ARTICLE IN PRESS

npj Digital Medicine

https://doi.org/10.1038/s41746-026-02862-5

Article in Press

Effects of digital versus face-to-face peer support on clinical skills: a three-arm randomized trial

Received: 22 January 2026

Accepted: 29 May 2026

Cite this article as: Yurdal, M.O., Kıncal, R.Y. Effects of digital versus face-to-face peer support on clinical skills: a three-arm randomized trial. *npj Digit. Med.* (2026). https://doi.org/10.1038/s41746-026-02862-5

Mustafa Onur Yurdal & Remzi Y. Kıncal

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

Effects of Digital Versus Face to Face Peer Support on Clinical Skills A Three Arm Randomized Trial

Mustafa Onur Yurdal[a]*, Remzi Y. Kıncal[b]

[a]*Medical Education, Çanakkale Onsekiz Mart University, Çanakkale, Turkey;* [b]*Faculty of Education, Çanakkale Onsekiz Mart University, Çanakkale, Turkey*

*Corresponding Author: monuryurdal@comu.edu.tr, Çanakkale Onsekiz Mart University, Department of Medical Education, Çanakkale/Turkey. ORCID ID: https://orcid.org/0000-0002-9632-7192

ARTICLE IN PRESS

ARTICLE IN PRESS

Effects of Digital Versus Face to Face Peer Support on Clinical Skills A Three Arm Randomized Trial

ABSTRACT

Digital interactive peer support may strengthen clinical skills when technology is coupled with pedagogy. This study was designed as a three-arm randomized controlled trial (N=98) involving medical students and compared a digital peer-feedback platform with face-to-face peer support and routine instructor feedback during training for intravenous injection, suturing/knot tying, and endotracheal intubation. OSCE performance was assessed post-intervention and at 7 weeks; reflection skills were measured with the Groningen Reflection Ability Scale (Turkish version). Reflective journals underwent content analysis and were integrated with quantitative findings. Complete outcome data were available for 82 participants. Post-test and retention OSCE scores varied by group in all skills (Kruskal–Wallis H=28.95–47.45, all $p<.001$), with the highest and most consistent score observed in the digital arm. Reflective ability scores also favoured the digital arm (all $p=.001$). Qualitative themes identified revisitable video evidence, explicit criteria, and peer coaching as affordances for self-regulation, while technical frictions impeded certain learners. Trial registration: ClinicalTrials.gov; registration number: NCT07186569; date of registration: September 22, 2025.

INTRODUCTION

High standards of clinical skills training are a foundation of undergraduate medical education, and yet there is an ongoing challenge in ensuring all learners achieve and maintain competence for core procedures. OSCEs are extensively used for evaluating clinical competence as they provide standardized and criterion-referenced assessment in multiple stations [1,2].

ARTICLE IN PRESS

Nevertheless, challenges in clinical skills training such as inadequate access to practice, variation in standardized instruction, and lack of integration of digital tools and technology-enhanced learning approaches into clinical skills instruction persist across institutions [3–5]. These challenges are particularly critical within competency-based curricula, such as Türkiye's National Core Curriculum (NCC) where procedural competence is mandated not just at the time of assessment but also in subsequent clinical encounters where reliable procedural performance is essential for patient safety [6]. One key to connecting skills training to competent performance is feedback. Recent literature on OSCEs recommends that evaluation should serve not only to measure performance but to promote learning as well; hence, formative feedback using criterion-referenced feedback should be incorporated into educational practice [2]. Specifically, feedback should be particularized, provided on a relevant timeline, and constructive, allowing students to adjust their goals or plans accordingly [7, 8]. Nevertheless, medical students enrolled in clinical skills programmes may not receive feedback that meets these requirements, thereby reducing the educational value of practice and post-evaluation review [9, 10].

Peer learning and peer feedback have accordingly been identified as complementary strategies potentially capable of improving the accessibility of and engagement with feedback. Peer learning refers broadly to the acquisition of knowledge and skills through active interaction with fellow students, whereas peer feedback denotes the specific evaluative information that one student provides to another regarding observed performance [11].

