Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Anxiety, self-efficacy and resilience following robotic mental health simulation among undergraduate nursing students: a quasi-experimental study with three-month follow-up

Fares Hameed Alshammari
fhalshammari@uhb.edu.sa
Hafer Al-batin University
Mahmoud Khedr
Hafer Al-batin University
Suliman Basahal
King Abdulaziz University
Salma Gomaa
Hafer Al-batin University
Magda Ahmed
Hafer Al-batin University
Ryam Alotaibi
Hafer Al-batin University
Essam Eltantawy
Hafer Al-batin University

Research Article

Keywords: Psychiatric nursing education, Nursing students, Robotic simulation, Anxiety, Self-efficacy, Resilience, Quasi-experimental study

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11500283/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Preparing nursing students for psychiatric practice requires educational approaches that address psychological demands alongside clinical learning. Evidence concerning psychological outcomes following robotic mental health simulation remains limited. This study compared anxiety, general self-efficacy and resilience between undergraduate nursing students participating in a structured robotic simulation program and those receiving standard preparation.

Methods

A nonrandomized, quasi-experimental study included 80 undergraduate nursing students at the University of Hafr Al-Batin, Saudi Arabia, with 40 students per condition allocated through existing clinical sections. The intervention comprised six robotic mental health simulation sessions over eight weeks, incorporating prebriefing, simulated encounters and reflective debriefing. Outcomes were assessed using the Beck Anxiety Inventory, General Self-Efficacy Scale and 25-item Connor–Davidson Resilience Scale at baseline, immediately after the program and three months later. Exploratory between-group comparisons used Mann–Whitney U tests, with Holm adjustment across six post-intervention comparisons. Analyses did not adjust for baseline scores or section-level clustering.

Results

Three-month assessments were available for 36 control and 37 intervention participants. One incomplete immediate intervention-group resilience questionnaire was excluded from scoring. Mean self-efficacy scores were higher in the intervention group immediately after the program (36.48 versus 29.40; Holm-adjusted $p < 0.001$) and at three months (35.54 versus 29.14; $p = 0.001$). Mean resilience scores were also higher immediately (76.69 versus 56.03; $p < 0.001$) and at follow-up (74.27 versus 61.33; $p = 0.022$). Anxiety scores were numerically lower in the intervention group, but between-group comparisons were not statistically significant at either post-intervention assessment (adjusted $p = 0.172$ for both).

Conclusions

Participation in robotic mental health simulation was associated with higher post-intervention self-efficacy and resilience scores, whereas anxiety comparisons were inconclusive. Nonrandom allocation, unadjusted analyses and unverified longitudinal linkage limit causal interpretation. These preliminary findings support further comparative evaluation of the program as an adjunct to psychiatric nursing education; they do not establish robot-specific benefits or improved clinical competence.

Background

Preparing nursing students for psychiatric practice requires attention to both clinical learning and the psychological demands of therapeutic engagement. Students must learn to interpret unfamiliar presentations, communicate sensitively and respond to distress while managing uncertainty about their own capabilities. Clinical placement anxiety can complicate this transition from theoretical preparation for patient care [1]. Within Saudi Arabia, qualitative research has identified apprehension, anxiety, fear and stress alongside opportunities to develop therapeutic communication and apply theoretical knowledge during mental health placements [2]. These experiences establish an educational need for structured preparation that addresses students' emotional responses while supporting respectful, person-centered engagement with people experiencing mental health difficulties.

Simulation offers opportunities to rehearse clinical encounters, receive feedback and reflect on decisions before or alongside supervised patient contact. A systematic review and meta-analysis found favorable effects of simulation on nursing students' anxiety and self-confidence compared with conventional teaching approaches [3]. Within psychiatric nursing, Piot et al. identified benefits for selected skills and attitudinal outcomes, although substantial variation existed across interventions and study designs [4]. More recently, a systematic review of mental health simulations reported improvements in nursing students' knowledge and performance, while highlighting heterogeneity and methodological variation [5]. This evidence supports simulation as an educational approach but does not establish that different simulation modalities produce equivalent outcomes. Improvements in knowledge or task performance cannot be assumed to represent broader changes in students' psychological resources.

Anxiety, general self-efficacy and resilience provide complementary perspectives for examining these resources. General self-efficacy concerns beliefs about the ability to manage challenging demands across situations, whereas resilience concerns adaptive capacity in the face of stress and adversity [6, 7]. These constructs are relevant to preparation for emotionally demanding encounters, but they should remain conceptually distinct. Students may perceive themselves as more capable of managing challenges without experiencing an equivalent reduction in anxiety. Similarly, higher self-reported resilience does not necessarily indicate stronger clinical performance. Assessing the three outcomes together therefore allows a more differentiated evaluation of psychological responses to an educational program than reliance on confidence or satisfaction alone.

Bandura's self-efficacy theory provides a rationale for examining how structured simulation could influence perceived capability [8]. Successful performance, observation of others, constructive feedback and the interpretation of emotional and physiological responses can inform efficacy beliefs. Applied to psychiatric nursing education, repeated encounters and reflective debriefing may help students recognize effective responses, reconsider perceived difficulties and approach subsequent situations with greater confidence. These processes offer plausible explanations for psychological outcomes, rather than established mechanisms of robotic simulation. Moreover, because learning occurs within specific scenarios, transfer from task-related success to general self-efficacy or resilience requires direct investigation [6, 8].

Humanoid patient robots introduce an embodied format for presenting selected psychiatric symptoms and interpersonal cues. Developmental research has demonstrated the technical feasibility of coordinating speech, facial expressions and movements to support psychiatric diagnostic training [9]. Such systems may

enable planned presentations to be repeated within structured learning encounters. However, physical embodiment and programmable behavior do not, by themselves, establish educational superiority. In a small pilot study comparing a humanoid robot with video presentations of a human actor, participants generally perceived the actor presentation as more realistic, while diagnostic outcomes showed little difference between groups [10]. These findings support further evaluation of robotic simulation while emphasizing the need to distinguish technical feasibility, perceived realism and psychological or clinical learning outcomes.

