

Enhancing Medical Students' Feedback Literacy through AI-Supported Virtual Reality: A Randomized Controlled Trial

Tayebeh Sadegh¹

tina_26_195@yahoo.com

Shiraz University

Mehdi Mohammadi

Shiraz University

Jafar Jahani

Shiraz University

Ghasem Salimi

Shiraz University

Research Article

Keywords:

Posted Date: August 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9692563/v2>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

EDITORIAL NOTE:

06 August 2026: Figure 1 on version 1 has been removed at the authors' request due to privacy concerns.

Enhancing Medical Students' Feedback Literacy through AI-Supported Virtual Reality: A Randomized Controlled Trial

Authors:

Tayebeh Sadegh^{1*}, Mehdi Mohammadi², Jafar Jahani^{3*}, Ghasem Salimi⁴

Affiliations:

¹PhD Candidate, Shiraz University, Fars Province, Shiraz, Iran. tina_26_195@yahoo.com

²Professor of Curriculum Studies, Shiraz University, Shiraz, Iran. M48r52@gmail.com

³Associate Professor of Higher Education, Shiraz University, Shiraz, Iran. jjahani37@gmail.com

⁴Assistant Professor of Higher Education, Shiraz University, Shiraz, Iran. salimi@shirazu.ac.ir

*¹Corresponding author: [tina_26_195@yahoo.com](mailto:tina_26_195@yahoo.com)

*²Co-corresponding author: jjahani37@gmail.com

Abstract

Background:

This study aimed to investigate whether participation in a multicomponent AI-supported virtual reality (VR) feedback-literacy program—comprising a preparatory workshop, mandatory AI-generated reflective prompt during VR clinical scenarios, and, in one condition, peer dialogue—was associated with greater improvements in supervisor-observed feedback-literacy behaviors than regular clinical clerkship activities.

Methods:

A randomized controlled study was conducted among 68 fourth-year medical students to examine the effects of multicomponent AI-supported VR feedback literacy program compared to a control condition, during the transition to clinical clerkships. Students' feedback literacy was assessed by supervisors as an observable and developmental capability before and after the intervention. Quantitative data were analyzed to compare changes in pre- and post-test scores across the groups, and qualitative data from 36 student reflections were analyzed thematically to contextualize the quantitative findings.

Results:

A significant improvement from pre- to post-test was observed in the experimental groups of AI–VR (24.3 ± 3.7) and AI–VR+peer (27.0 ± 5.0) compared with the control group (3.8 ± 7.8 ; both $p < 0.001$). Reflective activities suggested that students in the AI–VR+peer condition more consistently described collaborative sense-making, normalization of uncertainty, and emotional support, whereas students in the AI–VR condition reported more varied patterns of engagement with AI.

Conclusions:

The findings indicate the potential of multicomponent AI-supported VR environments to enhance feedback literacy in medical education. Furthermore, although incorporating peer dialogue did not produce significantly greater behavioral improvements, the qualitative findings suggest that it may enrich students' engagement with feedback by supporting collaborative meaning-making and reflective learning.

Keywords: Feedback, Feedback literacy, developmental trajectory of feedback literacy, Artificial intelligence, Virtual reality

Introduction

Participating in clinical clerkships is widely recognized to have a major influence on learning in medical education (1, 2). It is increasingly acknowledged by students and supervisors that feedback is crucial within this workplace-based learning and plays an indispensable role in improving clinical performance and achievement (3-6). Despite this recognition and targeted educational interventions designed to improve the quantity and quality of feedback, studies consistently reveal varying impacts of feedback on attainment during clerkship rotations (7-11).

There is growing recognition that the primary problem lies not in the provision of feedback, but in students' limited engagement with the feedback they receive. Even when feedback is timely and high-quality, meaningful improvement is unlikely unless students are able to effectively receive, understand, and act upon it (1, 7, 12).

Moreover, research increasingly indicates that feedback engagement is closely associated with students' feedback literacy—described as the knowledge, skills, and attitudes students must develop to engage effectively with feedback to improve their work and learning strategies (11, 13). Without well-developed feedback literacy, students' feedback engagement (e.g., reception, perception, interpretation, and use of feedback) tends to remain superficial, impeding its impact on their learning and development (14, 15).

This perspective has encouraged educational approaches that scaffold students' feedback literacy prior to their engagement with feedback to ensure that students are better equipped to make sense of, evaluate, and act on the feedback they later receive (16, 17). However, conventional approaches to scaffolding feedback literacy face practical challenges, such as time constraints, variability in supervisor training, and limited opportunities for sustained dialogic practice during busy clinical rotations (18-21). Recent advances in Generative Artificial Intelligence (GenAI) offer further opportunities to address these persistent limitations in feedback engagement in medical education (22-24).

Studies have shown that AI-generated real-time feedback can be accurate and can influence behavior, supporting students in skill acquisition and transfer to clinical practice (25). Yet, evidence also indicates that the mere provision of AI-generated feedback does not guarantee students' meaningful engagement with feedback (26). This is consistent with a social constructivist perspective, which emphasizes that dialogue and co-construction among participants are essential for developing shared and individual interpretations of feedback (11, 27). Furthermore, the development of feedback literacy—particularly the acquisition of tacit knowledge—requires observation, imitation, participation, and dialogue, rather than unilateral feedback delivery (11). Researchers have therefore explored the use of virtual reality simulators as ethically sound educational solutions that enable repeated, deliberate practice while providing safe learning environments that minimize direct patient risk (28). Building on this, researchers have developed AI-supported VR environments where GenAI generates real-time feedback on students' performance (29). However, a key limitation is that such AI-VR systems are often designed as autonomous evaluators rather than as dialogic scaffolds that support students' reception, perception, interpretation, and use of feedback (30). This highlights the need for intentional integration of peer interaction within AI-VR environments to not only enhance engagement but also directly support the development of feedback literacy as a socially situated and dialogic competence in clinical education (31, 32).

Statement of the problem

Despite the growing interest in VR simulators among medical educators and the increasing use of generative AI tools by students as quick-reference and information-synthesis resources (33-35), empirical investigations that explicitly examine such interventions through a feedback literacy lens—particularly in medical education—remain limited. Most existing studies focus on AI interventions solely as technology-mediated feedback tools per se (36), lacking a robust structured pedagogical feedback context in which AI facilitated dialogic interactions and feedback literacy growth (22, 37). Moreover, to our knowledge, the use of AI-generated feedback integrated into VR systems as a reflective prompt generator in medical education remains underexplored. This gap in evidence underscores the necessity of investigating the combined effectiveness of AI, VR, and peer interaction in optimizing the development of feedback literacy. Such an evidence-based approach can provide critical insights for improving feedback literacy as an observable, developmental capability that emerges through repeated opportunities to seek, interpret, evaluate, and act on feedback within authentic clinical contexts (28).

Conceptual framework

The study design is grounded in a contemporary feedback literacy framework, which conceptualizes feedback literacy as a developmental capability that grows through repeated participation in structured feedback-rich environments (11, 38). This capability is observable and manifested through students' feedback-seeking behaviors, interpretation of feedback, emotional regulation, and enactment of feedback in clinical practice.

To address the identified gap, we developed an integrated multicomponent program in which VR provides authentic clinical scenarios within a safe learning environment that enable repeated opportunities to practice clinical reasoning (28), AI functions as a reflective prompt generator intended to encourage students to make their thinking explicit by articulating their reasoning, acknowledging uncertainty, and critically examining their own assumptions through non-evaluative reflective dialogue (24), and peer interaction facilitates shared sense-making of feedback by exposing students to alternative perspectives through discussion (31, 32). The AI component was therefore expected to strengthen learners' readiness to engage with feedback by making their reasoning explicit, whereas peer interaction was expected to refine that reasoning through collaborative dialogue and shared interpretation. Together, these components were intended to create a structured environment enabling repeated participation in feedback processes, thereby supporting the progressive development of feedback literacy through learner-generated reflection and collaborative dialogue rather than passive receipt of feedback (11, 31, 32).