Within peer assessment, students make judgments about the work of others against specified criteria and offer focused comment, potentially building analytical thinking and expertise in feedback as well as task completion skills [11-13]. Within clinical skills learning, peer learning is potentially beneficial if well-designed and if learning participants are able to apply a common set

ARTICLE IN PRESS

of criteria [14-16]. Explanatory theory further suggests that peer-provided explanations and feedback may be especially accessible to novice learners when peers share a similar language, level of understanding, and learning context [17].

Beyond performance alone, clinical skills training is also closely related to learners' reflective ability. Reflective ability may be understood as making sense of experience through cognitive and affective processing in ways that generate meaning and guide future action [12,18-20]. In health professions education, reflective practice has been linked to clinical reasoning, self-awareness, communication, and professional development, and it may also help learners recognize and reduce cognitive bias in clinical work [21-24].

Retention of clinical skills is also a central concern in medical education because gains observed immediately after training may diminish over time if practice is not revisited [25–27]. Research on spaced practice, in which learning episodes are distributed over time rather than clustered together, and retrieval practice, in which learners actively recall previously learned information, suggests that repeated engagement with material may reduce knowledge decay. However, achieving sustained retention in medical training remains challenging [28–32]

Digital approaches may offer specific advantages for feedback, reflection, and retention in clinical skills training, but the existing literature remains difficult to interpret because interventions, outcomes, and comparison conditions vary considerably across studies [33-37]. In particular, prior work has often examined digital feedback or peer support in isolation, rather than directly comparing digital interactive peer support and face-to-face peer support under the same curricular conditions and across a shared set of outcomes. This leaves an important gap regarding whether feedback modality is associated with differences in immediate clinical skill performance, reflective ability, and subsequent skill retention. Accordingly, the present three-arm randomized controlled

ARTICLE IN PRESS

trial was designed to compare digital interactive peer support, face-to-face peer support, and routine training/instructor feedback in undergraduate medical students.

This three-arm randomized controlled trial aimed to compare the effects of digital interactive peer support, face-to-face peer support, and routine training/standard feedback on medical students' clinical skill performance immediately after the intervention and retention of clinical skills at the 7-week follow-up, assessed using OSCE-based ratings across intravenous injection, adult endotracheal intubation, and suturing/knot-tying. In addition, the study examined whether these conditions differed in students' reflective ability, assessed quantitatively using the Turkish version of the Groningen Reflection Ability Scale (GRAS-TR) [38,39], while reflective journals were analysed qualitatively and triangulated with the quantitative outcomes through joint displays to generate integrated meta-inferences [40,41].

RQ1 (Clinical skill performance). Do post-intervention OSCE performance scores differ among the digital interactive peer support, face-to-face peer support, and routine training groups?

RQ2 (Retention). Do OSCE performance scores at the 7-week follow-up (retention test) differ among the three groups?

RQ3 (Reflective ability). Do post-intervention GRAS-TR total scores differ among the three groups?

RQ4 (Triangulation and integration). How do themes derived from students' reflective journals help explain, complement, or qualify the observed between-group differences in OSCE performance, retention, and GRAS-TR scores when results are integrated via joint displays?

RESULTS

ARTICLE IN PRESS

Participant enrolment, allocation, follow-up, and analysis are summarized in the CONSORT flow diagram (Figure 1). Of 113 students assessed for eligibility, 98 consented and were randomized; 15 declined participation and none were excluded for not meeting eligibility criteria. Students who declined were not disadvantaged in any way: as the study procedures supplemented — rather than replaced — the existing first-year curriculum, and the target clinical skills are taught to all students as part of the Term 2 curriculum regardless of study participation, non-participation had no effect on students' curricular training or academic standing. Participants were allocated to routine training/instructor feedback (n = 33), face-to-face peer support (n = 33), or digital interactive peer support (n = 32). Complete-case outcome data were available for 82 participants (control n = 26; face-to-face n = 25; digital n = 31), and these participants were included in the primary analyses.

ARTICLE IN PRESS

ARTICLE IN PRESS

Across all three OSCE skills and at both time points, a consistent rank-order pattern emerged favouring the digital interactive peer support arm (Table 1). Notably, retention scores in the two peer-support arms exceeded post-test scores for IV injection and endotracheal intubation, while the digital arm maintained near-ceiling performance for suturing/knot-tying at the 7-week follow-up. Between-group differences in OSCE performance at post-test and at the 7-week retention assessment are shown in Figure 2.