The evidence reviewed therefore leaves a specific question unresolved: whether participation in a structured robotic mental health simulation program is associated with differences in nursing students' anxiety, general self-efficacy and resilience beyond the immediate training period. Existing simulation reviews establish a broader educational rationale, while robot-focused studies have principally examined symptom presentation, diagnostic performance and acceptance [3–5, 9, 10]. These findings do not establish the psychological outcomes of a program combining robotic encounters with prebriefing, repeated practice and reflective debriefing. Evaluating this combined program against standard preparation can inform its educational relevance, although such a comparison cannot isolate the independent contribution of the robot.

Accordingly, this quasi-experimental study compared anxiety, general self-efficacy and resilience among undergraduate nursing students participating in robotic mental health simulation and those receiving standard preparation at the University of Hafr Al-Batin, Saudi Arabia. Assessments were conducted at baseline, immediately after the program and three months later. The study aimed to provide preliminary comparative evidence on three distinct psychological outcomes and inform the design of subsequent evaluations of robotic simulation in psychiatric nursing education.

Methods

Study design and setting

A nonrandomized, two-group quasi-experimental study was conducted at the College of Nursing, University of Hafr Al-Batin, Saudi Arabia. Outcomes were assessed at baseline, immediately after an eight-week robotic mental health simulation program and three months after program completion. Recruitment took place during the first semester of the 2025–2026 academic year. Participants were enrolled in the Psychiatric and Mental Health Nursing course, which comprised 180 contact hours encompassing theoretical instruction, laboratory learning and supervised clinical practice.

Reporting was informed by the Transparent Reporting of Evaluations with Nonrandomized Designs (TREND) statement and the Template for Intervention Description and Replication (TIDieR) checklist [11, 12].

Participants and allocation

A convenience sample of 80 undergraduate nursing students participated, with 40 students in the intervention group and 40 in the control group. Recruitment occurred during scheduled classes, and written informed consent was obtained before enrolment. Eight students involved in preliminary pilot procedures were excluded from the main sample.

Allocation followed existing clinical practicum sections rather than individual randomization. Both conditions followed the same course specification, assignments and assessment requirements. The allocation arrangement accommodated established teaching schedules and was intended to reduce exchange of intervention content between groups; contamination was not formally assessed. The five groups of eight students used to deliver the intervention were instructional groups and should not be interpreted as independent allocation units.

The original eligibility specification required first-time enrolment in the psychiatric nursing course, no previous robotic simulation training and willingness to provide informed consent. Exclusion criteria included a reported diagnosed psychiatric disorder or advanced psychiatric clinical experience. However, the baseline dataset recorded previous robotic psychiatric simulation exposure for four control and six intervention participants. This discrepancy remained unresolved in the available documentation; the recorded categories were retained without retrospective reclassification.

Sample-size planning

The recruitment target was 80 students. The available planning record specified G*Power version 3.1, an assumed effect size of f = 0.25, a significance level of 0.05, statistical power of 80%, two groups, three measurement occasions, a repeated-measures correlation of 0.50 and a nonsphericity correction of 1.00 [13]. However, the exact statistical test and original calculation output supporting the previously reported minimum of 68 participants were unavailable. Consequently, that minimum could not be independently reproduced, and adequate power for the exploratory analyses reported here was not assumed.

Robotic mental health simulation program

The intervention comprised six sessions delivered over eight weeks in the nursing simulation laboratory. The 40 intervention participants were organized into five learning groups of eight students. Laboratory teaching was scheduled on Sundays and Tuesdays within a 09:00–13:00 period. This represented the available teaching window rather than the duration of individual exposure.

The program used a research-team-developed humanoid robotic patient capable of presenting selected psychiatric symptoms through prerecorded speech, facial expressions and limb movements. An operator controlled the system using a tablet-based graphical user interface developed in Python. The platform was operator-controlled rather than autonomous. Psychiatric nursing faculty reviewed the scenarios for clinical relevance and alignment with course objectives.

The six scenarios addressed therapeutic communication and relationship development; assessment and supportive care for anxiety; recognition of depressive symptoms; communication with individuals experiencing psychotic symptoms; de-escalation and crisis response; and suicide risk assessment and therapeutic emotional support. Table 1 summarizes the scenarios, activities and intended learning outcomes.

Table 1
Planned structure, educational activities, and intended learning outcomes of the robotic mental health simulation programme (intervention group, n = 40)

Session and scenario	Prebriefing: 10–15 min	Simulation: 20–30 min	Debriefing: 15–20 min	Intended learning outcomes
1. Therapeutic communication and relationship development	Introducing the environment, learning objectives, participant roles, and expectations for psychologically safe participation.	Practice initiating therapeutic interaction and establishing rapport with the simulated patient.	Reflect on communication choices, empathy, self-awareness, and factors that supported or hindered rapport.	Demonstrate communication techniques that support therapeutic engagement and identify opportunities to improve nurse–patient interactions.
2. Nursing assessment and support for anxiety	Review anxiety manifestations, assessment principles, and supportive communication approaches.	Assess the simulated patient's anxiety-related presentation and practice therapeutic communication and supportive nursing responses.	Discuss assessment findings, communication effectiveness, personal emotional responses, and the rationale for selected interventions.	Identify anxiety-related manifestations and select appropriate assessment and communication approaches.
3. Nursing assessment and support for depressive symptoms	Review depressive manifestations, principles of emotional support, and person-centered communication.	Practice mental status examination and therapeutic engagement with a simulated patient presenting with depressive symptoms.	Reflect on assessment findings, empathy, therapeutic presence, and supportive nursing responses.	Recognize depressive manifestations and demonstrate assessment and communication approaches that support therapeutic engagement.
4. Therapeutic engagement with psychotic symptoms	Review hallucinations, delusions, altered thought processes, communication principles, and relevant safety considerations.	Practice assessment and therapeutic interaction with a simulated patient presenting with psychotic symptoms.	Examine communication challenges, clinical reasoning, emotional responses, and person-centered alternatives.	Identify relevant psychotic symptoms and demonstrate communication approaches appropriate to the simulated presentation.
5. De-escalation and crisis	Review risk assessment, de-	Recognize escalating	Reflect on situational	Recognize behavioral