Research questions

Does participation in the multicomponent AI-supported VR feedback-literacy program (with or without peer dialogue) lead to greater short-term improvement in supervisor-observed feedback-literacy behaviors than usual clerkship activities?

How do students' qualitative reflections on their feedback experiences contextualize the quantitative patterns in supervisor-rated feedback literacy?

Methodology

Study design, setting, and participants

This randomized controlled study evaluated whether participation in two multicomponent AI-supported VR feedback interventions (AI–VR and AI–VR+peer) was associated with greater improvements in undergraduate medical students' feedback literacy than usual clinical clerkship activities during the transition to clinical training. Ethical approval was obtained from the Ethics Committee of Fasa University of Medical Sciences.

Fourth-year medical students enrolled in the physiopathology phase at Fasa University of Medical Sciences were eligible to participate because they were preparing to enter their first clinical clerkship rotations. Previous research suggests that feedback literacy continues to develop during this transition, making it an appropriate stage for interventions designed to strengthen students' engagement with feedback (39). As the accessible population comprised a single cohort of fourth-year medical students, recruitment was based on the available cohort rather than on a predetermined target sample size. Accordingly, all 102 eligible students were invited to participate through the university's official student communication channels by the first author. Participation was voluntary, and electronic informed consent—describing the study purpose, data collection procedures, and participants' right to withdraw at any stage—was obtained before enrolment. Ultimately, Sixty-eight students consented to participate (29 males and 39 females; mean age = 22.3 years, SD = 1.4). At the time of recruitment, students had already been assigned to fixed clerkship blocks and departmental rotations according to the university's routine educational schedule. Consequently, the research team knew in advance which of the five core clinical departments—Internal Medicine, Surgery, Pediatrics, Obstetrics and Gynecology, and the combined Infectious Diseases–Neurology block—each participating student would attend during the clinical clerkship.

Clinical supervisors were recruited separately from the clinical departments involved in the study. The first author subsequently contacted the heads of the clinical departments by telephone and invited each department to nominate one clinical supervisor (interns, resident, or attending physician) to serve as the assessor for students rotating through that department. The nominated supervisors were then contacted individually, provided with detailed information about the study, and invited to participate. All five nominated supervisors agreed to participate and provided written informed consent before data collection commenced.

The study was conducted over a six-week intervention and evaluation period. Before entering their clinical clerkships, students completed a two-week ward-based history-taking course as part of the physiopathology curriculum. This course, which took place one month before the start of clinical clerkships, allowed students to

practice systematic patient history-taking under faculty supervision, thereby providing supervisors with sufficient opportunity to observe students' feedback-related behaviors prior to the intervention. Upon completion of this course, supervisors completed the baseline (pre-test) assessment of students' feedback literacy. To minimize inter-assessor variability, the same clinical supervisor who would subsequently supervise each student's clerkship also completed that student's pre-test and post-test assessments.

The intervention was delivered over two weeks, with two sessions per week (four sessions in total). During each session, students completed one VR scenario lasting approximately 30 minutes. This duration was determined through pilot testing ($n = 5$ medical students) and was judged sufficient for completing the diagnostic process—including history-taking, physical examination, differential diagnosis, and initial management—without imposing undue time pressure.

Following the intervention, students started their eight-week clinical clerkship rotations, which included history-taking, documentation of laboratory findings, participation in ward rounds, and operating room attendance. To evaluate intervention effects, the post-test was administered two weeks after the start of the clerkship. This assessment point was chosen because feedback literacy is conceptualized as a developmental capability that becomes observable through authentic workplace behaviors, rather than as an immediate outcome of the intervention.

Randomization and blinding

Following the pre-test assessment, the first author performed stratified randomization by ranking students according to their baseline feedback literacy scores and dividing them into three approximately equal tertiles (lower, middle, and upper). Within each tertile, a computer-generated random sequence created in Microsoft Excel was used to allocate students across the three study groups, ensuring that each group included participants with a similar range of baseline feedback literacy scores and resulting in the final allocation of AI-VR ($n = 22$), AI-VR+peer ($n = 24$), and control ($n = 22$) in an approximately 1:1:1 ratio. The slightly larger AI-VR+peer group was intentional to accommodate peer dyads, which required an even number of participants. Following randomization, the first author informed participants of their group allocation by telephone. The flow of participants through each stage of the trial is presented in Fig. 1.

Clinical supervisors remained blinded to students' group allocation throughout the study. They were informed only that the study examined feedback literacy during clinical clerkships and were unaware that some students had participated in the intervention. Students were instructed not to disclose their allocation or discuss the intervention with supervisors during the assessment period. However, the success of assessor blinding was not formally evaluated and therefore cannot be confirmed.

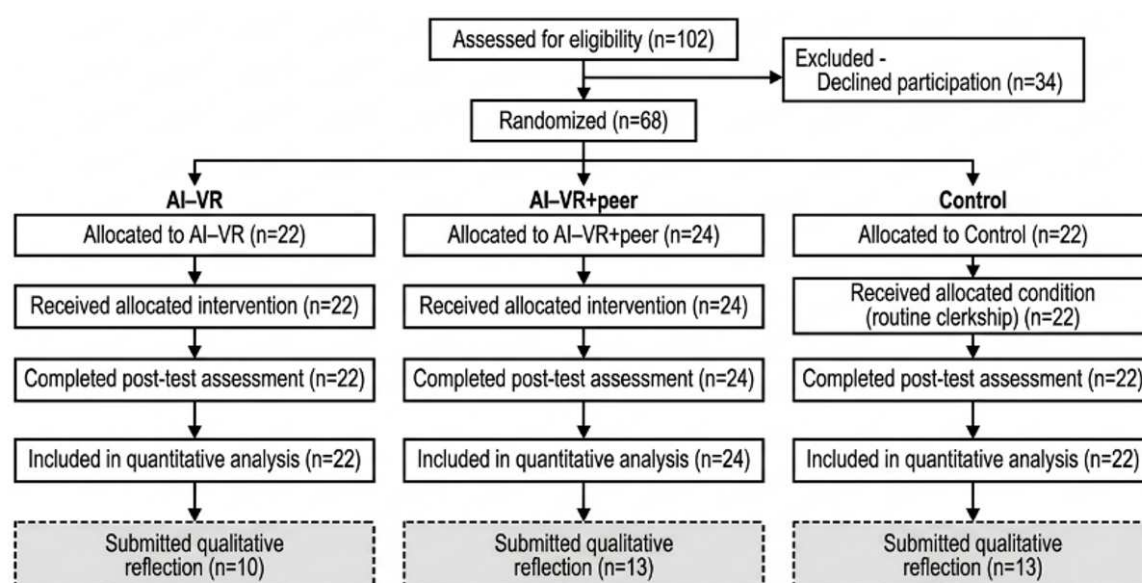


Fig. 1 CONSORT flow diagram of participant recruitment, randomization, intervention allocation, follow-up, and analysis

Preparatory phase

Before the intervention, the first author delivered two separate 1.5-hour preparatory workshops: one for students allocated to the intervention groups and another for the participating clinical supervisors.

