and mobile applications on nursing students’ clinical competence, self-efficacy, and cultural competence: A quasi-experimental study. *International Journal of Environmental Research and Public Health, 21*(4), Article 515. https://doi.org/10.3390/ijerph21040515

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care: Journal of the International Society for Quality in Health Care*, *19*(6), 349-357. https://doi.org/10.1093/intqhc/mzm042

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology, 54*(6), 1063-1070. https://doi.org/10.1037/0022-3514.54.6.1063

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Commun. ACM*, *9*(1), 36-45. https://doi.org/10.1145/365153.365168

Yu, J. H., Chang, H. J., Kim, S. S., Park, J. E., Chung, W. Y., Lee, S. K., Kim, M., Lee, J. H., & Jung, Y. J. (2021). Effects of high-fidelity simulation education on medical students’ anxiety and confidence. *PloS one, 16*(5), Article e0251078. https://doi.org/10.1371/journal.pone.0251078