Jiang et al. BMC Medical Education (2024) 24:1511
https://doi.org/10.1186/s12909-024-06530-8

BMC Medical Education

RESEARCH

Open Access

Enhancing medical education with chatbots: a randomized controlled trial on standardized patients for colorectal cancer

Yang Jiang[1†], Xinghua Fu[2†], Jing Wang[3], Qinling Liu[4], Xinyu Wang[5], Peijie Liu[6], Runchen Fu[7], Jiangpiao Shi[5*] and Yibo Wu[3]

Abstract

Introduction Combination of Standardized Patient (SP) and Case Based Learning (CBL) is a common method in medical education, but traditional SP (TSP) may not be conducive to students' mastery of basic medical knowledge and the cultivation of clinical thinking. Therefore, it is necessary to innovate SP to optimize SP combined with CBL teaching method.

Objective This study aims to explore the effectiveness of a chatbot utilizing standardized patients based on CBL (CSP-CBL) for colorectal cancer education.

Methods 61 medical students who have studied the theoretical knowledge of colorectal cancer were selected as the study objects and randomly divided into experimental group and control group. The experimental group used CSP-CBL, and the control group used traditional SP based on CBL (TSP-CBL). Before the intervention, basic knowledge test and clinical thinking ability assessment scale were used to investigate basic ability. After the intervention, we compared the effectiveness of two teaching methods in training colorectal cancer diagnosis and treatment skills through basic knowledge test, clinical thinking ability assessment scale, course experience questionnaire and client satisfaction questionnaire.

Result The majority of participants were female (62.3%, 38/61), 67.2% (41/61) were in the top 60% of school grades, and only 13.1% (8/61) had a medical background. There were no significant differences between the two groups in terms of demographic and sociological characteristics. There was no difference in pre-test of basic knowledge scores between the two groups ($P = 0.489$), but the CSP-CBL group scored significantly higher at post-intervention compared to the TSP-CBL group (SMD = -0.629, $P < 0.05$, 95% CI = (-0.789,-0.468)) and the CSP-CBL group also scored significantly higher than their baseline scores (SMD = − 0.991, $p < 0.05$, 95% CI = (-1.241,-0.740)). In terms of clinical thinking skills, the CSP-CBL group significantly improved their total score from (79.6 ± 15.9) to (86.2 ± 17.3) after the intervention (SMD = − 0.398, $p < 0.05$, 95% CI = (-0.498,-0.297)), but there was no significant difference between the two groups. The CSP-CBL group had a significantly lower academic load and course stress than the TSP-CBL group (SMD = -0.941,

[†]Yang Jiang and Xinghua Fu contributed equally to this work.

*Correspondence:
Jiangpiao Shi
1051114056@qq.com

Full list of author information is available at the end of the article

$p < 0.01$, 95% CI = (-1.181,-0.701); SMD = -0.6, $p < 0.05$, 95% CI = (-0.753,-0.447)) and scored significantly higher on the evaluation of future knowledge value (SMD = 0.603, $p < 0.05$, 95% CI = (0.449,0.757)). There was no significant difference between the two groups in terms of overall satisfaction, but the CSP-CBL group was significantly more satisfied than the TSP-CBL group in terms of meeting learning needs (SMD = 0.532, $P < 0.05$, 95% CI=(0.396,0.668)). There was no significant difference between the two groups in willingness to reuse the learning model.

Conclusion The results show that CSP-CBL teaching can significantly improve the basic knowledge and clinical thinking of clinical medical students learning colorectal cancer, better meet the learning needs of students and reduce the learning burden appropriately.

Trial registration ChiCTR2300072017 (registered on 31/05/2023).

Keywords Standardized patients, Teaching reform, Chatbot, Case-based learning, Colorectal cancer, Randomized controlled trial

Introduction

The landscape of medical education is continually evolving, with the concepts and methodologies underpinning it experiencing perpetual refinement. Crafting increasingly standardized and sophisticated teaching approaches has emerged as a critical priority within the realm of medical education in recent times. The paramount objective is to ensure that students of clinical medicine are furnished with a robust foundation in both medical knowledge and clinical competencies, while concurrently nurturing their capabilities for self-directed learning and clinical reasoning [1, 2].

Case-based learning (CBL) is a common way to teach in medical school. It gets students to think critically by looking at real patient cases and asking questions before class [3, 4]. The standardized patient (SP) concept, initially proposed by a neurosurgeon at the University of Southern California, is instrumental in training and assessing physicians' diagnostic abilities. SPs are capable of replicating the clinical presentation of actual patients, allowing undergraduate medical students to gain insights into clinical knowledge and skills without direct patient interaction, and to learn from errors within a controlled, safe environment [5, 6]. The utilization of SPs is a pivotal component of undergraduate medical education [7]. In China, the integration of SPs with CBL has been widely adopted in undergraduate clinical medical education. This combined teaching methodology demonstrates a notable enhancement over traditional approaches, effectively cultivating students' communication skills, self-directed learning capabilities, comprehension, and problem analysis proficiency [8–10].

Nevertheless, the traditional standardized patient (TSP) approach, as a mainstay of clinical medical education, has revealed its constraints within the contemporary educational landscape, particularly exacerbated during the COVID-19 pandemic. The widespread adoption of online learning has brought these limitations into sharp focus. TSP's reliance on a fixed teaching location restricts its ability to facilitate alternative communication methods and physical diagnosis training [11]. Moreover, the efficacy of TSP is compromised by the brevity of each session, typically lasting only about 4 h, which may insufficiently ensure the desired teaching outcomes. The temporal constraint is further compounded by the necessity for TSP instructors to undergo retraining with any shift in teaching content, limiting their capacity for rapid adaptation to the nuances of various diseases and thus compromising teaching quality [12]. Additionally, the recruitment, training standards, and professional experience of TSP instructors significantly influence the caliber of the course, leading to variability in students' comprehension levels and mastery of knowledge, ultimately affecting the overall pedagogical impact [13–15].

In this study, we define a 'chatbot' as an automated rule-based dialogue system that is linked to the WeChat platform through an API and interacts with users through a public service. Chatbots are commonly used in applications such as customer service and information access, where they can respond quickly to common questions and perform tasks without human intervention. In the field of medical education, chatbots can be used as interactive learning tools that simulate patient interactions and provide a platform for students to practice and improve their clinical skills. Such systems use natural language processing techniques to understand user input and generate responses based on a set of pre-defined dialogue paths and answer banks. Although it does not have a large language modelling engine, it is designed to simulate standardised patient interactions to support the learning and practice of clinical skills by medical students.

The integration of chatbot technology represents a novel and practical approach to enhancing the standardized patient (SP) teaching paradigm. This method provides an interactive online communication platform where users input information and receive text-based responses from chatbots, thereby creating an immersive environment for rapid and naturalistic information exchange [16]. In the realm of AI in medical education, the application of chatbots is particularly noteworthy. For instance, the work by Magdalena Gortz and colleagues exemplifies the use of an AI-driven chatbot focused on prostate cancer for doctor-patient interactions. This study highlights the advantages of consistent chatbot quality and the flexibility of training locations, which are crucial for the advancement of medical education [17].