Student session: The student workshop introduced the principles of feedback literacy, including the purposes and functions of feedback, proactive feedback-seeking strategies, common cognitive and emotional responses to feedback, and contextual influences such as time pressure, hierarchy, and clinical workload. Students also practiced engaging with multiple feedback sources through brief presentations, video demonstrations, and structured role-play activities. This preparatory session was considered an integral component of the intervention and aimed to reduce variability in prior exposure to feedback concepts and to support ethical and meaningful engagement with the intervention.

Supervisor session: The supervisor session focused on training the five participating clinical supervisors in the use of the Supervisor Observational Checklist for Feedback Literacy (SOC-FL), which was developed for this study (Appendix 1), to promote consistent scoring and reduce inter-rater variability. Supervisors participated only in outcome assessment and had no role in delivering the intervention or facilitating the student workshops.

Intervention Design and Study Conditions

Three study conditions were compared: AI–VR+peer, AI–VR, and a control condition receiving routine clinical clerkship activities. The intervention conditions shared the same preparatory workshop, VR clinical scenarios, and AI-generated reflective prompts but differed in whether learning occurred individually or with structured peer interaction (Table 1). Intervention fidelity was monitored throughout the study by recording attendance at the preparatory workshop and each VR session. In addition, students in both intervention groups were required to submit at least one screenshot of their peer and ChatGPT interactions during each VR scenario as evidence of engagement with the intervention. The first author reviewed these materials to verify students' engagement and did not influence the content or timing of interactions.

Four VR clinical scenarios—Stage II hypertension, Mitral Regurgitation with Congestive Heart Failure (Class II), Aortic stenosis, and Acute posterior wall myocardial infarction—were selected from the available VR scenarios in consultation with a cardiologist to represent a range of clinical complexity and acuity. The VR platform functioned solely as the clinical simulation environment and provided only procedural system notifications required for scenario progression (e.g., navigation and task-completion prompts).

Both intervention groups used ChatGPT-5.1 Auto (OpenAI) through the ChatGPT web interface (February 2026) with the default generation settings available at the time of the study. A fixed system prompt instructed the model to adopt a Socratic dialogue approach: *"In this group chat, medical students working in virtual reality will start a discussion with you regarding the scenarios they are facing. You should be aware that we expect you to use a structured Socratic dialogue method to encourage them to actively articulate their reasoning, examine their assumptions, and reflect on their decision-making process. Be aware, don't ever give them the correct answer immediately."*

The system prompt and clinical scenarios were pilot-tested with five medical students before the study to ensure that ChatGPT consistently adopted the intended Socratic approach. The system prompt remained unchanged throughout the study, whereas students' prompts were intentionally not standardized. Participants formulated their own questions according to their clinical reasoning, perceived uncertainties, and learning needs during each scenario.

Throughout the intervention, ChatGPT received only the text entered by students during the VR sessions and had no access to students' actions or performance within the VR environment. Accordingly, its responses were based solely on students' self-initiated dialogue rather than objective performance data from the simulation. Because the study evaluated students' feedback literacy outcomes rather than the performance of the AI system itself, individual AI conversations were not subjected to formal content analysis or quality evaluation.

In the AI–VR+peer condition, students worked in randomly assigned dyads. Both students had access to ChatGPT through a shared group chat interface using a researcher-provided ChatGPT account; however, only the student performing the scenario controlled the VR environment and initiated feedback-seeking during scenario completion, while the second student observed the performance and participated in the dialogue with both the performer and ChatGPT. Roles alternated across the four VR sessions. In Session 1, Student A performed while Student B observed; in Session 2, the roles were reversed. This alternation continued across all four sessions,

ensuring that each student completed two active performances and two observer experiences. Performer and observer experiences were intentionally analyzed as a single intervention condition rather than as separate roles for two reasons. First, the primary outcome measure (SOC-FL) assessed students' observable feedback engagement during authentic clinical clerkships rather than their performance within the VR scenarios. Second, active performance and peer observation were intentionally designed as complementary educational experiences within a single multicomponent intervention. Consequently, analyzing these roles separately would not have reflected the intended intervention design.

In the AI–VR condition, students completed the same four VR scenarios individually using the identical ChatGPT configuration, system prompt, and intervention procedures. As in the AI–VR+peer condition, students could initiate feedback-seeking at any point during the VR scenarios, and their prompts were not standardized (see Figure 2).

In the control condition, students did not participate in the preparatory workshop or the AI-supported VR intervention. The corresponding two-week period represented the routine interval between the physiopathology phase and the start of clinical clerkships, during which no additional educational activities were scheduled. All three groups then commenced their clinical clerkships simultaneously.

Fig. 2 Redacted due to privacy concerns. Students' interaction with the VR materials, ChatGPT and peer as part of the study activities

Table 1 Overview of the feedback settings and pedagogical differentiation of three study conditions

Dimension	AI–VR Group (n = 22)	AI–VR+Peer Group (n = 24)	Control Group (n = 22)
Experimental purpose	Evaluate the effect of AI-generated reflective prompts during VR	Examine the added value of peer interaction combined with AI-generated reflective prompts	Establish a baseline for comparison
Instructional setting	Individual engagement in immersive VR clinical scenarios	Collaborative engagement in immersive VR clinical scenarios	Routine clinical clerkship environment
Virtual Reality exposure	Students individually performed clinical tasks within standardized VR simulations	One student performed the task in VR while a peer observed the performance in real time	No VR exposure
AI support	Individual access to ChatGPT as an on-demand support tool	Shared access to ChatGPT via a group chat interface, visible to both members	No AI support
Learner configuration	Single learner	Group of two with clearly differentiated roles (performer and observer)	Participation within routine clinical practice
Feedback initiation	Learner-initiated (interaction required during each scenario)	Learner- and peer-initiated (interaction required during each scenario)	Externally provided as part of supervision
Feedback modality	AI-generated reflective prompts and questions (no performance evaluation or directive guidance)	Socially mediated, dialogic feedback supported by AI-generated reflective prompts and questions (no performance evaluation or directive guidance)	Informal clinical feedback
Pedagogical function of feedback	Support individual reflection and metacognitive monitoring.	Support shared reflection and collaborative sense-making	Support task completion and clinical exposure
Nature of reflection	Individual, internally regulated reflection	Joint reflection grounded in observation, dialogue, and shared interpretation	Largely implicit and unstructured

Data Collection

Two complementary data sources were used to evaluate the intervention: (1) supervisor ratings of students' observable feedback literacy behaviors and (2) students' self-reported reflections on their feedback engagement experiences.

Supervisor-rated feedback literacy

Observable feedback literacy was assessed using the SOC-FL, a 21-item instrument developed specifically for this study. The checklist was designed by operationalizing the feedback literacy framework of Carless and Boud (11) into observable behaviors suitable for assessment during clinical clerkships. An initial pool of items was drafted by the research team, refined through iterative discussion, and subsequently evaluated by an expert panel.

The instrument was administered before the intervention (pre-test) and two weeks after the start of the clinical clerkship (post-test). Items represented four domains: feedback-seeking behaviors (5 items), feedback interpretation (5 items), emotional regulation (5 items), and feedback enactment (6 items). Each item was rated on a 5-point Likert scale (1 = never observed to 5 = consistently observed), yielding total scores ranging from 21 to 105, with higher scores indicating greater observable feedback literacy.

Content validity was established through an expert panel of ten medical educators using Lawshe's Content Validity Ratio (CVR) (40). All 21 items exceeded the recommended threshold (CVR = 0.80–1.00). Item-level CVR values are presented in Table 2.