Table notes: RMHS, robotic mental health simulation. The programme comprised six sessions delivered over eight weeks. Each session included prebriefing, a simulated encounter, and facilitated debriefing. The stated durations represent planned timings, corresponding to 45–65 minutes per session and 270–390 minutes across six sessions, assuming participation in all phases. The intervention-group denominator (*n* = 40) does not establish attendance at every session. Intended learning outcomes describe educational objectives; they were not independently measured clinical-performance outcomes. The study measured self-reported anxiety, general self-efficacy, and resilience.

Session and scenario	Prebriefing: 10–15 min	Simulation: 20–30 min	Debriefing: 15–20 min	Intended learning outcomes
response	escalation principles, and safety-focused nursing responses.	behavior and rehearse therapeutic communication and de-escalation within the scenario.	awareness, emotional regulation, communication choices, and the rationale for crisis responses.	escalation and demonstrate appropriate de-escalation and safety-focused communication.
6. Suicide risk assessment and emotional support	Review suicide risk assessment procedures, ethical responsibilities, and supportive communication.	Practice suicide risk screening, recognition of warning signs, and therapeutic emotional support within the simulated encounter.	Discuss assessment reasoning, professional responsibilities, emotional responses, and communication effectiveness.	Identify relevant warning signs and demonstrate structured risk screening and supportive therapeutic communication.

Table notes: RMHS, robotic mental health simulation. The programme comprised six sessions delivered over eight weeks. Each session included prebriefing, a simulated encounter, and facilitated debriefing. The stated durations represent planned timings, corresponding to 45–65 minutes per session and 270–390 minutes across six sessions, assuming participation in all phases. The intervention-group denominator (n = 40) does not establish attendance at every session. Intended learning outcomes describe educational objectives; they were not independently measured clinical-performance outcomes. The study measured self-reported anxiety, general self-efficacy, and resilience.

Each session included three planned components:

1. **Prebriefing, 10–15 minutes**: introduction to learning objectives, participant roles, the simulated environment and expectations for respectful participation.
2. **Simulated encounter, 20–30 minutes**: rehearsal of assessment and communication responses relevant to the scenario.
3. **Debriefing, 15–20 minutes**: facilitated reflection on emotional responses, observations, clinical reasoning and alternative approaches.

The planned duration was therefore 45–65 minutes per session. This structure incorporated principles described in the Healthcare Simulation Standards of Best Practice for simulation design [14]. Learning objectives represented intended educational outcomes; they were not independently assessed clinical-performance endpoints.

Four faculty facilitators completed a two-day workshop covering robot operation, scenario delivery, prebriefing and debriefing. Standardized scenarios and facilitation guides were used to support consistency. Study documentation described independent observation of 10% of sessions using a structured checklist. However, numerical fidelity results, actual session durations and participant-level attendance records were unavailable for analysis. Delivered exposure and achieved fidelity could therefore not be quantified.

Comparison of the condition

Control participants received the routine psychiatric nursing curriculum, including lectures, demonstrations, case discussions, laboratory instruction, self-directed learning and supervised clinical practice. They did not receive the study robotic simulation program during the evaluation period and were offered access after data collection.

Equivalence of additional facilitator contact, active practice and debriefing time between conditions was not established. Accordingly, the comparison concerned the complete educational program rather than the independent contribution of the robotic platform.

Measures

Self-administered electronic questionnaires collected demographic and academic information and assessed anxiety symptoms, general self-efficacy and resilience. Background variables included age category, sex, marital status, recorded academic level, grade-point-average category, previous simulation exposure, personal or indirect experience of mental health issues, and familiarity with digital technologies.

The administered languages, exact translated versions and any instrument modifications were not established in the available documentation. Published psychometric findings therefore describe the source instruments and should not be interpreted as validation of an unidentified local adaptation.

Anxiety symptoms

The Beck Anxiety Inventory (BAI) comprises 21 items scored from 0 to 3, producing a total score of 0–63; higher scores indicate greater anxiety symptom severity [15]. The original development study reported Cronbach's alpha of 0.92 and one-week test–retest reliability of 0.75. In the combined baseline sample of the present study, Cronbach's alpha was 0.927.

The BAI was treated as a measure of general anxiety symptoms rather than a specific measure of anxiety associated with psychiatric placement.

General self-efficacy

The General Self-Efficacy Scale (GSE) comprises 10 items rated from 1 to 4, producing total scores of 10–40; higher scores reflect stronger perceived capability to manage challenging demands [16]. Multicultural research has examined the scale's psychometric properties across different populations [6]. In the present baseline sample, Cronbach's alpha was 0.792.

The GSE assessed broad efficacy beliefs. Scores were not interpreted as direct measures of psychiatric nursing competence or confidence in a specific clinical task.

Resilience

The 25-item Connor–Davidson Resilience Scale (CD-RISC-25) assesses self-reported resilience using items scored from 0 to 4. Total scores range from 0 to 100, with higher scores indicating greater resilience [7]. The original development study evaluated reliability, validity and sensitivity to change across community and clinical samples. In the present baseline sample, Cronbach's alpha was 0.940.

The 25-item instrument was used; psychometric findings for abbreviated versions were not substituted as evidence for this version.

Pilot procedures and assessment schedule

Eight students participated in pilot procedures examining questionnaire comprehensibility, electronic administration, simulation logistics and robot operation. Feedback informed procedural refinements, and pilot participants were excluded from the main study.

Assessments were administered before the program, immediately after its eight-week duration and three months after completion. No independently assessed clinical-performance or patient-care outcomes were collected. Figure 1 summarizes group sizes and the availability of assessment records.