Furthermore, research by Currie et al. has investigated the use of chatbots in medical imaging education, underscoring their potential integration into medical curricula. This work demonstrates how chatbots can be a transformative tool in the evolution of traditional SP (CSP) methodologies, offering a more interactive and personalized learning experience [18]. The growing body of literature on chatbots in medical education underscores the increasing importance of this technology in shaping the future of clinical training.

Undergraduate students in clinical medicine, who require early and extensive exposure to clinical training, are in urgent need of a more cost-effective and efficient learning tool. Such a tool, powered by chatbots, should provide students with access to diagnostic resources that enable repetitive training across a wide range of diseases. This ensures stable training quality and mitigates the potential for decision-making errors, which is a critical aspect of medical education. By highlighting the role of AI in this context, our study aims to better position itself within the current framework of medical educational research and practice.

In the present investigation, we engineered a chatbot to transition TSP approach to an enhanced, CSP model. Furthermore, we amalgamated this CSP model with CBL to create a more sophisticated and potent educational framework for clinical instruction. The study entailed a comparative analysis of the CSP-CBL hybrid method versus the conventional TSP-CBL approach, aiming to evaluate the efficacy of CSP-CBL in promoting clinical medical students' comprehension of knowledge and development of clinical reasoning skills. Additionally, we assessed student satisfaction with the respective teaching modalities to gauge their perceptions and preferences.

Method

Participant

This investigation constitutes a prospective, single-blind, single-center randomized controlled trial (trial registration number: ChiCTR2300072017) (registered on 31/05/2023). Ethical clearance for the study was granted by the Ethics Committee of the Fourth Central Hospital of Baoding (approval number: 2023032). A total of 64 fourth-year undergraduate students pursuing clinical medicine at the Baoding Fourth Central Hospital were recruited into the study. The reporting of this study adheres to the CONSORT (Consolidated Standards of Reporting Trials) statement guidelines.

Inclusion criteria:

1) age ≥ 18 years;
2) Skilled use of smart phones;
3) Have learned the theoretical knowledge of colorectal cancer.

Exclusion criteria:

1) Participating in or having completed standardized patient training for colorectal cancer cases
2) Did not fill in the informed consent form.

Sample size

This study is a randomized controlled trial where the intervention group is exposed to an artificial intelligence-based standardized patient (CSP-CBL) model, while the control group receives the conventional standardized patient (TSP-CBL) intervention. The primary outcome measure for this study is the performance of participants in a theoretical examination. Based on a literature review [19], the projected average score for the intervention group is 73.9 ± 7.0, and for the control group, it is 66.2 ± 6.5. Utilizing PASS 15 software for sample size calculation, it was determined that the intervention group (N1) should comprise 22 cases, and the control group (N2) should also include 22 cases. Accounting for potential attrition due to loss to follow-up and refusal, a 20% contingency was added. Consequently, the final sample includes at least 28 subjects in each group, totaling at least 56 participants to ensure statistical power and allow for robust analysis.

Study randomization

Following the obtaining of informed consent from the participants, a randomization procedure was conducted. For each participant, a random integer was generated and allocated according to a pre-determined randomization table. Participants were subsequently allocated to either the experimental group (CSP-CBL) or the control group (TSP-CBL) based on the parity of the final digit of their assigned random integer. Participants assigned an odd-numbered random integer were allocated to the CSP-CBL group, while those assigned an even-numbered random integer were allocated to the TSP-CBL group. This method ensured an equitable distribution of participants between the two study groups.

Instrument

Basic knowledge test

The basic knowledge assessment comprises two components: a pre-class and a post-class evaluation. Each component consists of 30 multiple-choice questions, with a maximum possible score of 30 points for each section. The knowledge points covered in the test were selected in alignment with the criteria of the Chinese Medical Practitioner Qualification Examination, encompassing areas such as the etiology, pathology, and staging, clinical presentation, diagnosis, treatment, and prognosis of colorectal cancer.

Once the test questions were drafted, they were scrutinized by a multidisciplinary team comprising experts in preventive medicine, basic medical sciences, and medical psychology, in collaboration with the Department of Gastroenterology. The final selection of pre-class and post-class test questions was made by two deputy chief physicians to ensure that both sets of questions were of equivalent difficulty, thereby maintaining the integrity and fairness of the assessment.

Course experience questionnaire-28

The Curriculum Experience Questionnaire (CEQ) was developed by Paul Ramsden [20]. In Australia, the questionnaire is widely used in higher education quality assessment. Lin Peng et al.translated and revised China 's CEQ-28 scale on the basis of the Australian version. The scale consists of four dimensions: good teaching, reasonable learning, classroom quality and classroom harvest. The split-half reliability of the questionnaire was 0.898, and the internal consistency reliability was 0.928, indicating that the questionnaire had high stability and reliability [21].

CEQ-28 has not been used in the field of SP teaching, and the teaching process involves the participation of non-teachers. Therefore, we further modified CEQ-28. First of all, it is adapted for specific disciplines. The subject of the project was uniformly modified from the original " teacher " to " this teaching method. " This modification ensures that students can understand the intention of the problem and provide a more accurate evaluation of the teaching method. Secondly, the four dimensions of the original scale are adopted. In addition, the project expression was modified according to the characteristics of clinical medical students. Through these projects, we can further understand the teaching effect in detail. Finally, we have further refined the language to ensure that the students can understand it. In this study, the Cronbach 's α of the questionnaire was 0.954. The Likert five-level scoring method was used in the questionnaire (1 point-very inconsistent, 2 points-comparatively inconsistent, 3 points-uncertain, 4 points-comparatively consistent, 5 points-very consistent); the higher the score, the better the quality of education.

Clinical thinking ability assessment scale

In order to assess the clinical thinking skills of medical students, an evaluation scale developed by Song Junyan was used in this study. The scale was constructed after exploring the determinants of clinical reasoning and was built around three key assessment dimensions, including 3 dimensions of critical thinking skills, systems thinking skills, and evidence-based thinking skills, with a total of 24 questions that collectively measure medical students' proficiency in clinical reasoning. These dimensions include 6 items related to critical thinking skills, 11 items related to systems thinking, and 7 items focused on evidence-based thinking. The scale is rated on a 5-point scale from 'very good' to 'very poor,' with higher scores indicating better clinical thinking skills (see Annex 1 for the original scale). In the original study conducted by Song Junyan, the scale showed a high degree of internal consistency, as indicated by a Cronbach's alpha coefficient of 0.969 [22].

Client satisfaction questionnaire-3

To evaluate student satisfaction with the CSP-CBL and TSP-CBL teaching methods, this study employed the CSQ-3 scale, which was developed by Roger, D et al. in 1993. The CSQ-3 scale is recognized as a concise, cost-effective, user-friendly, and sensitive instrument for assessing client satisfaction [15, 23]. The scale utilizes a Likert scale format with four response options, ranging from 1 point (indicating low satisfaction) to 4 points (indicating high satisfaction), with higher scores denoting greater levels of satisfaction. Within the context of this study, the questionnaire demonstrated a high degree of internal consistency, as evidenced by a Cronbach's α coefficient of 0.913.