Table 2 Level of agreement among experts for each question

Question	Expert										Number of “Essential”	CVR	Result
	1	2	3	4	5	6	7	8	9	10			
1	E	E	E	E	E	E	E	E	E	E	10	1.00	Accepted
2	E	E	E	E	E	E	U	E	E	E	9	0.80	Accepted
3	E	E	E	E	E	E	E	E	E	E	10	1.00	Accepted
4	E	E	E	E	E	E	E	E	E	E	10	1.00	Accepted
5	E	E	E	E	E	E	E	E	E	E	10	1.00	Accepted
6	E	E	E	E	E	E	NN	E	E	E	9	0.80	Accepted
7	E	E	E	E	E	E	E	E	E	E	10	1.00	Accepted
8	E	E	E	E	E	E	E	E	E	E	10	1.00	Accepted
9	E	E	E	E	E	E	E	E	E	E	10	1.00	Accepted
10	E	E	E	E	E	E	E	E	E	E	10	1.00	Accepted
11	E	E	E	E	E	E	U	E	E	E	9	0.80	Accepted
12	E	E	E	E	E	E	E	E	E	E	10	1.00	Accepted
13	E	E	E	E	E	E	E	E	E	E	10	1.00	Accepted
14	E	E	E	E	E	E	E	U	E	E	9	0.80	Accepted
15	E	E	E	E	E	E	E	E	E	E	10	1.00	Accepted
16	E	E	E	E	E	E	E	E	E	E	10	1.00	Accepted
17	E	U	E	E	E	E	E	E	E	E	9	0.80	Accepted
18	E	E	E	E	E	E	E	E	NN	E	9	0.80	Accepted
19	E	E	E	E	E	E	E	E	E	E	10	1.00	Accepted
20	E	E	E	E	E	E	E	U	E	E	9	0.80	Accepted
21	E	E	E	E	E	E	U	E	E	E	9	0.80	Accepted

Note: E = Essential; U = Useful but not Essential; NN = Not Necessary.

Inter-rater reliability was evaluated by having two clinical supervisors independently assess a subset of 15 students (22% of the sample) during the pre-test assessment. The intraclass correlation coefficient (ICC) was 0.84 (95% CI: 0.71–0.92), indicating good agreement between raters. Internal consistency was good at both assessment points (Cronbach's α = 0.85 at pre-test and 0.87 at post-test). Preliminary construct validity is supported by the instrument's theoretical grounding in the feedback literacy framework and its ability to distinguish between study groups following the intervention. Evidence of responsiveness was demonstrated by the instrument's sensitivity to changes over time, detecting substantially greater improvements in the intervention groups than in the control group. Because the instrument was developed specifically for this study and the available sample was insufficient for stable factor analysis, evidence regarding its underlying factor structure has not yet been established and should be examined in future validation studies.

Students' reflective activities

To complement the quantitative assessment, two weeks after commencing clinical clerkship, all 68 students were invited to participate in self-reflection activities within one week of receiving the survey invitation. The reflective activity consisted of four open-ended questions designed to explore students' experiences of feedback engagement, including hesitation in seeking feedback, collaborative learning with peers, emotional responses to feedback, and the role of AI in supporting clinical reasoning.

Participation in the reflection activity was voluntary and was not linked to participation in the intervention or outcome assessment. Reflections were collected anonymously through an online survey. Participants were asked to indicate only their intervention group (AI–VR, AI–VR+peer, or Control) to enable group-level analysis; no names, student identification numbers, or other personally identifying information were collected.

Four open-ended reflection questions were as follows:

1. During the last two weeks, did you ever feel uncertain about a clinical decision but didn't seek clarification? What factors help you consider this situation as a normal, shared condition?
2. Do you remember a situation you discussed a case with peers which made you change your clinical judgement? How was this experience different from discussing it with supervisors?
3. How do you typically respond to critical feedback directed toward yourself or another student? Has observing or receiving critical feedback influenced your emotional response to feedback? If so, how?
4. During your clinical clerkship, did you use AI tools to support your clinical reasoning? If so, did you treat their responses as authoritative answers or as prompts for further reflection?

During the clinical clerkship—which formed part of the routine curriculum—students in all three study groups were free to use AI tools independently as part of their usual learning activities. This use was neither encouraged nor restricted, as the study sought to examine students' authentic engagement with AI in a real clinical learning environment. Accordingly, the reflection questions explored whether and how students incorporated AI into their clinical reasoning following the intervention.

Data Analysis

Descriptive statistics (mean and standard deviation) were calculated for each group at pre-test and post-test. Group baseline differences in feedback literacy were examined using One-way ANOVA. Paired-samples t-test was done to assess within-group changes over time. To compare post-test scores while adjusting for pre-test scores, ANCOVA was used. Data were analyzed using SPSS 24.

All student-written reflections underwent thematic analysis following Braun and Clarke's (41) framework. The analysis adopted a realist orientation, focusing on participants' explicit (semantic-level) accounts of their experiences with feedback engagement during the early weeks of clinical clerkship.

Two researchers (TS and MM), both experienced in medical education research and qualitative analysis, independently familiarized themselves with the dataset and generated initial codes. The initial coding framework was informed by existing feedback literacy literature (11, 24), while remaining open to concepts emerging from the data. Then the codes were collated into potential themes which underwent reviewing phase to generate a thematic map of the analysis. Discrepancies in coding resolved through discussion or by a third author (GhS). The final themes were used to contextualize, rather than explain causally, the quantitative findings by providing insight into possible reasons for the differences observed in supervisor-rated feedback literacy across study groups.

No participants withdrew after enrolment, and no outcome data were missing. Therefore, the intention-to-treat and per-protocol populations were identical, and all analyses were conducted using complete-case data.

Results

Baseline characteristics and pre-test feedback literacy scores

Baseline supervisor-rated feedback literacy was assessed before the intervention. The students' feedback literacy mean score in the AI–VR+peer experimental group was 46.75 (SD = 12.73), the AI–VR experimental group scored 47.68 (SD = 7.99), and the control group had a mean score of 45.95 (SD = 15.66). One-way ANOVA showed no statistically significant difference across the three groups, $F(2, 65) = 0.105$, $p = .901$ (Table 3).

Table 3. Descriptive statistics and within-group comparisons of feedback literacy scores

Group	n	Pre-test M (SD)	Post-test M (SD)	Mean difference (95% CI)	t	df	p
Control	22	45.95 (15.66)	49.73 (8.65)	3.77 (0.33, 7.22)	2.28	21	.033
AI-VR	22	47.68 (7.99)	72.00 (9.73)	24.32 (22.69, 25.95)	31.08	21	<.001
AI-VR+peer	24	46.75 (12.73)	73.71 (9.28)	26.96 (24.84, 29.08)	26.32	23	<.001

Within-group changes in feedback literacy

Paired-samples *t* tests demonstrated significant improvements in feedback literacy over time in all three groups (Table 3). Improvements were modest in the control group (mean difference = 3.77, 95% CI 0.33–7.22, $p = .033$) but substantially larger in the AI-VR (24.32, 95% CI 22.69–25.95) and AI-VR+peer groups (26.96, 95% CI 24.84–29.08), both $p < .001$.

Between-group differences in post-test feedback literacy

Assumptions of normality and homogeneity were met (Shapiro–Wilk $p = 0.057$ – 0.637 ; Levene’s test $p = 0.593$). Homogeneity of regression slopes was also supported, as the interaction between study group and pre-test score was not statistically significant, $F(2, 62) = 2.00$, $p = .176$. Between-group differences in post-test feedback literacy were then examined using ANCOVA controlling for baseline (pre-test) scores (Table 4). After adjustment for baseline feedback literacy, study group had a statistically significant effect on post-test feedback literacy, $F(2, 64) = 240.27$, $p < .001$, partial $\eta^2 = .882$.