Data verification and missing observations

Item-level records were examined for permissible values, missing responses and agreement between recorded totals and item sums. Observed numeric responses were within the specified scoring ranges. One immediate post-intervention CD-RISC-25 questionnaire in the intervention group lacked item 15. Its stored total represented the sum of 24 responses and was not accepted as a complete 25-item score. This questionnaire was excluded from resilience scoring for that assessment; no missing response was imputed or inferred.

Analyses retained available complete scores separately for each outcome and assessment. An incomplete score for one outcome did not automatically exclude valid scores for other outcomes or occasions. Immediate assessment denominators were 40 per group for anxiety and self-efficacy, and 40 control and 39 intervention participants for resilience. At three months, records were available for 36 control and 37 intervention participants.

Statistical analysis

The original analyses used IBM SPSS Statistics version 26. Item-level verification and the recalculated comparisons reported in this manuscript used Python with NumPy, pandas, openpyxl and SciPy.

Categorical variables were summarized as frequencies and percentages. Outcome distributions were described using means and standard deviations together with medians and first and third quartiles. Cronbach's alpha was calculated from complete baseline item responses for each instrument.

Exploratory between-group comparisons at each assessment used two-sided Mann–Whitney U tests with asymptotic inference, tie correction and continuity correction. Reported U statistics refer to the intervention

group. These tests assess differences in score distributions and were not interpreted specifically as tests of mean differences.

Holm adjustment was applied across the six post-intervention comparisons comprising three outcomes at two assessment times. A two-sided Holm-adjusted p value below 0.05 was considered statistically significant. Baseline comparisons were reported separately and were not used to establish group equivalence. Multiplicity adjustment did not constitute adjustment for baseline scores or confounding.

The analytical workbooks lacked explicit participant and clinical-section identifiers. Consequently, baseline-to-follow-up linkage could not be verified, and individual change analyses, baseline-adjusted estimates and analyses accounting for section-level dependence were not finalized. Spreadsheet row order was not accepted as proof of longitudinal identity. The reported results therefore represent exploratory, occasion-specific group comparisons rather than estimates of a causal intervention effect. Their inferential interpretation is limited by the unaddressed allocation structure.

Ethical considerations

Ethical approval

was obtained from the Local Committee for Research Ethics on Living Creatures at the University of Hafr Al-Batin, Saudi Arabia (approval number: HPO-05-FT-25/29). The study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments.

Participants received information about the study objectives, procedures and their rights before providing written informed consent. Participation was voluntary, and students were informed that declining participation or withdrawing would not affect their grades or academic progression. Electronic data were stored in password-protected files accessible to authorized research personnel. Facilitators provided supportive debriefing and guidance when students experienced emotional discomfort.

Results

Participant characteristics and availability of outcome data

The study included 80 undergraduate nursing students, with 40 participants in each group. Most participants were aged 20–22 years, female, single and in the grade-point-average category above 3.0. Academic level 5 was the most frequently recorded category. Moderate familiarity with digital technologies was reported by 26 participants (65.0%) in each group. Previous robotic psychiatric simulation exposure was recorded for four control participants (10.0%) and six intervention participants (15.0%), despite the original eligibility restriction. Baseline demographic and academic characteristics are presented in Table 2.

Table 2
Baseline demographic and academic characteristics

Characteristic	Control, n = 40	RMHS, n = 40
Age category, years		
< 20	5 (12.5%)	6 (15.0%)
20–22	31 (77.5%)	29 (72.5%)
> 22 to < 25	2 (5.0%)	3 (7.5%)
≥ 25	2 (5.0%)	2 (5.0%)
Sex		
Male	14 (35.0%)	16 (40.0%)
Female	26 (65.0%)	24 (60.0%)
Marital status		
Single	38 (95.0%)	37 (92.5%)
Married	2 (5.0%)	3 (7.5%)
Academic level, as coded		
1	1 (2.5%)	1 (2.5%)
3	3 (7.5%)	1 (2.5%)
4	4 (10.0%)	4 (10.0%)
5	23 (57.5%)	21 (52.5%)
6	6 (15.0%)	9 (22.5%)
7	2 (5.0%)	3 (7.5%)
8	1 (2.5%)	1 (2.5%)
GPA category		
< 2.0	2 (5.0%)	2 (5.0%)
2.0–3.0	7 (17.5%)	8 (20.0%)
> 3.0	31 (77.5%)	30 (75.0%)
Previous robotic psychiatric simulation, as labelled in the original table		
Yes	4 (10.0%)	6 (15.0%)
No	36 (90.0%)	34 (85.0%)

Characteristic	Control, n = 40	RMHS, n = 40
Age category, years		
Personal or indirect experience with mental health issues		
Personal experience	3 (7.5%)	5 (12.5%)
Through family or friends	8 (20.0%)	8 (20.0%)
No experience	29 (72.5%)	27 (67.5%)
Familiarity and comfort with digital technologies		
Low	3 (7.5%)	4 (10.0%)
Moderate	26 (65.0%)	26 (65.0%)
High	11 (27.5%)	10 (25.0%)

Complete baseline scores were available for all 80 participants across the three outcomes. Immediately after the intervention, anxiety and self-efficacy scores were complete for 40 participants in each group. One intervention-group resilience questionnaire contained an unanswered item and was excluded from complete-item scoring, leaving 39 intervention and 40 control resilience scores. Over the three months, outcome records were available for 36 control participants (90.0%) and 37 intervention participants (92.5%). These denominators represent available records at each assessment; individual longitudinal linkage remained unverified.

Anxiety symptoms

At baseline, mean anxiety scores were 8.85 (SD 8.79) in the control group and 11.78 (SD 11.08) in the intervention group; the between-group rank comparison yielded p = 0.250. Immediately after the program, mean scores were 8.60 (SD 8.46) and 4.63 (SD 4.00), respectively. Over the three months, corresponding means were 6.83 (SD 7.60) and 3.62 (SD 3.09).