Interventions

Development of CSP

The research team, composed of professionals in clinical medicine, nursing, rehabilitation therapy, and two associate senior instructors, meticulously reviewed and refined the case studies. This process involved a comprehensive examination of patients' background information, diagnostic procedures, symptoms, physical signs, and laboratory or imaging results, focusing on selecting cases with clinically representative features to optimize teaching outcomes. Based on the pathological attributes of the CSP cases, the research team formulated five key questions covering preliminary diagnosis, diagnostic rationale, laboratory examinations, differential diagnosis, and treatment protocols. Subsequently, the team developed a corresponding question bank and crafted detailed CSP responses. These responses were organized into dialogues and integrated into a CSP platform hosted on the WeChat subscription platform.

The technical specifications of the CSP platform are as follows:

- Platform Architecture: The frontend user interface is built using WeChat Mini Programs, with server-side processing handled by Node.js, and MongoDB is used for storing case data and dialogue logic in the backend.
- Natural Language Processing (NLP): The platform integrates a machine learning-based NLP engine to parse user queries and is capable of processing natural language input to provide accurate responses.
- Dialogue Management System: The platform uses a rule-based system combined with machine learning algorithms to maintain meaningful dialogues with users, simulating interactions with 'patients' even when information is incomplete.
- User Interaction: The platform supports text input and output and has an error handling mechanism that guides users to supplement missing information during dialogues.

The CSP platform is designed not only to understand users' queries and provide precise responses but also to sustain meaningful dialogue in the face of incomplete information, thereby simulating interactions with 'patients'. This innovative tool aims to enhance students' self-directed learning, clinical reasoning, doctor-patient communication, medical history collection, and diagnostic skills.

Teaching implementation

The experimental protocol of this study is delineated in Fig. 1. Prior to the commencement of the course, all participants were provided with basic information sheets and electronic questionnaires to evaluate their baseline knowledge and demographic/sociological attributes. Ahead of the lesson, the students were administered a basic knowledge test (lasting 15 min, with a maximum possible score of 30 points).

The detailed teaching plan was structured as follows:

Preliminary to the class, in alignment with the curriculum objectives, the instructor compiled a comprehensive dataset pertaining to colorectal cancer patients. This information was synthesized into a teaching case involving a '65-year-old male patient presenting with weight loss, constipation, and altered stool consistency'. Digital rectal examination revealed a rectal mass characterized by a firm texture and a nodular surface. Laboratory investigations indicated a hemoglobin level of 96 g/L (decreased), red blood cell count of 3.87×10^{12}/L, mean corpuscular volume (MCV) of 84 fl (decreased), with alanine aminotransferase (ALT), creatinine (Cr), blood urea nitrogen (BUN), and fecal occult blood test (FOBT) within normal limits. Colonoscopic evaluation revealed a protuberant lesion located 10 cm from the rectum to the anus, exhibiting irregular and ulcerated surrounding mucosa. Abdominal ultrasound imaging detected multiple hyperechoic masses within the liver. To facilitate an in-depth exploration of colorectal cancer, seven question sets were crafted for students to investigate, including inquiries such as: 'What are the current diagnostic considerations for this patient?'; 'What are the typical clinical manifestations of colorectal cancer?'; 'Which laboratory tests are commonly utilized in the diagnosis of colorectal cancer?'; 'How is colorectal cancer diagnosed and distinguished from other conditions?'; and 'What are the differences in clinical presentation between colon and rectal cancer?'. Students were then prompted to consider these questions as if they were the consulting physician and to formulate their lines of inquiry accordingly.

One week prior to the classroom discussion, the instructor disseminated the case studies and associated questions to both the CSP and TSP groups. The groups were tasked with independently researching textbooks, scholarly literature, and other relevant materials to arrive at answers and construct diagnostic and treatment plans in preparation for the classroom debate.

The experimental group was exposed to the CSP-CBL teaching methodology, with the following specific teaching plan implemented:

1) Teaching Phase (1 class hour): Students initiated the CSP application on their mobile devices. Guided by the instructors, they sequentially posed questions to the CSP, adhering to the structure of the chief complaint, current medical history, past medical history, personal history, marital history, and family history, as outlined in the question bank. Through interaction with the CSP, students gathered medical histories, formulated diagnostic hypotheses, and constructed treatment plans. Following the response to several queries, the CSP presented students with concealed multiple-choice questions pertaining to the diagnosis of colorectal cancer. Students entered their selected answers into an online test system integrated within the teaching software, facilitating convenient review of student responses by instructors through the backend interface.

2) Synthesis and Reporting Phase (1 class hour): Students engaged in group discussions to deliberate over the cases and associated challenges. The instructor then compiled a summary of the discussion outcomes and the collective performance, elucidated prevalent issues, dissected the cases, assessed the identification of latent risks, and directed students in isolating key issues from complex pathologies. Furthermore, the instructor guided students in ascertaining a robust theoretical foundation to underpin their analytical processes.