Table 4. ANCOVA results for post-test feedback literacy scores controlling for baseline feedback literacy

Source	Sum of Squares	df	Mean Square	F	p	Partial η^2
Pre-test (covariate)	4546.09	1	4546.09	292.35	< .001	0.820
Group	7472.49	2	3736.25	240.27	< .001	0.882
Error	995.23	64	15.55			
Total	304381.00	68				

*Note: $R^2 = 0.927$ (Adjusted $R^2 = 0.923$).

Adjusted post-test means and Bonferroni-adjusted pairwise comparisons are presented in Table 5 and 6. After controlling for baseline scores, adjusted mean scores were 50.29 (SE = 0.84) for the control group, 71.41 (SE = 0.84) for the AI-VR group, and 73.74 (SE = 0.81) for the AI-VR+peer group.

Bonferroni-adjusted pairwise comparisons demonstrated that both intervention groups achieved significantly higher adjusted post-test scores than the control group. The adjusted mean difference between the AI-VR+peer and control groups was 23.45 points (95% CI 20.59–26.31, $p < .001$). Likewise, the AI-VR group scored 21.12 points higher than the control group (95% CI 18.19–24.05, $p < .001$). The adjusted mean difference between the two intervention groups was 2.33 points and was not statistically significant (95% CI -0.53 to 5.19 , $p = .149$).

Table 5. Adjusted post-test means estimated from the ANCOVA model

Group	Adjusted Mean	SE	95% CI
Control	50.29	0.84	48.61–51.97
AI-VR	71.41	0.84	69.73–73.09
AI-VR+peer	73.74	0.81	72.13–75.35

Table 6. Bonferroni-adjusted pairwise comparisons of adjusted post-test means

Comparison	Mean Difference	SE	95% CI	p	Cohen's d
AI-VR+peer vs. Control	23.45	1.16	20.59–26.31	<.001	2.63
AI-VR vs. Control	21.12	1.19	18.19–24.05	<.001	2.41
AI-VR+peer vs. AI-VR	2.33	1.17	-0.53 – 5.19	.149	0.19

Within-dyad dependence

Because participants in the AI–VR+peer condition worked in dyads (12 dyads), a sensitivity analysis was conducted to examine potential within-dyad dependence. Using a two-way mixed-effects intraclass correlation coefficient (ICC) with a consistency definition, the single-measures ICC was 0.295 (95% CI –0.308 to 0.729, $p = .164$). Although the point estimate suggested modest similarity between dyad members, the wide confidence interval and non-significant result indicated substantial uncertainty regarding the magnitude of within-dyad dependence, reflecting the limited number of dyads. Consequently, multilevel modelling was unlikely to yield reliable parameter estimates and was therefore not undertaken.

Qualitative Results

Of the 68 participants, 36 submitted post-intervention reflections (response rate: 52.9%), including 13 students from the AI–VR+peer group, 10 from the AI–VR group, and 13 from the control group. Reflection submission was voluntary, and several students informally cited time constraints during the early clerkship period as the primary reason for non-submission.

Thematic analysis identified four themes that helped explain the quantitative findings by illustrating how students experienced feedback engagement during the transition to clinical clerkships: (1) normalization of uncertainty, (2) social calibration of judgment, (3) transformation of AI from authority to instrument, and (4) vicarious desensitization to affective threat.

Normalization of uncertainty: Students in the AI–VR+peer group commonly described uncertainty as becoming a shared and acceptable aspect of clinical learning rather than an indication of personal inadequacy. One participant explained: *"In VR, my partner said 'I don't know' openly, and that gave me permission. Now I tell myself: uncertainty means I'm thinking, not failing."* Another reflected: *"Seeing my peer hesitate too made me realize I'm not alone. I'm getting better at saying 'I'm not sure' to my peers first."* Students in the AI–VR group also described becoming more comfortable with uncertainty, although these accounts emphasized individual reflection rather than social interaction: *"The AI kept asking 'What else could this be?' That taught me to sit with uncertainty rather than panic. Now in the ward, I see it as a thinking process, not a personal failure."* In contrast, many control participants continued to describe uncertainty as evidence of insufficient competence: *"I interpreted uncertainty as a personal knowledge gap—proof that I don't belong here."* (Control)

Not all students experienced uncertainty in the same way. One participant in the AI–VR+peer group reported that shared uncertainty increased anxiety rather than reducing it, whereas one control participant described gradually becoming more comfortable with clinical ambiguity through routine ward experience alone. Together, these accounts suggest that students in the intervention groups, particularly those in the AI–VR+peer condition, were more likely to describe uncertainty as a constructive part of clinical reasoning than were students in the control group.

Social calibration of judgment: Students in the AI–VR+peer group frequently described peer dialogue as helping them reconsider clinical decisions and refine their reasoning through collaborative discussion. One participant commented: *"A peer asked, 'Did you notice she had her appendix removed five years ago?' I had missed that. I now systematically check for prior surgeries before locking in a diagnosis."* Another explained: *"After a peer and I disagreed on antibiotics, I realized I was treating a lab number. I calibrated my risk perception—I now ask, 'What's the harm of waiting?'"* Students in the AI–VR group more often described refining their reasoning through individual reflection supported by AI: *"The AI prompted me to justify my reasoning. When I found a discrepancy, I asked ChatGPT why."* Several AI–VR students also reported transferring this questioning approach to interactions with peers during clerkship: *"When I was on the ward, I found myself asking my peers to explain why they preferred one approach over another instead of just accepting the correction."* Control participants generally described fewer opportunities for collaborative calibration: *"I can't remember a time I genuinely changed my judgment after a peer discussion. I just wanted them to stop questioning me."* Nevertheless, one control participant described an isolated experience in which peer discussion substantially changed subsequent clinical reasoning, illustrating that productive calibration also occurred outside the intervention. Overall, peer dialogue appeared to provide more consistent opportunities for collaborative refinement of clinical judgment than either AI-supported reflection alone or routine clerkship activities.

Transformation of AI from authority to instrument: Students described markedly different ways of engaging with ChatGPT across the intervention groups. Participants in the AI–VR+peer group most commonly portrayed AI as a resource for collaborative inquiry rather than a source of definitive answers: *"I used ChatGPT as a tool to interrogate, not authority. I asked for references and cross-checked with UpToDate. I interrogate, not obey."* Another noted: *"I took ChatGPT's list and asked my resident, 'Does this seem complete?' That taught me to use AI for options, but humans for context."* Experiences in the AI–VR group were more variable. Some students adopted similarly reflective approaches: *"I asked ChatGPT to explain its reasoning, then compared it to UpToDate. I used it like a senior resident who might be wrong—I listen, then I verify."* Others, however, described relying on AI primarily for confirmation: *"I treated its output as authoritative and stopped thinking. I used AI as a shortcut, not an instrument."* Most control participants also reported using ChatGPT during clerkship, although primarily to confirm existing judgments: *"I used ChatGPT to check my own diagnosis. It was enough to tell me I was right."* Only two control participants described using AI as a reflective guide rather than a verification tool. These findings indicate that students differed not only in whether they used AI, but also in how they engaged with it.