Although observed anxiety scores were lower in the intervention group at both post-intervention assessments, neither comparison was statistically significant. The immediate comparison yielded U = 649.5 and an unadjusted p value of 0.145; the three-month comparison yielded U = 511.0 and an unadjusted p value of 0.086. Both Holm-adjusted p values were 0.172 (Table 3).

Table 3
Anxiety scores by group and assessment occasion

Assessment	Control n	Control mean ± SD	Control median [Q1, Q3]	RMHS n	RMHS mean ± SD	RMHS median [Q1, Q3]	U	Raw p	Holm p
Baseline	40	8.85 ± 8.79	7.00 [2.75, 11.75]	40	11.78 ± 11.08	8.00 [3.00, 21.00]	919.5	.250	—
Immediate	40	8.60 ± 8.46	6.00 [2.00, 14.00]	40	4.63 ± 4.00	3.50 [2.00, 4.00]	649.5	.145	.172
Three months	36	6.83 ± 7.60	4.50 [2.00, 8.00]	37	3.62 ± 3.09	3.00 [1.00, 6.00]	511.0	.086	.172

BAI range: 0–63; lower scores indicate fewer anxiety symptoms.

General self-efficacy

Baseline mean self-efficacy scores were 29.33 (SD 5.63) in the control group and 27.85 (SD 4.08) in the intervention group. The exploratory baseline comparison indicated a difference in score distributions (U = 597.5, p = 0.0495).

Immediately after the program, mean self-efficacy scores were 29.40 (SD 7.31) in the control group and 36.48 (SD 2.93) in the intervention group. At three months, corresponding means were 29.14 (SD 8.52) and 35.54 (SD 2.88). Between-group comparisons favored the intervention group immediately after the program (U = 1333.0, Holm-adjusted p < 0.001) and at three months (U = 993.0, Holm-adjusted p = 0.001). These comparisons were not adjusted for baseline self-efficacy or clinical-section clustering (Table 4).

Table 4
General self-efficacy scores by group and assessment occasion

Assessment	Control n	Control mean ± SD	Control median [Q1, Q3]	RMHS n	RMHS mean ± SD	RMHS median [Q1, Q3]	U	Raw p	Holm p
Baseline	40	29.33 ± 5.63	30.00 [27.75, 32.00]	40	27.85 ± 4.08	29.00 [27.00, 30.00]	597.5	.0495	—
Immediate	40	29.40 ± 7.31	31.50 [26.00, 34.25]	40	36.48 ± 2.93	37.00 [34.75, 39.00]	1333.0	< .001	< .001
Three months	36	29.14 ± 8.52	32.00 [24.75, 35.25]	37	35.54 ± 2.88	35.00 [33.00, 39.00]	993.0	< .001	.001

GSE range: 10–40; higher scores indicate greater general self-efficacy. Baseline p-values are descriptive exploratory comparisons, not tests establishing equivalence.

Resilience

Baseline mean resilience scores were 49.15 (SD 20.53) in the control group and 51.18 (SD 13.88) in the intervention group; the baseline rank comparison yielded p = 0.992. Immediately after the program, mean scores were 56.03 (SD 23.33) among 40 control participants and 76.69 (SD 11.87) among 39 intervention participants. At three months, corresponding means were 61.33 (SD 21.40) and 74.27 (SD 12.26).

Between-group comparisons favored the intervention group immediately after the program (U = 1228.0, Holm-adjusted p < 0.001) and at three months (U = 908.5, Holm-adjusted p = 0.022). These findings describe differences between the available group records at each assessment and do not establish individual changes in resilience (Table 5).

Table 5
Resilience scores by group and assessment occasion

Assessment	Control n	Control mean ± SD	Control median [Q1, Q3]	RMHS n	RMHS mean ± SD	RMHS median [Q1, Q3]	U	Raw p	Holm p
Baseline	40	49.15 ± 20.53	50.00 [23.00, 66.75]	40	51.18 ± 13.88	50.00 [44.00, 55.75]	798.5	.992	—
Immediate	40	56.03 ± 23.33	64.50 [23.00, 73.00]	**39**	**76.69 ± 11.87**	81.00 [70.00, 83.00]	1228.0	< .001	< .001
Three months	36	61.33 ± 21.40	71.00 [50.00, 79.00]	37	74.27 ± 12.26	75.00 [70.00, 84.00]	908.5	.007	.022

CD-RISC-25 range: 0–100; higher scores indicate greater resilience. One immediate RMHS questionnaire had a missing item and was excluded from this complete-item analysis.

Discussion

This study examined anxiety, general self-efficacy and resilience among undergraduate nursing students participating in a structured robotic mental health simulation program compared with those receiving standard preparation. Intervention participants had higher self-efficacy and resilience scores immediately after the program and at three months, with between-group differences remaining statistically significant after adjustment for multiple comparisons. Anxiety scores were numerically lower in the intervention group, but the corresponding comparisons were not statistically significant. These findings suggest a differentiated pattern across the psychological outcomes examined. However, nonrandom allocation, unadjusted comparisons and unverified participant linkage mean that the observed differences cannot be interpreted as established intervention effects.

The findings contribute preliminary evidence to a broader literature supporting simulation in psychiatric nursing education. Piot et al. identified benefits for selected skills and attitudinal outcomes across

psychiatric simulation studies, while emphasizing heterogeneity in interventions and research methods [4]. Reangsing et al. similarly reported improvements in nursing students' knowledge and performance following mental health simulation [5]. The present study extends attention to general self-efficacy and resilience alongside anxiety, with observations beyond the immediate training period. Nevertheless, these outcomes differ from knowledge, observed skills and attitudes. Agreement with the broader simulation literature therefore supports the rationale for further investigation rather than demonstrating equivalent benefits across technologies or outcome domains.