Boxplots (Figure 2) display post-test and 7-week retention scores for intravenous injection (0–44), suturing/knot-tying (0–40), and endotracheal intubation (0–40) in the control (routine training/instructor feedback), face-to-face peer support, and digital interactive peer support arms. Boxes represent the interquartile range (IQR), horizontal lines indicate the median, whiskers indicate dispersion, and points denote outliers. Across skills and time points, Figure 2 shows a consistent pattern in which the digital interactive peer support group achieved higher OSCE scores than the face-to-face peer support and control groups. This separation is evident at post-test for all three skills and is maintained at the 7-week assessment, with particularly high and tightly clustered retention scores in the digital group for suturing/knot-tying and endotracheal intubation. In

ARTICLE IN PRESS

contrast, the control group displays lower central tendency and greater variability—most notably for intubation—suggesting less consistent skill acquisition and retention under routine training alone. Between-group differences in OSCE performance/retention and reflection outcomes are summarized in Table 2.

Kruskal–Wallis tests showed significant between-group differences for all OSCE outcomes at post-test and at 7-week retention (all $p<.001$), with the digital interactive peer support arm consistently outperforming the control arm and, in most outcomes, also the face-to-face arm; the clearest graded pattern (Control < F2F < Digital) was observed for IV injection (post-test and retention) and intubation retention. Baseline GRAS-TR subscales and total scores did not differ by group (all $p\geq.356$), whereas follow-up GRAS-TR self-reflection, reflective communication, and total scores were higher in the digital arm than in both comparison arms (all $p<.001$). Between-group differences in reflective ability at follow-up are presented in Figure 3.

Boxplots (Figure 3) show self-reflection, reflective communication, and total GRAS-TR scores in the control (routine training/instructor feedback), face-to-face peer support, and digital interactive peer support arms. Boxes represent the interquartile range (IQR), horizontal lines indicate the median, whiskers indicate dispersion, and points denote outliers. Figure 3 shows higher follow-up reflective ability in the digital interactive peer support arm compared with both the control and face-to-face peer support arms across all GRAS-TR outcomes, with a visibly higher central tendency for self-reflection, reflective communication, and total GRAS-TR scores. In contrast, the control and face-to-face groups display similar distributions with greater spread, whereas the digital group scores are more tightly clustered near the upper end of the scale, consistent with the between-group differences summarized in Table 2.

ARTICLE IN PRESS

In reporting below, participant excerpts are presented as English translations (participant IDs preserved). Consistent with the code–theme structure (Table 3), "Educational Gain" emerged as the core theme, accounting for 37% of initial codes and functioning as the central hub linking the remaining five themes (Specialized Feedback; Self-Efficacy & Confidence; Learning Strategy & Sustainability; Emotional Climate & Motivation; Technical-Structural Barriers/Facilitators) (Figure 4). Table 3 presents the full code–theme structure, and the complete qualitative codebook is provided in Supplementary Note 4. Briefly, Educational Gain emerged as the central theme reflecting observable learning outcomes across all groups, driven by the specificity and modality of feedback received (Specialized Feedback). Students' affective responses to feedback shaped their confidence and learning motivation (Self-Efficacy & Confidence; Emotional Climate & Motivation), while their post-session behaviors reflected varying degrees of self-regulated practice (Learning Strategy & Sustainability). Technical and structural conditions acted as either facilitators or barriers to these processes, particularly in the digital arm (Technical-Structural Barriers & Facilitators). Detailed theme descriptions follow below.

The network takes a star-shaped form in which Educational Gain is the central node, and five peripheral themes connect directly to it, indicating Educational Gain as the primary linkage point across themes.

The Educational Gain theme emerged as the core finding, with entries emphasizing observable learning gains (skill improvement, error correction, increased speed, and perceived transfer), with Educational Gain-related statements appearing in 89% of digital-group journals, 71% of face-to-face peer-feedback journals, and 48% of routine training/instructor feedback group journals. Students described how feedback supported micro-level error detection and iterative correction—e.g., the ability to slow down and replay performance footage to identify a specific knot error and

ARTICLE IN PRESS

correct it later ("…I could slow down and watch each step… I noticed the reverse wrap at the last knot, and correcting it afterward became easier," P5). In contrast to this, face-to-face feedback was felt to be beneficial though sometimes restricted due to a lack of opportunities to 're-view' ("…it was shown step by step, but I didn't have a chance to watch it a second time…," P17).