Vicarious desensitization to affective threat: Students in the AI–VR+peer group frequently described observing peers receive feedback as reducing the emotional threat associated with criticism. One participant reflected: *"I watched my peer nod and say 'Thank you, I will.' That moment desensitized me. Now when I receive direct criticism, I don't feel attacked."* Another commented: *"Watching my peer receive feedback without defensiveness showed me that was possible."* Students in the AI–VR group reported more varied emotional experiences. Several described AI's neutral tone as reducing anxiety associated with feedback: *"The AI never sounded angry or disappointed. After dozens of neutral prompts, I stopped associating feedback with shame. Now I hear the content, not the tone."* Others felt that AI-mediated practice did not fully prepare them for emotionally charged workplace feedback: *"The VR feedback didn't desensitize me because it wasn't real. In real clinic, a supervisor's tone can ruin my week."* Control participants more frequently described continuing to perceive feedback as threatening: *"Watching these events has made direct criticism more threatening. Every time I see it happen, I imagine it happening to me."* However, one control participant reported becoming less emotionally reactive through repeated clinical exposure, suggesting that adaptation also occurred outside the intervention. Collectively, these accounts suggest that students in the AI–VR+peer condition more consistently described emotionally supportive experiences of feedback than did participants in the other groups.

Discussion

Although interest in AI-supported feedback is increasing, empirical research examining such interventions through a feedback literacy perspective in medical education remains limited. Addressing this gap, the present study examined whether participation in a multicomponent AI-supported VR feedback-literacy intervention was associated with improvements in feedback literacy compared with usual clerkship training, how such participation influenced observable feedback-related behaviors, and how students' reflections helped explain supervisor-observed patterns. The findings, based on qualitative and quantitative analyses, yield several points for discussion.

Regarding the first research question, students who attended the experimental conditions demonstrated greater improvements in observable feedback literacy-related behaviors than those receiving usual clinical training. These findings suggest that participation in a structured multicomponent feedback-learning environment, rather than routine clinical exposure alone, was associated with greater short-term enactment of productive feedback engagement behaviors. One possible explanation is that early clinical environments primarily prioritize patient care and service delivery, leaving comparatively few structured opportunities for students to rehearse productive feedback engagement behaviors through guided reflection and dialogue (42-47). Accordingly, preparatory simulation-based activities may provide opportunities for students to practice these behaviors before encountering authentic clinical challenges (39, 48). This interpretation is consistent with previous work suggesting that feedback literacy can be intentionally developed through structured and safe opportunities for feedback engagement, reflection, and learner participation (11, 22, 24, 32, 49).

It is important to note that, unlike previous studies, the intervention was not designed simply as a technology-mediated feedback tool (36), but as a structured pedagogical feedback context in which AI facilitated dialogic interactions and supported the behavioral enactment of feedback literacy processes (22, 36, 37). Specifically, the intervention used ChatGPT as a Socratic dialogue partner, inviting students to articulate their

reasoning, examine assumptions, and initiate reflective dialogue. Such learner-initiated interaction aligns closely with theoretical descriptions of feedback literacy as an active rather than passive process (11,17,38,50–52).

Regarding the second research question, the qualitative findings provide important insight into how these quantitative improvements may have been achieved. Although no statistically significant differences were observed between the AI–VR and AI–VR+peer groups in supervisor-rated feedback literacy, qualitative data highlighted substantively different orientations toward feedback engagement across conditions. Students in the AI–VR condition reported more varied patterns of AI use, ranging from reflective questioning and justification of clinical reasoning to predominantly verification-oriented use. In contrast, students in the AI–VR+peer condition more consistently described feedback as a collaborative process characterized by shared sense-making, negotiated uncertainty, calibration of clinical judgment, and reduced emotional defensiveness. These findings suggest that similar short-term behavioral outcomes may emerge through different learning processes. While AI-supported reflection appeared sufficient to promote observable feedback engagement during early clerkship, peer dialogue may have influenced the quality of that engagement by shaping learners' orientation toward uncertainty, collaborative reasoning, and emotional regulation (31, 50). Rather than producing substantially greater observable behavioral gains within the study period, peer interaction appeared to enrich the way students experienced and interpreted feedback.

The qualitative findings also help explain the pattern observed in the control group. Whereas many students in the intervention groups described uncertainty as a legitimate component of clinical reasoning and feedback as an opportunity for learning, several control participants continued to interpret uncertainty as evidence of incompetence and described using AI primarily to confirm existing judgments rather than to interrogate their reasoning. These accounts provide a plausible explanation for why routine clinical exposure produced only modest improvements in observable feedback-literacy behaviors despite ongoing workplace experience. At the same time, the presence of negative cases across all groups—including students who found peer uncertainty anxiety-provoking or who developed greater tolerance for uncertainty through clinical exposure alone—illustrates that learners did not respond uniformly to either the intervention or the clinical environment. These disconfirming cases strengthen the credibility of the qualitative interpretation by demonstrating that alternative experiences were actively considered rather than excluded.

Importantly, the qualitative findings should be viewed as providing interpretive depth rather than evidence of causal mechanisms. The reflective activities cannot determine whether students' different orientations toward feedback resulted directly from the intervention or reflected pre-existing individual differences. Nevertheless, the convergence between supervisor-observed behavioral improvements and students' descriptions of changing approaches to uncertainty, dialogue, and feedback engagement strengthens confidence in the overall interpretation of the findings.

These findings are consistent with Carless and Boud's (11) argument that feedback literacy extends beyond receiving information to actively making sense of, evaluating, and acting upon feedback. They also align with research suggesting that peer dialogue promotes collaborative interpretation and encourages learners to seek clarification and challenge their own assumptions (51). More broadly, the present findings support recent work indicating that AI is likely to be most educationally valuable when embedded within dialogic learning processes rather than positioned as a replacement for human interaction (13, 22, 52-54). For example, Rad et al., (55) suggest that when AI is not designed as a substitute for human creativity, it can serve as a source of inspiration. Similarly, according to Kim et al., (56), ChatGPT has shown potential to influence the feedback process by facilitating dialogic interactions to enhance the learning experience and real-time feedback engagement. Rather than viewing AI as the primary educational agent, the findings suggest that its value may lie in supporting structured reflection within learning environments that also foster learner agency, psychological safety, and collaborative meaning-making.

Finally, the findings should be interpreted in light of the study's relatively short follow-up period. Although improvements were observed in supervisor-rated feedback behaviors during the first two weeks of clerkship, future longitudinal studies incorporating repeated behavioral observations, self-report measures, and workplace performance indicators are needed to determine whether the changes observed here are sustained over time and translate into broader feedback literacy capabilities.

Conclusion

This study suggests that participation in a multicomponent AI-supported VR feedback-literacy program was associated with short-term improvements in supervisor-observed feedback-literacy behaviors compared with usual clerkship activities. Because the intervention integrated AI-generated reflective prompt, VR simulation, peer dialogue, and preparatory instruction, these improvements should be interpreted as reflecting the educational value of the multicomponent program rather than the independent effect of any single component. Although the difference between the two experimental groups was small and did not reach statistical significance, the qualitative findings suggest that peer dialogue may enrich students' engagement with feedback by supporting collaborative sense-making, normalization of uncertainty, and emotional regulation. Importantly, the qualitative data also demonstrated substantial within-group variability and disconfirming cases, indicating that students did not engage with the intervention uniformly. Future research should examine how learner characteristics interact with AI-supported educational designs and should seek to disentangle the respective contributions of AI, VR simulation, and peer dialogue within multicomponent feedback-literacy interventions.