General self-efficacy and perceived capability

Higher post-intervention self-efficacy scores are theoretically consistent with the opportunities for rehearsal, observation and feedback incorporated into the program. Bandura's self-efficacy account identifies performance accomplishments, vicarious experiences, verbal persuasion and interpretation of emotional responses as sources of efficacy beliefs [8]. Within the present program, encountering a challenging scenario and discussing possible responses may have helped students identify strategies for managing unfamiliar demands. This offers a plausible interpretation of the findings, although mastery experiences, feedback quality and other proposed mechanisms were not measured.

Related qualitative evidence supports the educational relevance of these processes. In a study of standardized-patient simulation in a mental health nursing course, students described opportunities to practice therapeutic communication, receive immediate feedback and develop perceived readiness for clinical placement [17]. These accounts provide insight into how structured encounters may be experienced by learners, but they concern a different simulation modality and cannot establish the mechanism underlying the present findings.

Two considerations limit interpretation of the self-efficacy differences. First, intervention participants had lower baseline self-efficacy scores than controls. Their higher subsequent scores cannot be interpreted as greater individual improvement without verified longitudinal matching and baseline-adjusted analysis. Second, the General Self-Efficacy Scale measures broad beliefs about managing demanding situations rather than psychiatric nursing competence [6]. Students' perceptions of capability may be educationally relevant, but their relationship with observed communication, assessment and clinical reasoning requires direct evaluation.

Resilience and adaptation to challenging encounters

Resilience scores were also higher in the intervention group at both post-intervention assessments. Structured encounters and reflective discussion may provide opportunities to consider emotional responses, explore alternative approaches and identify ways of coping with difficult situations. Such processes are compatible with the adaptive capacity represented by the CD-RISC-25 [7]. However, the present findings do not establish that the program strengthened coping strategies or altered students' responses to adversity in clinical practice.

Mean resilience scores were descriptively higher after baseline in both groups. This pattern leaves open the contribution of routine coursework, clinical exposure, increasing familiarity with psychiatric care and other experiences occurring during the study period. Because individual trajectories cannot be verified, these changes in group means should not be treated as evidence that each participant improved. Furthermore, no threshold for educationally meaningful change was specified. Statistical differences in resilience scores therefore require consideration alongside their practical relevance, which remains uncertain in the absence of independently assessed behavior or functioning.

Interpreting the anxiety findings

The anxiety findings provide a necessary qualification to any broad claim of improved psychological preparedness. Although the intervention group had lower observed anxiety scores immediately after the program and at three months, neither between-group rank comparison was statistically significant. The findings do not establish equivalence between conditions or demonstrate that robotic simulation has no effect on anxiety. They indicate that the reported analyses did not provide sufficient evidence of a between-group difference.

This result differs from the overall favorable anxiety findings reported in the simulation meta-analysis by Silva et al. [3]. However, differences in intervention content, comparators, populations and measurement instruments limit direct comparison. Ko et al. also found reductions in clinical placement-related stress and anxiety across educational and coping interventions, while reporting substantial heterogeneity [18]. Such variation suggests that psychological outcomes depend on what an intervention addresses and how those outcomes are measured, rather than on the presence of simulation alone.

Measurement specificity is particularly relevant. The Beck Anxiety Inventory assesses general anxiety symptoms and was not developed as a measure of concerns specific to psychiatric placement [15]. It may therefore capture experiences extending beyond therapeutic encounters, such as broader academic or personal demands. Conversely, students could feel more capable of managing a clinical situation while continuing to experience anxiety. The observed pattern is compatible with this possibility, but the study did not test the relationships among changes in anxiety, self-efficacy and resilience. Future evaluations should distinguish general symptoms from placement-specific concerns and select measures that correspond closely to the intervention's intended targets.

The contribution of the robotic platform

The embodied robotic patient is a distinctive component of the program, but its independent contribution remains unresolved. Developmental research has demonstrated that humanoid platforms can present selected psychiatric cues through coordinated speech and movement [9]. In a small comparison of a humanoid robot with an actor presented through video, diagnostic outcomes showed little difference between groups, while the actor presentation was generally perceived as more realistic [10]. These findings indicate that technical feasibility, perceived realism and educational outcomes are related but separate considerations.

The present intervention combined robotic encounters with preparation, repeated practice, facilitator attention and debriefing. Because these components were delivered together, and the control condition was not matched for additional practice or contact time, observed differences cannot be attributed specifically to robotic embodiment. Nor do the findings establish superiority over standardized patients, role-play or virtual simulation. A more informative comparison would hold scenario content, exposure time and facilitation constant while varying the simulation modality.

The program should therefore be understood as an educational package containing a robotic component. This distinction matters for curriculum decisions: a resource-intensive platform requires evidence of added educational value relative to accessible alternatives. The present study provides a basis for further comparative investigation but does not establish scalability, cost-effectiveness or reduced demand for faculty supervision.

Implications for nursing education

The findings support further evaluation of robotic mental health simulation as a complementary component of psychiatric nursing education. Curriculum planning should begin with the encounters students need to manage and the learning processes required to support them. Clear objectives, preparation, appropriate challenge and constructive debriefing remain central considerations in simulation design [14]. Institutions considering robotic platforms should examine whether technology enables educational activities that are difficult to deliver effectively through existing approaches.

Evaluation should extend beyond self-reported psychological outcomes. Observed therapeutic communication, mental status assessment, clinical reasoning and responses to distress would help determine whether perceived capability corresponds with performance. Students' experiences of realism, emotional safety and technological discomfort should also be examined. These outcomes would provide a stronger basis for judging educational value than if greater confidence necessarily indicates better preparation.

Scenario development should preserve person-centered and recovery-oriented approaches to mental health care. Involving people with lived experience in reviewing scenarios could help identify stereotyped portrayals and strengthen the authenticity of communication demands. Robotic encounters should complement supervised engagement with patients, whose experiences and responses cannot be fully represented by programmed scenarios. Decisions about implementation should additionally consider accessibility, technical reliability, facilitator workload and costs, none of which were evaluated as outcomes in this study.