The second theme, Specialized Feedback identified the specificity, concreteness, and actionability of feedback, which was categorized in three code groups: visual demonstration (SF-1, n=19), "millimetric" guidance (SF-2, n=15), and verbal detail (SF-3, n=11). In the digital interactive peer support group, students could specifically locate the error and its correction on the screen through markers, comments, and video playback: (P5). In the face-to-face peer support group, feedback was very specific when demonstrated in person: (P19). However, face-to-face feedback could not be revisited after the session, which some students found limiting: ("I'm not able to watch it again, so details might be lost," P17). In the routine training/instructor feedback group, feedback was not always specific, thus becoming hard to act upon: ("a little higher" isn't enough guidance," P24).

Beyond the quality of feedback content, reflective journals also revealed how feedback shaped students' sense of capability — a pattern captured under the theme of Self-Efficacy and Confidence, which encompassed mastery beliefs, reduced anxiety, and confidence gained through peer validation. The digital interactive peer support group students reported greater confidence and composed behavior while performing correctly when they had examined the right technique ("My hand is no longer shaking while performing intubation… after seeing the right angle, now I think I can do it right," P3; "After the explanation, now I can remain calmer in the second trial," P12). Face-to-face students emphasized encouragement and immediate correction as confidence-building ("…he showed the wrong movement on my wrist… it reduced my hesitation," P19; "Hearing where I did it right reduced my anxiety," P17), though some noted uncertainty when

ARTICLE IN PRESS

details could not be revisited (P21). The routine training/instructor feedback group excerpts suggested confidence could be fragile when feedback was brief or unclear ("…I can't be sure where I am wrong… I remain indecisive," P30).

Alongside confidence-related responses, journals also revealed how students translated feedback into concrete post-session behaviors — a pattern termed Learning Strategy and Sustainability — encompassing repeat planning (LS-1, n=23), self-testing and self-assessment (LS-2, n=17), and peer teaching or joint practice (LS-3, n=15).Digital-group students described structured cycles of replay, checklist-based self-scoring, timing attempts, and peer coaching ("I score myself with the same checklist and try to do it within 3 minutes," P6; "I timed myself… reduced the time each attempt," P12; "I marked my friend's errors with screenshots… we recorded several more practice videos," P9). In the face-to-face group, strategy depended more on access to space/time for practice ("…we planned to repeat when the lab is available," P21). Traditional-group students reported more improvised repetition and uncertainty in self-evaluation ("…I try when there's an opening, but I don't know if I'm improving," P28; "I just tick each step… but can't ask when I don't understand," P24).

ARTICLE IN PRESS

The strategies students described were also embedded in a broader affective context — captured under the theme of Emotional Climate and Motivation — which encompassed positive affect (ECM-1, n=20), challenge orientation (ECM-2, n=11), and barriers or stressors (ECM-3, n=16), with positive affect more prominent in the digital and face-to-face conditions. Digital-group students often framed learning as engaging and playful ("…it felt like a game… friends commented on every version… sometimes with funny emojis," P12) and described motivation to repeat after seeing peer comments (P9). Face-to-face students similarly highlighted enjoyment and encouragement ("…practicing with my peer was fun… motivation increased," P19; "'Great' /

ARTICLE IN PRESS

'now it's correct' was encouraging," P17). Challenge-oriented narratives described goal-like "competition" with oneself or peers (e.g., aiming for correct angles or faster suturing, P3/P12/P21), whereas stressors were more salient in the traditional condition (e.g., practicing alone; limited detailed feedback), and in the digital condition were more often technical (e.g., slow uploads disrupting flow, P15).

The affective and strategic patterns described above were further shaped by contextual conditions that either enabled or constrained engagement — captured under the theme of Technical-Structural Barriers and Facilitators — which encompassed platform speed and access (TS-1, n=10), interface usability and "trace" features (TS-2, n=8), and material or environment access (TS-3, n=14), particularly salient in the digital arm. Digital-group excerpts showed how slow loading could interrupt engagement ("Videos sometimes load late; my flow breaks," P15), while interface tools (scrubbing/rewind, bookmarking, on-screen marking) helped restore continuity and support targeted review (P9/P5). In the traditional group, limited materials and lack of reference points were reported as barriers to repetition and self-assessment (P28/P30), sometimes prompting ad hoc compensations (creating a personal checklist, P24).