Limitations and future implication

Despite providing valuable insights into the integration of peer feedback and GenAI within VR-based learning environments, several limitations should be acknowledged when interpreting these findings. A key limitation of this study is the absence of an active, time-equivalent control group. Therefore, the observed effects cannot be attributed solely to the intervention itself, as they may partly reflect differences in time on task, attention, or engagement across groups. Second, the study was conducted at a single institution with a relatively small cohort of fourth-year medical students. Although appropriate for this initial evaluation, the findings may not generalize to learners at different stages of training, institutions, or educational contexts. Replication across larger and more diverse populations is therefore warranted. Third, although the study moved beyond self-report by incorporating supervisor observations and qualitative reflections, these outcome measures primarily captured students' observable engagement with feedback during the early clerkship period. They may not fully reflect broader changes in learners' underlying beliefs, dispositions, or long-term feedback literacy capabilities. Future research could strengthen evaluation by combining behavioral observations with validated self-report measures, learning analytics, or other longitudinal indicators of feedback engagement.

An additional limitation relates to the alignment between the intervention and the primary outcome measure. Several SOC-FL items assessed behaviors that were intentionally rehearsed during the intervention, such as seeking clarification, requesting feedback, and implementing feedback in subsequent performance. While this alignment appropriately reflects the intervention's educational objectives, it also raises the possibility that the observed improvements represent successful acquisition of specifically practiced behaviors rather than broader, transferable feedback literacy. To address this concern, the post-test was conducted two weeks after students entered authentic clinical clerkships rather than immediately following the intervention, and qualitative reflections explored how students described applying feedback strategies in real clinical settings. Although these findings provide preliminary evidence that some students transferred elements of their learning into workplace practice, the relatively short follow-up period precludes conclusions regarding the durability or broader transferability of these changes across clinical contexts. Future research should examine convergence with independent measures, established feedback literacy instruments, and longer-term workplace outcomes to determine whether these behavioral improvements generalize across contexts.

Declarations

Ethical approval and consent to participate Ethical approval was obtained from the Ethics committee of Fasa University of Medical Sciences (Ethics code: IR.FUMS.AEC.1404.34), and all research procedures were conducted in accordance with ethical standards for educational research. Electronic informed consent (describing the study purpose, data collection, and the right to withdraw at any stage of the study) was obtained from all participants.

Consent for publication Consent for publication was obtained from all participants. However, images of students were not permitted to be published publicly.

Availability of data and materials The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests No conflict of interest was reported related to this work.

Funding No funding was required for this research.

Authors' contributions: All authors contributed to the study design, data collection, analysis, and manuscript preparation. All authors approved the final manuscript.

Acknowledgements Not applicable

References

1. Delva D, Sargeant J, Miller S, Holland J, Alexiadis Brown P, Leblanc C, et al. Encouraging residents to seek feedback. *Medical teacher*. 2013;35(12):e1625-e31.
2. Mukhia S, Tenzin K. Training academic staff for effective feedback in workplace-based assessment: a study in Bhutan. *BMC Medical Education*. 2025;25(1):748.
3. Archer JC. State of the science in health professional education: effective feedback. *Medical education*. 2010;44(1):101-8.
4. Teunissen PW, Stapel DA, van der Vleuten C, Scherpbier A, Boor K, Scheele F. Who wants feedback? An investigation of the variables influencing residents' feedback-seeking behavior in relation to night shifts. *Academic medicine*. 2009;84(7):910-7.
5. Watling C, Driessen E, van der Vleuten CP, Vanstone M, Lingard L. Understanding responses to feedback: the potential and limitations of regulatory focus theory. *Medical education*. 2012;46(6):593-603.
6. Shafian S, Ilaghi M, Shahsavani Y, Okhovati M, Soltanizadeh A, Aflatoonian S, et al. The feedback dilemma in medical education: insights from medical residents' perspectives. *BMC Medical Education*. 2024;24(1):424.
7. Telio S, Ajjawi R, Regehr G. The "educational alliance" as a framework for reconceptualizing feedback in medical education. *Academic Medicine*. 2015;90(5):609-14.
8. Johnson NR, Pelletier A, Royce C, Goldfarb I, Singh T, Lau TC, et al. Feedback Focused: a learner-and teacher-centered curriculum to improve the feedback exchange in the obstetrics and gynecology clerkship. *MedEdPORTAL*. 2021;17:11127.
9. Grau TC. Enhancing Feedback by Fostering Resident Feedback Literacy: University of Pittsburgh; 2022.
10. Hattie J. Visible learning: A synthesis of over 800 meta-analyses relating to achievement: routledge; 2008.
11. Carless D, Boud D. The development of student feedback literacy: Enabling uptake of feedback. *Assessment & evaluation in higher education*. 2018;43(8):1315-25.
12. Winstone NE, Nash RA, Parker M, Rowntree J. Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational psychologist*. 2017;52(1):17-37.
13. Zhan Y, Yan Z. Students' engagement with ChatGPT feedback: Implications for student feedback literacy in the context of generative artificial intelligence. *Assessment & Evaluation in Higher Education*. 2025:1-14.
14. Han Y, Xu Y. Student feedback literacy and engagement with feedback: A case study of Chinese undergraduate students. *Teaching in Higher Education*. 2021;26(2):181-96.
15. van der Kleij FM. Evaluation of the 'Feedback Engagement Enhancement Tool' to examine and enhance students' engagement with feedback on their writing. *Studies in Educational Evaluation*. 2020;66:100907.
16. Evans C. Enhancing assessment feedback practice in higher education: The EAT framework. 2016.

17. Winstone NE, Mathlin G, Nash RA, editors. Building feedback literacy: Students' perceptions of the developing engagement with feedback toolkit. *Frontiers in Education*; 2019: Frontiers Media SA.
18. Carless D. Teacher feedback literacy, feedback regimes and iterative change: towards enhanced value in feedback processes. *Higher Education Research & Development*. 2023;17;42(8):1890-904.
19. Zhan Y. Beyond technology: Factors influencing the effects of teachers' audio feedback on students' project-based learning. *Technology, Pedagogy and Education*. 2023;1;32(1):91-103.
20. Henderson M RT, Phillips M. The challenges of feedback in higher education. *Assessment & Evaluation in Higher Education*. 2019.
21. Winstone Nash CD. Designing effective feedback processes in higher education. A learning-focused approach Routledge. 2019.
22. Jin FJ-Y, Maheshi B, Lai W, Li Y, Gasevic D, Chen G, et al. Students' perceptions of generative AI-powered learning analytics in the feedback process: A feedback literacy perspective. *Journal of Learning Analytics*. 2025;12(1):152-68.
23. Yan L, Martinez-Maldonado R, Jin Y, Echeverria V, Milesi M, Fan J, et al. The effects of generative AI agents and scaffolding on enhancing students' comprehension of visual learning analytics. *Computers & Education*. 2025:105322.
24. Zhan Y, Boud D, Dawson P, Yan Z. Generative artificial intelligence as an enabler of student feedback engagement: a framework. *Higher Education Research & Development*. 2025;44(5):1289-304.
25. Matthews A, Hall M, Parra JM, Hayes MM, Beltran CP, Ranchoff BL, et al. Receiving real-time clinical feedback: a workshop and OSTE assessment for medical students. *Advances in Medical Education and Practice*. 2020:861-7.
26. Hicke Y, Geathers J, Rajashekar N, Chan C, Jack AG, Sewell J, et al. MedSimAI: simulation and formative feedback generation to enhance deliberate practice in medical education. *arXiv preprint arXiv:250305793*. 2025.
27. Price M, Handley K, & Millar J. Feedback: focusing attention on engagement. *Studies in Higher Education*. 2011;36(8):879-96.
28. Khakpaki A. Advancements in artificial intelligence transforming medical education: a comprehensive overview. *Medical Education Online*. 2025;30(1):2542807.
29. Hicke Y, Geathers J, Rajashekar N, Chan C, Jack A, Sewell J, et al. MedSimAI: Simulation and formative feedback generation to enhance deliberate practice in medical education. *arXiv*. Preprint posted online March. 2025;1.
30. Mergen M, Meyerheim M, Graf N. Reviewing the current state of virtual reality integration in medical education—a scoping review protocol. *Systematic Reviews*. 2023;12(1):97.
31. Filius RM, de Kleijn RA, Uijl SG, Prins FJ, van Rijen HV, Grobbee DE. Strengthening dialogic peer feedback aiming for deep learning in SPOCs. *Computers & education*. 2018;125:86-100.
32. Hoo H-T, Deneen C, Boud D. Developing student feedback literacy through self and peer assessment interventions. *Assessment & Evaluation in Higher Education*. 2022;47(3):444-57.
33. Boscardin CK, Gin B, Golde PB, Hauer KE. ChatGPT and generative artificial intelligence for medical education: potential impact and opportunity. *Academic Medicine*. 2024;99(1):22-7.
34. Chan CKY, Hu W. Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*. 2023;20(1):43.
35. Nazaretsky T, Mejia-Domenzain P, Swamy V, Frej J, Käser T, editors. AI or human? Evaluating student feedback perceptions in higher education. *European conference on technology enhanced learning*; 2024: Springer.
36. Lin X. A preliminary study investigating adult learners' perspectives of using ChatGPT to seek effective assignment feedback. *The Australian Educational Researcher*. 2025;52(5):3393-420.