Strengths and limitations

The inclusion of a comparison group, three distinct psychological measures and a three-month assessment provide a broader evaluation than an immediate post-session satisfaction survey. Outcome-specific denominators and item-level verification also make the available evidence more transparent. Nevertheless, several limitations substantially constrain interpretation.

The single-institution convenience sample and nonrandom allocation introduce risks of selection bias and confounding. The number and distribution of original clinical sections were not established in the analytical

records, and the reported tests did not account for section-level dependence. If only one section received each condition, section and intervention effects would be inseparable. Adjustment for multiple comparisons does not address these design limitations.

Explicit participant identifiers were absent from the analytical workbooks, preventing verified baseline-to-follow-up matching. Consequently, the study reports occasion-specific comparisons rather than individual change or baseline-adjusted intervention effects. Higher scores at the three-month assessment do not establish maintenance of individual gains. Missing follow-up observations and the incomplete immediate resilience questionnaire may also introduce bias; similar retention percentages do not demonstrate that missingness was unrelated to outcomes.

The discrepancy between the eligibility criterion excluding previous robotic simulation and recorded prior exposure remains unresolved. Instrument languages, adaptation procedures and the interpretation of recorded academic levels also require clarification. Internal consistency estimates do not resolve these questions or establish cross-cultural validity. Moreover, the original sample-size calculation could not be independently reproduced, and attendance, delivered session duration and numerical fidelity results were unavailable.

Finally, all outcomes were self-reported and may have been influenced by participant expectations or socially desirable responding. Clinical competence, patient outcomes and educationally meaningful change thresholds were not assessed. Future multicenter evaluations should establish reliable participant linkage, document allocation and intervention delivery, use an analysis appropriate to the number of allocation units, and report comparative effect estimates with confidence intervals. Matched simulation comparators and longer observation would help clarify the added value of the robotic platform and the durability and practical relevance of any benefits.

Conclusions

In this single-institution quasi-experimental study, nursing students participating in robotic mental health simulation had higher self-efficacy and resilience scores than controls immediately after the program and at three months. Between-group anxiety comparisons were not statistically significant. These exploratory findings support further investigation of the program as an adjunct to psychiatric nursing education, but nonrandom allocation, unadjusted analyses and unverified longitudinal linkage prevent attribution of the observed differences to the intervention.

The educational value of robotic simulation should be established through comparative studies that distinguish the contribution of the platform from practice, facilitation and debriefing. Verified longitudinal analyses, appropriately matched comparators and independently assessed clinical performance are needed before recommending wider curricular adoption.

Abbreviations

Abbreviation	Full term
BAI	Beck Anxiety Inventory
CD-RISC-25	25-item Connor–Davidson Resilience Scale
GPA	Grade point average
GSE	General Self-Efficacy Scale
RMHS	Robotic mental health simulation
SD	Standard deviation
TIDieR	Template for Intervention Description and Replication
TREND	Transparent Reporting of Evaluations with Nonrandomized Designs

Declarations

Ethics approval and consent to participate

Ethical approval was obtained from the Local Committee for Research Ethics on Living Creatures at the University of Hafr Al-Batin, Saudi Arabia (approval number: HPO-05-FT-25/29). The study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments. All participants received information about the study objectives, procedures and their rights before providing written informed consent. Participation was voluntary, and students were informed that declining participation or withdrawing would not affect their grades or academic progression.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request, subject to applicable ethical requirements and protection of participant confidentiality.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number [0033-1446-S]

Authors’ contributions

FA, MK, SB, SG, MA, RA and EE contributed to the conception and design of the study. All authors contributed to data collection, analysis and interpretation, and to drafting and critically revising the manuscript. All authors read and approved the final manuscript and agreed to be accountable for all aspects of the work, including ensuring that questions concerning its accuracy or integrity are appropriately investigated and resolved.

Acknowledgements

The authors thank the participating nursing students for their time and contribution to the study. The authors also acknowledge the support of the Deanship of Graduate Studies and Scientific Research at the University of Hafr Al-Batin, Saudi Arabia.

References

1. Simpson MCG, Sawatzky JAV. Clinical placement anxiety in undergraduate nursing students: a concept analysis. Nurse Educ Today. 2020;87:104329. 10.1016/j.nedt.2019.104329.
2. Alhamidi SA, Alyousef SM. Undergraduate nursing student mental health clinical experiences in Saudi Arabia: findings from a qualitative study. J Am Psychiatr Nurses Assoc. 2023;29(2):136–45. 10.1177/10783903211000112.
3. Silva GO, Oliveira FSE, Coelho ASG, Cavalcante AMRZ, Vieira FVM, Fonseca LMM, et al. Effect of simulation on stress, anxiety, and self-confidence in nursing students: systematic review with meta-analysis and meta-regression. Int J Nurs Stud. 2022;133:104282. 10.1016/j.ijnurstu.2022.104282.
4. Piot MA, Dechartres A, Attoe C, Romeo M, Jollant F, Billon G, et al. Effectiveness of simulation in psychiatry for nursing students, nurses and nurse practitioners: a systematic review and meta-analysis. J Adv Nurs. 2022;78(2):332–47. 10.1111/jan.14986.
5. Reangsing C, Hayes CC, Maneekunwong K, Sopajaree C, Alexander L. The effects of mental health simulations on nursing students' knowledge and performance: a systematic review and meta-analysis. Nurse Educ Pract. 2026;92:104755. 10.1016/j.nepr.2026.104755.
6. Luszczynska A, Scholz U, Schwarzer R. The General Self-Efficacy Scale: multicultural validation studies. J Psychol. 2005;139(5):439–57. 10.3200/JRLP.139.5.439-457.
7. Connor KM, Davidson JRT. Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). Depress Anxiety. 2003;18(2):76–82. 10.1002/da.10113.
8. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev. 1977;84(2):191–215. 10.1037/0033-295X.84.2.191.
9. Schwarz P, Hellmers S, Spanknebel S, Hurlemann R, Hein A. Humanoid patient robot for diagnostic training in medical and psychiatric education. Front Robot AI. 2024;11:1424845. 10.3389/frobt.2024.1424845.
10. Schwarz P, Hellmers S, Spanknebel S, Immel D, Hurlemann R, Hein A. Comparing patient simulation with a humanoid robot or a human actor in terms of training success and acceptance: pilot questionnaire study. JMIR Form Res. 2025;9. 10.2196/70363.