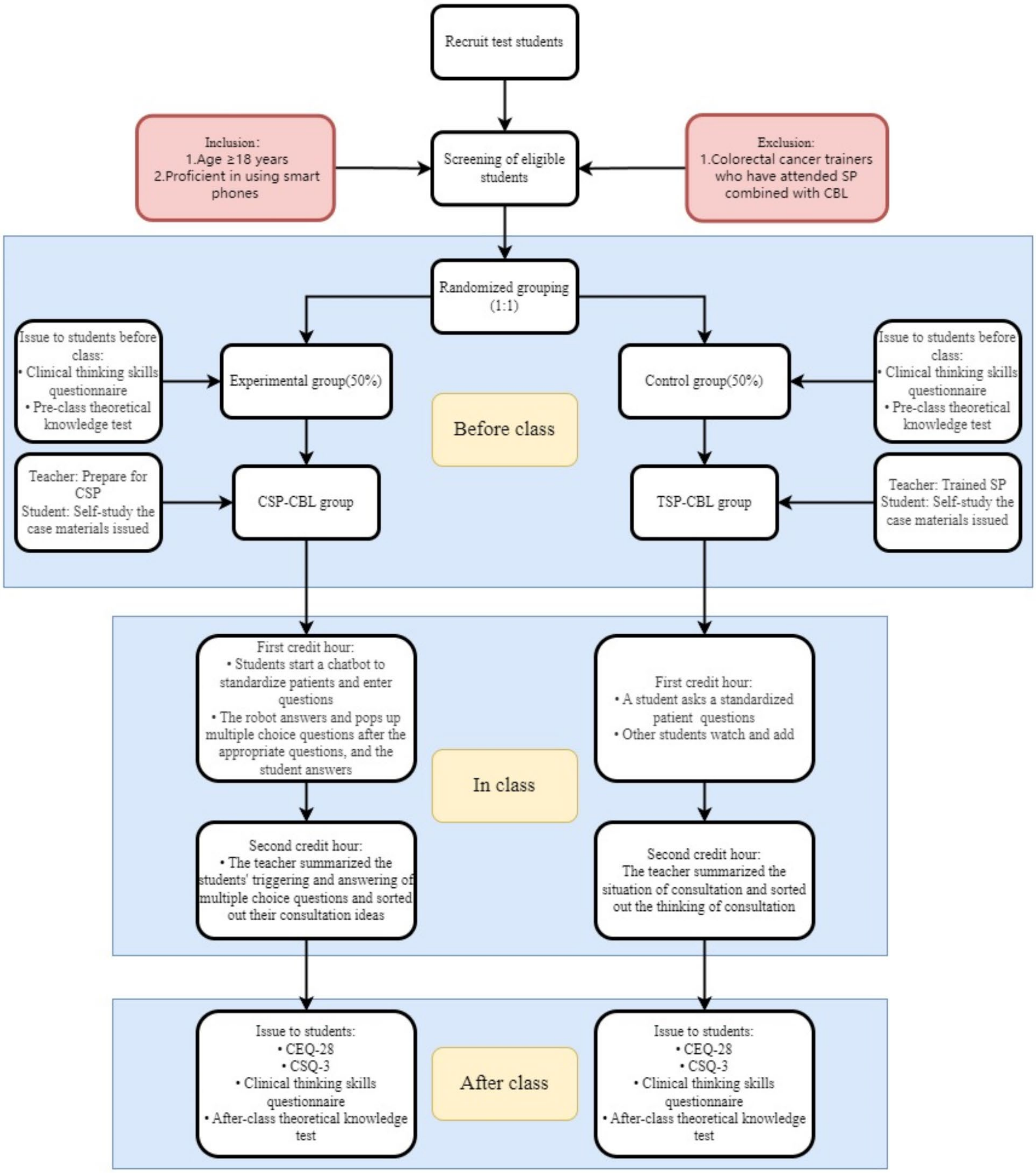

Fig. 1 Flow chart of the trial in this study

The control group was instructed using the TSP-CBL teaching method, with the following specific teaching plan:

1) Standardized Patient (SP) Preparation: In alignment with the guidelines set forth by Lisa Howley et al. for SP reporting [15], the script for the case, identical to that used in the experimental group, was provided to the SP instructor for pre-teaching preparation. Following the completion of this preparation, a deputy chief physician or a panel of physicians conducted simulated interrogations to assist the SP instructor in refining the content.

2) Teaching Phase (1 class hour): The SP instructor presented the case according to the simulated lecture, with one student posing questions while other students contributed to the inquiry and identified existing issues.

3) Synthesis and Reporting Phase (1 class hour): Students engaged in group discussions to deliberate and report on the cases and associated problems. The instructor then summarized the outcomes of the discussions and the overall performance of the inquiry, providing targeted explanations for the identified issues and highlighting the key focal points of the inquiry process.

The instructional duration for both methods was two hours. Upon completion of the course, a post-class basic knowledge test was administered, which consisted of 30 multiple-choice questions (lasting 15 min, with a maximum possible score of 30 points). Following the test, an assessment of course quality, clinical reasoning abilities, and student satisfaction was conducted for both groups. This evaluation aimed to determine whether the teaching innovation enhanced students' foundational and clinical competencies, as well as to gauge student acceptance of the curriculum.

Statistical analysis

All the collected data were cleaned and sorted after the study was completed, and then imported into SPSS (version 26.0, IBM) for statistical analysis. Continuous variables were reported in the form of mean±standard deviation (X±S), and categorical variables were expressed as number and percentage (frequency %). For the main outcome indicators, we used the independent sample t test to compare the statistical differences between the two groups, and the paired sample t test was used to compare the statistical differences within the two groups. For categorical variables, chi-square test ($\chi 2$ test) was used to compare differences between groups. All statistical tests were performed using a two-tailed test and the corresponding P values were reported. The significance level of $\alpha = 0.05$ was used for the hypothesis test of main outcome indicators. All P values were reported in the form of a two-tailed test. When the P value is less than 0.05, we believe that the difference is statistically significant. The Standard Mean Difference (SMD) effect size was employed to quantify the magnitude of difference between the two groups, with its 95% confidence interval computed. SMD ranged from −1 to 1, and a value approaching zero indicated no discernible distinction between the two groups. The disparities were categorized as small, medium, or large based on thresholds of 0.2, 0.5, and 0.8 respectively.

Result

Baseline characteristics of the students

From March 2023 to September 2023, a total of 64 participants were recruited in the study, 62 participants completed pre-class surveys and theoretical tests, and 62 were included in further trials. Among them, 31 people were assigned to the CSP-CBL group and 31 people were assigned to the TSP-CBL group, all of whom completed the course training. 1.6% (1 / 62) of the students did not complete the theoretical test and after-school survey, and finally 61 students completed all the research process. Most of the participants were female, accounting for 62.3% (38 / 61). 67.2% (41 / 61) of the participants ranked above 60% in school performance points, and only 13.1% (8 / 61) of the participants had medical background. There was no significant difference in demographic and sociological characteristics between the two groups (Table 1).

Comparison of the basic knowledge test scores between the two groups

No significant difference was observed in the pre-instruction basic knowledge test scores between the two groups (n.s). Following the CSP-CBL teaching intervention, the scores of the CSP-CBL group exhibited a statistically significant improvement compared to those of the TSP-CBL group (SMD=-0.629, $P<0.05$, 95% CI=(−0.789,–0.468)), and the CSP-CBL group's scores were also significantly higher than their baseline scores (SMD=−0.991,$P<0.05$ 95% CI=(−1.241,–0.740)) (Fig. 2).

Comparison of the clinical thinking ability between the two groups

Post-intervention, the CSP-CBL group achieved a total score of (86.2±17.3), a significant increase from their pre-intervention score of (79.6±15.9) (SMD=-0.398,$P<0.05$,95% CI= (−0.498, −0.297)) (Table 2). At baseline, there was no significant difference in the scores of clinical thinking ability between the two groups (n.s). Furthermore, after the intervention, no significant difference was observed in the scores of clinical thinking ability between the CSP-CBL and TSP-CBL groups (n.s) (Table 3). The CSP-CBL group's scores for systematic thinking ability and evidence-based thinking ability improved from (36.9±7.9) and (20.8±5.8) points pre-intervention to (40.6±7.6) and (24.7±5.3) points post-intervention, respectively, representing significant enhancements (SMD=-0.477, $P<0.05$, 95% CI=(−0.598, −0.356)), (SMD=−0.702, $P<0.01$, 95% CI=(−0.8796, −0.5244)).