In this convergent mixed-methods trial, quantitative outcome results (OSCE post-test and 7-week retention scores; GRAS-TR reflection pre–post scores) were integrated with qualitative themes derived from reflective journals. We used a joint display to merge "statistics-by-themes" evidence across research questions and to generate meta-inferences by assessing the fit between strands (e.g., convergence, complementarity, or dissonance) (Table 4). This approach followed established methodological guidance on mixed-methods integration and joint displays [40,41].

This study compared three feedback modalities—digital interactive peer support, face-to-face peer support, and routine training/instructor feedback—in relation to medical students' clinical skill

ARTICLE IN PRESS

performance, 7-week skill retention, and reflective ability. Overall, the findings suggest that the pedagogical structuring of feedback is relevant to learning outcomes and that feedback designs which make process and self-regulatory elements visible and actionable may be associated with stronger outcomes than feedback limited to task confirmation alone.

Across the assessed OSCE skills—intravenous injection, suturing/knot-tying, and endotracheal intubation—the digital interactive peer support condition was associated with the highest post-test and follow-up scores. The face-to-face peer support condition generally showed intermediate results, whereas the routine training/instructor feedback condition yielded lower scores overall. For reflective ability, the combination of brief reflection tasks and feedback was associated with higher scores on self-reflection, reflective communication, and total reflection measures.

These conclusions should not be interpreted as evidence that a digital medium is inherently superior in all contexts or for all clinical tasks. Rather, the pattern observed here is more consistent with the value of a structured peer-feedback design in which learners can review performance evidence, use criterion-referenced guidance, and translate feedback into continued self-regulated practice. Replication in other settings, with longer follow-up and fuller assessment of implementation processes, is needed before stronger curricular conclusions are drawn.

ARTICLE IN PRESS

Discussion

In this study, the digital interactive peer support group achieved the highest OSCE scores across intravenous injection, suturing/knot-tying, and endotracheal intubation. The face-to-face peer support group generally scored between the digital interactive peer support group and the routine training/instructor feedback group, whereas the routine training/instructor feedback group had the lowest scores overall. Although score dispersion appeared narrower in the digital interactive peer

ARTICLE IN PRESS

support group, variability was not formally compared; therefore, this pattern should be interpreted cautiously.

These findings are compatible with social-cognitive and sociocultural learning perspectives, where observing peers' performances and explanations can enrich process cues and support learning through modelling and guided participation [42,43]. They are also consistent with self-regulated learning accounts in which learners benefit from opportunities to monitor performance evidence, calibrate self-judgments, and revise plans during iterative practice cycles [44]. From a feedback theory perspective, the pattern fits with the notion that feedback can be most productive not only in confirming a task is done correctly but rather in confirming levels of process and self-regulation [7].

Our interpretation does not attribute the observed differences to the "digital" medium alone. Rather, the results suggest that a pedagogically structured peer-feedback ecology—with explicit criteria, opportunities to revisit performance evidence (e.g., video), and dialogic cycles—may create conditions that facilitate both improvement and convergence in performance. This is consistent with contemporary accounts emphasizing learners' capacity to make sense of and use feedback (feedback literacy) and the importance of relational and contextual features that enable feedback to be acted upon [9,45,46].

The present findings are also broadly in line with evidence that video-based and peer-video feedback approaches can support skill development in health professions education, while noting that effects can be context- and task-dependent [37,47]. At the same time, randomized studies in procedural training have sometimes reported equivalence between technology-mediated and face-to-face formats, or advantages for face-to-face interaction for visuospatial subcomponents [48,49]. This body of evidence supports a cautious conclusion: design quality, task characteristics, and

ARTICLE IN PRESS

feedback structure likely shape whether digital peer support yields incremental gains over face-to-face approaches.

At the 7-week follow-up, the digital interactive peer support group retained higher OSCE scores than the comparison groups across the three assessed procedures. The face-to-face peer support group showed intermediate retention scores, whereas the routine training/instructor feedback group generally had the lowest follow-up scores. This pattern is consistent with the view that durable learning is supported when feedback prompts learners to revisit performance evidence, refine strategies, and re-engage in goal-directed practice—mechanisms emphasized in process/self-regulation feedback frameworks [7] and self-regulated learning models [44].