11. Des Jarlais DC, Lyles C, Crepaz N, TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health. 2004;94(3):361–6. 10.2105/AJPH.94.3.361.
12. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ. 2014;348. 10.1136/bmj. g1687.
13. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods. 2007;39(2):175–91. 10.3758/BF03193146.
14. Watts PI, McDermott DS, Alinier G, Charnetski M, Ludlow J, Horsley E, et al. Healthcare Simulation Standards of Best Practice™ Simulation Design. Clin Simul Nurs. 2021;58:14–21. 10.1016/j.ecns.2021.08.009.
15. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol. 1988;56(6):893–7. 10.1037/0022-006X.56.6.893.
16. Schwarzer R, Jerusalem M. Generalized Self-Efficacy Scale. In: Weinman J, Wright S, Johnston M, editors. Measures in health psychology: a user's portfolio. Causal and control beliefs. Windsor: NFER-NELSON; 1995. pp. 35–7.
17. Fawaz M, Alreshidi SM, Thultheen I, Samaha A, Kavuran E. Undergraduate student nurses' experience of using standardized patient simulation in mental health course: a qualitative study. BMC Nurs. 2024;23:936. 10.1186/s12912-024-02601-9.
18. Ko KY, Montayre JR, Chiu PL, Lai TKH, Chan K, Chan HHT. Effects of interventions to reduce clinical placement-related psychological distress among nursing students: a systematic review with meta-analysis. Int J Nurs Stud Adv. 2025;9:100435. 10.1016/j.ijnsa.2025.100435.

Tables

Tables 1 to 5 are available in the Supplementary Files section.

Figures

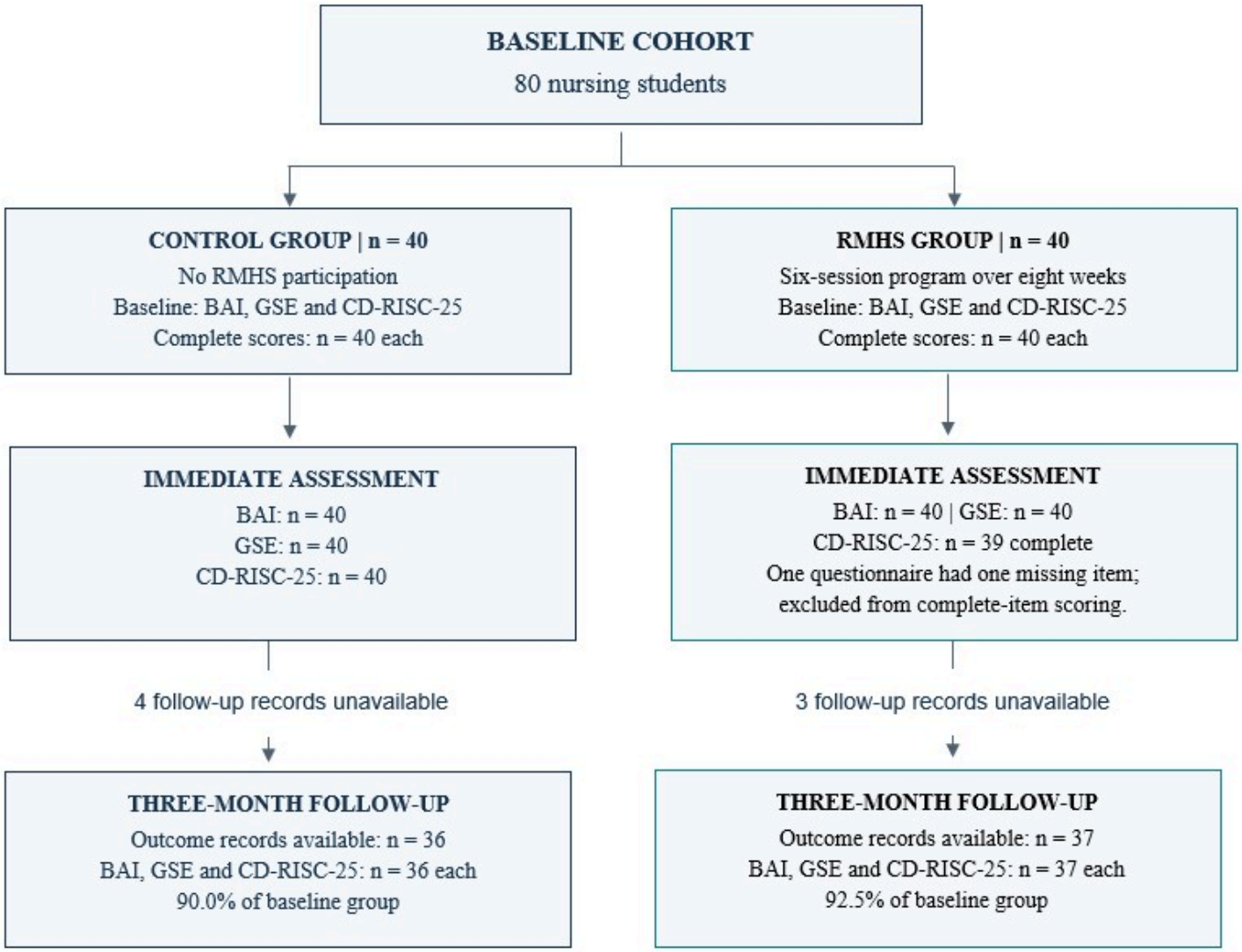

Figure 1

Study group sizes and availability of complete outcome scores at baseline, immediately after the intervention, and at three-month follow-up.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Tables.docx