37. Bearman M, Ajjawi R. Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology*. 2023;54(5):1160-73.
38. Molloy E, Boud D, Henderson M. Developing a learning-centred framework for feedback literacy. *Assessment & Evaluation in Higher Education*. 2020;45(4):527-40.
39. Ebihara H, Kasai H, Shimizu I, Shikino K, Tajima H, Kimura Y, et al. Development of a clinical clerkship mentor using generative AI and evaluation of its effectiveness in a medical student trial compared to student mentors: 2-part comparative study. *JMIR Medical Education*. 2025;11:e76702.
40. Lawshe CH. A quantitative approach to content validity. *Personnel psychology*. 1975;28(4):563-75.
41. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative research in psychology*. 2006;3(2):77-101.
42. Bing-You R, Hayes V, Palka T, Ford M, Trowbridge R. The art (and artifice) of seeking feedback: clerkship students' approaches to asking for feedback. *Academic Medicine*. 2018;93(8):1218-26.
43. Marambe KN, Edussuriya DH, Jayawickrama D. Feedback provided by clinical teachers during undergraduate medical training: a Sri Lankan experience. *The Asia Pacific Scholar*. 2020;5(3):20.
44. Natesan S, Jordan J, Sheng A, Carmelli G, Barbas B, King A, et al. Feedback in medical education: an evidence-based guide to best practices from the council of residency directors in emergency medicine. *Western Journal of Emergency Medicine*. 2023;24(3):479.
45. Ramani S, Post SE, Könings K, Mann K, Katz JT, van der Vleuten C. "It's just not the culture": a qualitative study exploring residents' perceptions of the impact of institutional culture on feedback. *Teaching and learning in medicine*. 2017;29(2):153-61.
46. Molloy E. Time to pause: giving and receiving feedback in clinical education. *Clinical education in the health professions*. 2009;1:128-46.
47. Noble C, Billett S, Armit L, Collier L, Hilder J, Sly C, et al. "It's yours to take": generating learner feedback literacy in the workplace. *Advances in Health Sciences Education*. 2020;25(1):55-74.
48. Darvishi A, Khosravi H, Sadiq S, Gašević D, Siemens G. Impact of AI assistance on student agency. *Computers & Education*. 2024;210:104967.
49. Winstone NE, Nash RA, Rowntree J, Parker M. 'It'd be useful, but I wouldn't use it': barriers to university students' feedback seeking and reciprocity. *Studies in Higher Education*. 2017;42(11):2026-41.
50. Qian J, Li D. Toward a better understanding of student engagement with peer feedback: A longitudinal study. *International Review of Applied Linguistics in Language Teaching*. 2025;63(1):709-34.
51. Tripodi N, Vaughan B, Wospil R. Peer feedback as a strategy to foster feedback literacy in first-year allied health students. *MedEdPublish*. 2021;10.
52. Amir A. AI-Generated Dialogic Feedback: Designing a Pedagogical Chatbot Grounded in Literacy Resilience Principles. *Education Sciences*. 2026;16(2):318.
53. Saini AK, Cope B, Kalantzis M, Zapata GC. The future of feedback: Integrating peer and generative AI reviews to support student work. *EdArXiv preprint*. 2024:1-57.
54. Wood J. Making peer feedback work: the contribution of technology-mediated dialogic peer feedback to feedback uptake and literacy. *Assessment & Evaluation in Higher Education*. 2022;47(3):327-46.
55. Rad HS, Alipour R, Jafarpour A. Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of Wordtune application. *Interactive Learning Environments*. 2024;32(9):5020-40.
56. Kim H-S, Cha Y, Kim NY. Effects of AI chatbots on EFL students' communication skills. *영어학*. 2021;21:712-34.

Appendix 1:

Supervisor Observational Feedback Literacy Scale (SOFLS-21)

Scale: 1= Never observed, 2= Rarely observed, 3= Sometimes observed, 4= Often observed, 5= Consistently observed

Domain 1: Feedback-Seeking Behaviors

1. This student actively asks supervisors for feedback on specific aspects of their clinical performance. 1 2 3 4 5
2. This student seeks feedback from peers on their clinical reasoning or communication. 1 2 3 4 5
3. This student requests clarification or elaboration when feedback is unclear. 1 2 3 4 5
4. This student proactively identifies areas of uncertainty and asks for guidance. 1 2 3 4 5
5. This student seeks feedback before completing a clinical task (rather than only after). 1 2 3 4 5

Domain 2: Interpretation of Feedback

6. This student accurately paraphrases or summarizes the feedback they receive. 1 2 3 4 5
7. This student asks clarifying questions to better understand feedback content. 1 2 3 4 5
8. This student connects feedback to specific clinical actions or decisions. 1 2 3 4 5
9. This student distinguishes between feedback on clinical reasoning vs. communication vs. procedural skills. 1 2 3 4 5
10. This student identifies the key learning point from feedback without needing explicit direction. 1 2 3 4 5

Domain 3: Emotional Regulation

11. This student remains composed when receiving critical or corrective feedback. 1 2 3 4 5
12. This student does not display defensive behaviors (e.g., excuse-making, blaming, interrupting) when receiving feedback. 1 2 3 4 5
13. This student thanks the supervisor for feedback, even when it is critical. 1 2 3 4 5
14. This student maintains eye contact and open body language during feedback conversations. 1 2 3 4 5
15. This student recovers quickly from perceived feedback-related disappointment and continues engagement. 1 2 3 4 5

Domain 4: Enactment of Feedback

16. This student implements specific changes to their clinical practice based on previous feedback. 1 2 3 4 5
17. This student demonstrates improvement on previously identified areas of weakness. 1 2 3 4 5
18. This student references prior feedback when discussing current clinical decisions (e.g., "Last time you told me to... so this time I..."). 1 2 3 4 5
19. This student develops their own strategies to improve learning based on feedback received. 1 2 3 4 5
20. This student integrates feedback from multiple sources (supervisors, peers, AI) into their practice. 1 2 3 4 5
21. This student applies feedback from one clinical context to new or different clinical situations. 1 2 3 4 5

Supplementary Files

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

- [supplementGPTinteractions.rar](#)