Comparison of the course experience of the two groups

In the dimension of reasonable learning quantity, the learning burden and course pressure of CSP-CBL group

Table 1 Comparison of basic data between the two groups

	CSP-CBL group(N=31)	TSP-CBL group (N=30)	P value
Sex			n.s
Male (n,%)	11(35.5%)	12(40.0%)	
Female (n,%)	20(64.5%)	18(60.0%)	
Age (years) (X ± S)	19.55 ± 0.72	19.60 ± 0.67	n.s
Top percentage of school performance			n.s
≤ 40% (n,%)	11(35.5%)	13(43.3%)	
40< x ≤ 60 (n,%)	8(25.8%)	9(30.0%)	
60< x ≤ 100 (n,%)	12(38.7%)	8(26.7%)	
Whether the parents are engaged in clinical medicine related majors			n.s
Yes (n,%)	2(6.5%)	6(20.0%)	
No (n,%)	29(93.5%)	24(80.0%)	
Father's education background			n.s
primary education (n,%)	9(29.0%)	6(20.0%)	
secondary education (n,%)	14(45.2%)	19(63.3%)	
higher education (n,%)	8(25.8%)	5(16.7%)	
Mother's education background			n.s
primary education (n,%)	9(29.0%)	7(23.3%)	
secondary education (n,%)	18(58.1%)	16(53.4%)	
higher education (n,%)	4(12.9%)	7(23.3%)	

* $p<0.05$ ** $p<0.01$ n.s: not significant

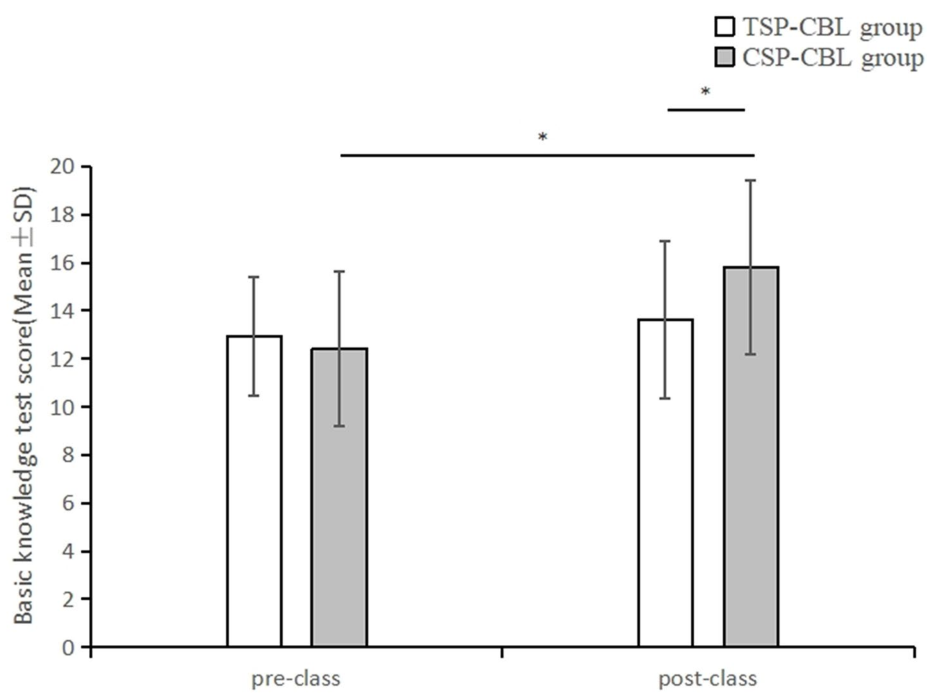

Fig. 2 Evaluation of the training effect. The basic knowledge test scores of the two groups of students before and after training (full score of 30) were shown and compared.* $p<0.05$ ** $p<0.01$,by a t-test

were significantly lower than those of TSP-CBL group (SMD=-0.941, P<0.01, 95% CI=(−1.181,–0.701)), (SMD=-0.6, P<0.05, 95% CI=(−0.753,–0.447)). In the course quality dimension, the CSP-CBL group scored significantly higher than the TSP-CBL group in terms of the future value of the knowledge learned (SMD=0.603, P<0.05, 95% CI=(0.449,0.757)) (Table 4).

Comparison of the satisfaction scores between the two groups

Concerning overall satisfaction, no significant disparity was found between the CSP-CBL and TSP-CBL groups (n.s). However, in terms of meeting students' learning needs, the CSP-CBL group exhibited significantly higher satisfaction than the TSP-CBL group (SMD=0.532,

Table 2 Comparison of the differences of clinical thinking ability between the CSP-CBL group and the TSP-CBL group before and after treatment between the two groups

	CSP-CBL group(N = 31)		P value	TSP-CBL group (N = 30)		P value
	pre-class	post-class		pre-class	post-class	
Critical thinking ability (point)	21.9 ± 3.5	20.9 ± 5.5	n.s	22.1 ± 4.5	20.1 ± 6.5	n.s
Systematic thinking ability(point)	36.9 ± 7.9	40.6 ± 7.6	*	37.9 ± 8.8	38.2 ± 9.2	n.s
Evidence-based thinking ability(point)	20.8 ± 5.8	24.7 ± 5.3	**	22.9 ± 6.8	23.4 ± 5.9	n.s
Total scores (point)	79.6 ± 15.9	86.2 ± 17.3	*	82.9 ± 19.1	81.7 ± 19.6	n.s

* $p<0.05$ ** $p<0.01$ n.s: not significant

Table 3 Comparison of the differences of clinical thinking ability between CSP-CBL group and TSP-CBL group before and after intervention

	pre-class		P value	post-class		P value
	CSP-CBL group(N = 31)	TSP-CBL group (N = 30)		CSP-CBL group(N = 31)	TSP-CBL group (N = 30)	
Critical thinking (point)	21.9 ± 3.5	22.1 ± 4.5	n.s	20.9 ± 5.5	20.1 ± 6.5	n.s
Systematic thinking ability(point)	36.9 ± 7.9	37.9 ± 8.8	n.s	40.6 ± 7.6	38.2 ± 9.2	n.s
Evidence-based thinking ability(point)	20.8 ± 5.8	22.9 ± 6.8	n.s	24.7 ± 5.3	23.4 ± 5.9	n.s
Total scores (point)	79.6 ± 15.9	82.9 ± 19.1	n.s	86.2 ± 17.3	81.7 ± 19.6	n.s

* $p<0.05$ ** $p<0.01$ n.s: not significant

Table 4 Evaluation of the course experience of the two groups of students

	CSP-CBL group(N = 31)	TSP-CBL group (N = 30)	P value
Good teaching (point)	40.3 ± 7.7	39.1 ± 8.1	n.s
Reasonable learning amount (point)	5.5 ± 1.8	6.9 ± 1.9	**
Heavy study load (point)	2.7 ± 0.8	3.5 ± 0.9	**
Great pressure on study (point)	2.7 ± 1.0	3.3 ± 1.0	*
Course quality (point)	20.6 ± 3.3	20.0 ± 3.6	n.s
The study of this course is very inspiring(point)	4.0 ± 0.7	3.9 ± 0.9	n.s
I find this course motivating me to learn(point)	3.9 ± 1.0	4.0 ± 0.7	n.s
This course has broadened my field of expertise(point)	4.2 ± 0.7	4.1 ± 0.6	n.s
Overall, I am satisfied with the quality of the course(point)	4.1 ± 0.8	3.9 ± 0.9	n.s
I think the knowledge I have learned will be valuable for my future(point)	4.3 ± 0.5	3.9 ± 0.8	*
classroom gains (point)	43.9 ± 7.1	43.1 ± 8.0	n.s

* $p<0.05$ ** $p<0.01$ n.s: not significant

$P<0.05$, 95% CI=(0.396,0.668)). When it came to the willingness of students to employ the respective learning mode again, no significant difference was observed between the CSP-CBL and TSP-CBL groups (n.s) (Table 5).