One possible explanation is that digital workflows may have supported repeated retrieval, reconstruction of performance steps, and more distributed engagement with feedback over time. From a cognitive perspective, longer-term performance is often supported by desirable difficulties such as retrieval effort and distributed practice [50]. Evidence syntheses also suggest that digitally delivered instruction may support more distributed learning than massed approaches in some contexts, although heterogeneity and risk of bias remain important considerations [51]. Third, there is meta-analytic evidence showing the positive impacts of video-based methods for health education skills outcomes, suggesting the general plausibility of video-mediated mechanisms of learning sustainment effects [47].

Significantly, however, the current findings should not be taken as suggesting an expectation of benefit generalizable to all digital environments. Rather, the data indicate that when peer feedback is criterion-referenced and revisitable and when tied to improvement plans, such feedback may be more effective than less structured peer feedback experiences at supporting long-term performance.

ARTICLE IN PRESS

The findings suggest that peer-support approaches, particularly when combined with structured prompts and feedback cycles, were associated with higher reflection-related scores in this sample. This is consistent with theoretical accounts positioning reflection as a regulated learning process supported by cues, standards, and opportunities to translate appraisal into forward-looking action [44,52,53]. This also fits feedback models emphasizing learners' movement from "what happened" toward "what to improve next," and with feedback literacy perspectives that foreground the learner's interpretive and action capacities [7, 45].

This is empirically interpreted on the basis that reflective capacity in medical students is best developed through structured, supported reflection, which involves triggers, guidance, discussion, and feedback. Literature shows that interventions which combine multiple components are commonly associated with stronger reflective outcomes, for example, prompts plus feedback and facilitation [54]. Broader reviews also identify that reflective writing and reflection activities tend to be most meaningful when longitudinally embedded and accompanied by structured feedback and facilitation [55,56].

Taken together, the reflective findings have suggested that the value of the interventions may lie less in modality per se but more in the consistent integration of reflection prompts with feedback processes which enable students to make judgments, manage affect, and translate insight into action-core features of feedback-literate engagement [7].

The present findings suggest that peer feedback may have educational value in procedural skills training when it is linked to clear criteria and to process- and self-regulation-oriented guidance about how to improve and what to do next. In this context, feedback is likely to be more educational when learners identify the gap between current and desired performance and supports action on that gap, rather than merely confirming whether a task was completed. Accordingly, medical

ARTICLE IN PRESS

schools may consider integrating structured peer-feedback cycles into skills-lab teaching as a supplement to, rather than a replacement for faculty feedback, particularly when faculty resources are limited.

A second implication concerns curriculum design. The pattern observed across immediate performance and 7-week follow-up is consistent with the educational value of distributed practice and repeated engagement with performance evidence. Skills curricula that support spaced practice, review, and revision may help reduce skill decay and strengthen consolidation over time, particularly when learners can revisit performance evidence and calibrate subsequent practice.

Finally, implementation is everything. Digital and face-to-face peer-feedback interventions will tend to be most effective when combined with (i) brief training in giving and using feedback (feedback literacy), (ii) attention to the relational conditions that support receptivity (educational alliance), and (iii) sufficient technical and institutional infrastructure. These design considerations might also enable mitigating known risks of peer feedback (e.g., variable quality, discomfort, or reduced engagement) and advance scalability across varied educational contexts.

Several limitations should be considered when interpreting these findings. First, the trial was conducted at a single medical school, and voluntary participation may limit transferability to other institutions, learner levels, and curricular settings. Second, although randomization supports baseline comparability, procedural skill outcomes were assessed without a baseline OSCE pretest to reduce pretest sensitization; therefore, residual baseline differences in procedural performance cannot be fully excluded. Third, participant blinding was not feasible because of the nature of the educational interventions, and informal sharing across groups may have introduced expectancy or contamination effects.

ARTICLE IN PRESS

Additional measurement limitations should also be noted. OSCE scores were strengthened using multiple raters and average judgments; however, an inter-rater reliability coefficient (e.g., ICC) was not reported, which limits inference about scoring consistency. Reflective ability outcomes were assessed using a validated self-report scale, but self-report is susceptible to social desirability and may not fully capture enacted reflective practice. Qualitative findings were derived from reflective journals completed by a subset of participants and may under-represent students who were less inclined to write extensively or disclose difficulties.