Discussion

This study conducted a comparative analysis of the impacts of CSP-CBL and TSP-CBL on the acquisition of basic knowledge and the development of clinical reasoning skills among clinical medical students focusing on colorectal cancer. The findings indicate that the CSP-CBL group significantly outperformed the TSP-CBL group in terms of mastery of foundational knowledge, and also demonstrated reduced learning stress and a lighter cognitive load. These results suggest that CSP-CBL teaching holds promise in enhancing student learning outcomes. Moreover, in the enhancement of clinical thinking abilities, the CSP-CBL group exhibited a marked improvement in the post-class scores for systematic thinking and evidence-based thinking, which may be attributed to the interactive and personalized learning experience facilitated by CSP-CBL teaching. The results of the study showed that students in the TSP-CBL group did not show significant improvement in clinical thinking. We hypothesise that this may be due to the fact that students in the TSP-CBL group may be less individually engaged, as interaction with the CSP requires a higher level of personal engagement, where students need to take the initiative to ask questions and obtain information, whereas the TSP-CBL is more reliant on group discussion, which may result in insufficient engagement of individual students.The CSP provides immediate feedback, which helps students to revise their clinical reasoning process in a timely manner, whereas the feedback from the TSP-CBL may not be immediate and personalised enough. In addition the assessment tool we used may not have adequately captured all aspects of clinical

Table 5 Comparison of the satisfaction scores between the two groups

	CSP-CBL group($N=31$)	TSP-CBL group ($N=30$)	*P* value
1.degree of satisfaction (point)			n.s
4	11(35.5%)	5(16.7%)	
3	18(58.1%)	19(63.3%)	
2	2(6.4%)	5(16.7%)	
1	0(0%)	1(3.3%)	
2. degree of meeting learning needs(point)			*
4	9(29.0%)	6(20.0%)	
3	19(61.3%)	13(43.3%)	
2	2(6.5%)	10(33.3%)	
1	1(3.2%)	1(3.3%)	
3.degree of the teaching model was selected again(point)			n.s
4	10(32.3%)	7(23.3%)	
3	18(58.1%)	19(63.3%)	
2	3(9.6%)	3(10.0%)	
1	0(0%)	1(3.3%)	

* $p<0.05$ ** $p<0.01$ n.s: not significant

thinking, resulting in the actual performance of students in the TSP-CBL group on clinical thinking not being adequately captured.

The current study differs from other AI-based teaching tools in its use of a dialogue-driven CSP platform that simulates patient interactions, providing a more immersive and personalized learning experience. The outcomes of this study, particularly in the domains of systematic thinking and evidence-based thinking ability, align with those reported by Marsal-Hasan Al Amrani et al. [24] in their investigation of critical thinking skills among nursing students. This similarity could be attributed to the inherent challenges associated with small sample sizes, limited statistical power, or insufficient exposure to the teaching methods. Despite the CSP-CBL group's superior performance in addressing students' learning needs, there was no discernible difference in satisfaction or the inclination to repeat the teaching mode when compared to the TSP-CBL group. This discrepancy may be due to the CSP-CBL approach's emphasis on comprehensive knowledge acquisition over student engagement, unlike other AI-based tools that often prioritize user interaction and engagement. However, similar teaching aids, such as clinical virtual simulation tools, have demonstrated enhanced student satisfaction in other studies. For instance, in the study conducted by Jose Miguel Padilla et al., students expressed higher satisfaction with clinical virtual simulation tools, underscoring the potential for further development and application of such technologies in education [25].

As a prevalent instructional modality in medical education, TSP holds significance across various domains, including diagnostics, emergency medicine, and internal medicine [26]. Frank Herbstright et al. have demonstrated that the incorporation of SP (Standardized Patient) can enhance the clinical performance of medical students in emergency management scenarios [27]. Michael K Turner's research underscores the effectiveness of SP in bolstering students' diagnostic abilities and reducing educational expenses [28]. Tulay Basak et al. have further validated the role of SP in boosting students' confidence during clinical practice [29]. TSP instruction plays a pivotal role in areas such as physical examination, basic medical sciences, and clinical medicine [30]. Nevertheless, the cultivation of clinical thinking skills in inquiry and other communicative contexts is profoundly contingent upon the expertise and instructional strategies employed by the SP instructors. The CSP-CBL model, in contrast, offers a standardized and repeatable approach that complements the expertise of SP instructors by providing a consistent learning experience across different students and scenarios.

Despite the advantages of TSP teaching in bolstering foundational knowledge, its limitations in terms of comprehensive learning coverage, the capacity for repetitive practice, and the stimulation of teaching interest hinder its ability to fully accommodate the learning requirements of clinical medical students. The CSP teaching model employed in this study addresses these limitations of TSP by offering repeatability, broad coverage, standardized uniformity, and a targeted exchange of knowledge content, which are characteristics that set it apart from other AI-based instructional tools. This integration forms a more efficient and comprehensive teaching approach, synergizing with CBL, compared to the traditional TSP method [31, 32]. The practical impact of integrating CSP-CBL into the medical curriculum includes the potential for a more scalable and accessible form of clinical education, which can be particularly beneficial in resource-limited or remote learning environments.

The CSP teaching model used in this study addresses these limitations of TSP by providing repeatability, broad coverage, standardised uniformity and targeted knowledge content exchange, which makes it unique compared to other AI-based teaching tools. For example, Yu, P et al. developed a virtual reality (VR)-based simulation system for training medical students in surgical skills, which provides an immersive learning experience through highly realistic surgical simulations [33]. On the other hand, Nagi, F et al. explored the application of a machine learning-based educational platform in personalised medical education, which is able to adjust the content and difficulty of teaching according to students' learning progress and abilities [34].

In contrast to these studies, the CSP-CBL model in this study emphasises the development of clinical reasoning and diagnostic skills through dialogue-driven interactions. This approach is unique in that it not only provides standardised case-based learning, but also allows students to practice and consolidate their clinical skills through interaction with simulated patients. This pedagogical approach is similar to the study by Holderried, F, who developed a Natural Language Processing (NLP)-based chatbot to assist medical students with history taking and diagnostic training [35]. However, the CSP model in this study further extends this concept by integrating a CBL approach that facilitates deeper learning and the development of clinical thinking skills.

The CSP-CBL teaching model demonstrates efficacy in enhancing the learning outcomes of clinical medical students' comprehension of colorectal cancer. Future research should endeavor to extend the application of this instructional approach to other clinical medical fields, thereby contributing to the accumulation of empirical evidence for the advancement and reform of medical education. It is advisable to extend the implementation of this teaching model across a broader spectrum of learning modules and clinical medicine disciplines, thereby facilitating students'comprehension of complex clinical scenarios and the development of their adaptive skills.

Limitations

The limitations of this study include: 1) it is conducted in a single clinical teaching center, and the results may be affected by specific environment and culture; 2) The small sample size limits the generalisability and generalisation of the results and the strength of interpretation; 3) The intervention period is short, which is not enough to evaluate the long-term changes of clinical thinking ability; 4) It is impossible to completely exclude the influence of students' after-class communication on the results, which may affect the internal effectiveness of the research. 5) Focusing on the teaching effect of colorectal cancer, the results may not be directly extended to the study of other diseases. Future research should consider these limitations and design more rigorous studies to verify the preliminary findings.

Conclusion

The outcomes of this study revealed that medical students who were exposed to CSP-CBL teaching experienced not only a marked enhancement in their foundational theoretical knowledge of colorectal cancer but also a significant advantage in their clinical reasoning skills when compared to those receiving traditional instruction. The CSP-CBL teaching method is well-suited to addressing students' individualized learning requirements and is effective in alleviating their academic load.

Abbreviations

SP	Standardized patient
CBL	Case Based Learning
TSP	Traditional Standardized patient
CSP	Chatbot utilizing standardized patient
CSP-CBL	Chatbot utilizing standardized patients based on CBL
TSP-CBL	Traditional Standardized patient based on Case Based Learning; NLP: Natural Language Processing

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12909-024-06530-8.

Supplementary Material 1

Acknowledgements

We are grateful to the professionals and students who provided enormous support in the recruitment and data collection of this study, and the participants who took the time to cooperate with the survey.

Author contributions

YJ, XHF, JW and JPS conceived and designed the study. XHF, YJ, JW, QLL, XYW, PJL, RCF, JPS and YBW reviewed and revised the manuscript. JPS and YBW provided theoretical and methodological guidance. XYW, JPS and XHF are responsible for the study management. YJ, XHF, QLL and JW designed the questionnaire. XHF, YJ, XYW and JPS are responsible for the study implementation. All authors provided revisions and approved the final manuscript.

Funding

This project is supported by the Baoding city Science and Technology Program project (2340ZZ003).

Data availability

The data sets collected and analyzed in this study can be provided by the corresponding correspondent authors according to reasonable requirements.

Declarations

Ethics approval and consent to participate

The study was approved by the Ethics Committee of the The Fourth Central Hospital of Baoding, Baoding, Hebei Province (2023032), and the registration of the study protocol with the China Clinical Trial Registry (ChiCTR2300072017) was completed. This study complied with the' Helsinki Declaration '.Written Informed consent was obtained from all study participants.

Consent for publication

Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Jitang College, North China University of Science and Technology, Tangshan, Hebei, China
[2]The Fourth Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang, China
[3]School of Public Health, Peking University, Beijing, China
[4]College of Nursing, North Sichuan Medical college, Nanchong, Sichuan, China
[5]The Fourth Central Hospital of Baoding, Baoding, Hebei, China
[6]Xi'an Jiaotong University, Xi'an, Shaanxi, China
[7]School of Pharmaceutical Sciences& Institute of Materia Medica, Shandong First Medical University, Shandong Academy of Medical Sciences, Jinan, Shandong, China

Received: 25 April 2024 / Accepted: 13 December 2024
Published online: 20 December 2024

References
1. Li H, Zhou X. Medical education reform in China. Postgrad Med J. 2018;94(1118):673–4. 10.1136. Epub 2018 Dec 15. PMID: 30554170. / postgradmedj-2018-136145.
2. Badrawi N, Hosny S, Ragab L, Ghaly M, Eldeek B, Tawdi AF, Makhlouf AM, Said ZNA, Mohsen L, Waly AH, El-Wazir Y. Radical reform of the undergraduate medical education program in a developing country: the Egyptian experience. BMC Med Educ. 2023;23(1):143. https://doi.org/10.1186/s12909-023-04098-3. PMID: 36869307; PMCID: PMC9983512.
3. Jamkar AV, Burdick W, Morahan P, Yemul VY, Sarmukadum, Singh G. Proposed model of case based learning for training undergraduate medical student in surgery. Indian J Surg. 2007;69(5):176–83.
4. Wei F, Sun Q, Qin Z, Zhuang H, Jiang G, Wu X. Application and practice of a step-by-step method combined with case-based learning in Chinese otoendoscopy education. BMC Med Educ. 2021;21(1):89.
5. Han SG, Kim YD, Kong TY, Cho J. Virtual reality-based neurological examination teaching tool(VRNET) versus standardized patient in teaching neurological examinations for the medical students: a randomized, single-blind study. BMC Med Educ. 2021;21(1):493. https://doi.org/10.1186/s12909-021-02920-4. PMID: 34526004; PMCID: PMC8444400.
6. Sakakushev BE, Marinov BI, Stefanova PP, Kostianev SS, Georgiou EK. Striving for Better Medical Education: the Simulation Approach. Folia Med (Plovdiv). 2017;59(2):123–131. https://doi.org/10.1515/folmed-2017-0039. PMID: 28704187.
7. Wilbur K, Elmubark A, Shabana S. Systematic Review of Standardized Patient Use in Continuing Medical Education. J Contin Educ Health Prof. 2018 Winter;38(1):3–10.
8. Li Shujuan N, Junyi. Chen Fengying, etc. Research on the application of CBL combined with SP teaching method in emergency clinical teaching [J]. Volume 15. Continuing Medical education in China; 2023. pp. 50–4. 10.
9. Li Xian W. Chaokui. Value analysis of standardized patient combined CBL teaching in oral and maxillofacial surgery [J]. Continuing Med Educ China 2023,15(18):113–7.
10. Zhang Yanmin L, Luchang W, Yujia. etc. The application of wechat combining CBL and SP teaching methods in pediatric practice teaching [J]. Higher medical education in China2023(05):66–67.
11. Jeong Y, Lee H, Han JW. Development and evaluation of virtual reality simulation education based on coronavirus disease 2019 scenario for nursing students: A pilot study. Nurs Open. 2022;9(2):1066–76. https://doi.org/10.1002/nop2.1145. Epub 2021 Dec 1. PMID: 34851042; PMCID: PMC8859080.
12. Downar J, McNaughton N, Abdelhalim T, Wong N, Lapointe-Shaw L, Seccareccia D, Miller K, Dev S, Ridley J, Lee C, Richardson L, McDonald-Blumer H, Knickle K. Standardized patient simulation versus didactic teaching alone for improving residents' communication skills when discussing goals of care and resuscitation: A randomized controlled trial. Palliat Med. 2017;31(2):130–9. Epub 2016 Jul 10. PMID: 27307057.
13. Liu Y, Qie D, Wang M, Li Y, Guo D, Chen X, Li L, Yu H, Wang J. Application of role reversal and standardized patient simulation (SPS) in the training of new nurses. BMC Med Educ. 2023;23(1):373. https://doi.org/10.1186/s12909-023-04294-1. PMID: 37226138; PMCID: PMC10207783.
14. Howley L, Szauter K, Perkowski L, Clifton M, McNaughton N, Association of Standardized Patient Educators (ASPE). Quality of standardised patient research reports in the medical education literature: review and recommendations. Med Educ. 2008;42(4):350–8. https://doi.org/10.1111/j.1365-2923.2007.02999.x. Epub 2008 Feb 20. PMID: 18298448.
15. Attkisson CC, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. Eval Program Plann. 1982;5(3):233-7. https://doi.org/10.1016/0149-7189(82)90074-x. PMID: 10259963.
16. Ghorashi N, Ismail A, Ghosh P, Sidawy A, Javan R. AI-Powered Chatbots in Medical Education: Potential Applications and Implications. Cureus. 2023;15(8):e43271. https://doi.org/10.7759/cureus.43271. PMID: 37692629; PMCID: PMC10492519.
17. Görtz M, Baumgärtner K, Schmid T, Muschko M, Woessner P, Gerlach A, Byczkowski M, Sültmann H, Duensing S, Hohenfellner M. An artificial intelligence-based chatbot for prostate cancer education: Design and patient evaluation study. Digit Health. 2023;9:20552076231173304.
18. Currie G, Singh C, Nelson T, Nabasenja C, Al-Hayek Y, Spuur K. ChatGPT in medical imaging higher education. Radiography (Lond). 2023;29(4):792–9.
19. Shen J, Yuan L, Ge R, Shao X, Jiang X. Improving medical student recruitment into neurosurgery through teaching reform. BMC Med Educ. 2022;22(1):656. https://doi.org/10.1186/s12909-022-03722-y.
20. Paul Ramsden. (1991) A performance indicator of teaching quality in higher education: The Course Experience Questionnaire, Studies in Higher Education, 16:2, 129–150, https://doi.org/10.1080/03075079112331382944
21. Lin Peng H, Wang, Zhi, Liu et al. Development of Classroom Teaching quality evaluation questionnaire based on course experience [J]. Journal of Tianjin Vocational and Technical Normal University,2012,22(03):74–78.
22. Li Z. A study on the evaluation of medical students' clinical thinking ability and its influencing factors.2020.Jilin University,MA thesis. https://doi.org/10.27162/d.cnki.gjlin.2020.003498
23. Roger D, Jarvis P, Najarian B. Detachment and coping: the construction and validation of a new scale for measuring coping strategies. Personal Individ Differ. 1993;15:619–26. https://doi.org/10.1016/0191-8869(93)90003-L.
24. Alamrani MH, Alammar KA, Alqahtani SS, Salem OA. Comparing the Effects of Simulation-Based and Traditional Teaching Methods on the Critical Thinking Abilities and Self-Confidence of Nursing Students. J Nurs Res. 2018;26(3):152–157. https://doi.org/10.1097/jnr.0000000000000231. PMID: 29016466.
25. Padilha JM, Machado PP, Ribeiro A, Ramos J, Costa P. Clinical Virtual Simulation in Nursing Education: Randomized Controlled Trial. J Med Internet Res. 2019;21(3):e11529. doi: 10.2196/11529. Erratum in: J Med Internet Res. 2019;21(6):e14155. PMID: 30882355; PMCID: PMC6447149.
26. Khan R, Plahouras J, Johnston BC, Scaffidi MA, Grover SC, Walsh CM. Virtual reality simulation training in endoscopy: a Cochrane review and meta-analysis. Endoscopy. 2019;51(7):653–64. https://doi.org/10.1055/a-0894-4400. Epub 2019 May 9. PMID: 31071757.
27. Herbstreit F, Merse S, Schnell R, Noack M, Dirkmann D, Besuch A, Peters J. Impact of standardized patients on the training of medical students to manage emergencies. Med (Baltim). 2017;96(5):e5933. https://doi.org/10.1097/MD.0000000000005933. PMID: 28151877; PMCID: PMC5293440.
28. Turner MK, Simon SR, Facemyer KC, Newhall LM, Veach TL. Web-based learning versus standardized patients for teaching clinical diagnosis: a randomized, controlled, crossover trial. Teach Learn Med. 2006 Summer;18(3):208 – 14. https://doi.org/10.1207/s15328015tlm1803_4. PMID: 16776607.
29. Basak T, Demirtas A, Iyigun E. The effect of simulation based education on patient teaching skills of nursing students: A randomized controlled study. J Prof Nurs 2019 Sep-Oct;35(5):417–24. https://doi.org/10.1016/j.profnurs.2019.02.004. Epub 2019 Feb 8. PMID: 31519347.
30. Nayak KR, Nayak V, Punja D, Badyal DK, Modi JN. Simulated patient videos to supplement integrated teaching in competency-based undergraduate medical curriculum. Adv Physiol Educ. 2023;47(2):296–306. https://doi.org/10.1152/advan.00167.2022. Epub 2023 Mar 23. PMID: 36951628.
31. Kononowicz AA, Woodham LA, Edelbring S, Stathakarou N, Davies D, Saxena N, Tudor Car L, Carlstedt-Duke J, Car J, Zary N. Virtual Patient Simulations in Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. J Med Internet Res. 2019;21(7):e14676. https://doi.org/10.2196/14676. PMID: 31267981; PMCID: PMC6632099.
32. Lo Sasso AA, Lane JL, Malloy RB. Using standardized patient outcome to measure the effect of teaching asthma-related patient education and information-giving skills to medical students. Teach Learn Med. 2005

Summer;17(3):228 – 32. https://doi.org/10.1207/s15328015tlm1703_5. PMID: 16042517.
33. Yu P, Pan J, Wang Z, Shen Y, Li J, Hao A, Wang H. Quantitative influence and performance analysis of virtual reality laparoscopic surgical training system. BMC Med Educ. 2022;22(1):92. https://doi.org/10.1186/s12909-022-03150-y.
34. Nagi F, Salih R, Alzubaidi M, Shah H, Alam T, Shah Z, Househ M. Applications of Artificial Intelligence (AI) in Medical Education: A Scoping Review. Stud Health Technol Inform. 2023;305:648–51. https://doi.org/10.3233/SHTI230581 .
35. Holderried F, Stegemann-Philipps C, Herrmann-Werner A, Festl-Wietek T, Holderried M, Eickhoff C, Mahling M. A Language Model-Powered Simulated Patient With Automated Feedback for History Taking: Prospective Study. JMIR Med Educ. 2024;10:e59213. https://doi.org/10.2196/59213.

Publisher's note