Future research should prioritize multicentre replication and longer follow-up intervals to examine whether retention advantages persist beyond seven weeks and across additional procedural domains. Studies that collect platform/process analytics (e.g., frequency and timing of review, feedback exchanges, dose of engagement) could test mediation hypotheses consistent with feedback and self-regulation models (e.g., whether repeated review and action-planning behaviours explain retention). Component or factorial designs could also help disentangle which elements (e.g., video evidence, asynchrony, peer dialogue, spacing) are most responsible for observed effects. Finally, mixed methods work that explicitly evaluates fit between strands through joint displays may strengthen explanatory claims by clarifying when qualitative themes converge with, complement, or diverge from quantitative patterns.

ARTICLE IN PRESS

METHODS

Study Design

We conducted a three-arm, parallel-group randomized controlled trial (RCT) comparing digital interactive peer support, face-to-face peer support, and routine training/standard feedback

ARTICLE IN PRESS

conditions, and reported the trial using the CONSORT 2025 framework [57] (including a participant flow diagram). The trial was registered at ClinicalTrials.gov; registration number: NCT07186569; date of registration: September 22, 2025. For procedural/psychomotor clinical skill outcomes (OSCE-based performance ratings), we used a randomized post-test-only control-group design with a delayed retention assessment (post-test + 7-week follow-up), rather than administering a baseline OSCE pretest. This decision was made to minimize threats to internal validity that are particularly salient in procedural skills research, including pretest sensitization and the testing effect [58-60]. In OSCE-style assessments, exposure to structured checklists and performance criteria can itself function as an instructional event—prompting learners to infer the required step sequence and thereby producing an artificial performance gain unrelated to the intervention [61]. Additionally, for ethical and practical reasons, administering an invasive/procedural OSCE to students before they had received skills training was not considered appropriate due to safety and training-ethics concerns. Accordingly, learning was evaluated immediately after instruction (post-test) and again at 7 weeks (retention), which provides evidence for both short-term acquisition and longer-term maintenance (Figure 5). Baseline comparability was supported through random allocation and eligibility criteria designed to exclude students with prior relevant experience. Outcome ratings were performed under a blinded assessment procedure to reduce scoring bias. Because reflective ability is a primarily cognitive/metacognitive construct rather than a procedural skill, it was assessed using a pretest–post-test approach, as the measurement itself was not expected to directly train the target clinical procedures. In parallel, we embedded a convergent mixed-methods component: qualitative data (including reflective journals and process-related records) were analysed to illuminate learners' experiences during the intervention and then integrated with the quantitative outcomes using joint displays to generate meta-inferences [40,41,62]. Additional details on the procedures implemented to support internal

ARTICLE IN PRESS

validity, reliability, and qualitative trustworthiness across the quantitative and qualitative strands are provided in Supplementary Table 1.

Ethical Approval

This study was a randomized controlled trial (RCT) registered at ClinicalTrials.gov (NCT07186569). These procedures were all conducted in accordance with the guidelines and standards. Written informed consent was obtained from all participants for participation and for the publication of any study-related images, videos, audio recordings, and personal details included in the manuscript or supplementary materials. No identifiable patient images, videos, or case details are included. This study was approved by the Scientific Research Ethics Committee of Canakkale Onsekiz Mart University on October 26, 2023 (Decision No: 13/58), and then approved by the Faculty of Medicine, Canakkale Onsekiz Mart University, on November 8, 2023 (Document No: E-30566003-604.02-2300274597).

Setting, participants, recruitment, and sampling frame

Eligible participants were Term 1 medical students at ÇOMÜ during the 2024–2025 academic year. Students were not eligible if they were enrolled in Term 2 or later, or if they reported prior knowledge or experience with the target clinical skills. Participation was voluntary; students who declined were not penalized in any way, and non-participation had no effect on their academic standing or progression within the programme.

This single-center study was conducted at the Çanakkale Onsekiz Mart University (ÇOMÜ) Faculty of Medicine within the context of the faculty's Professional/Clinical Skills training, using the Professional Skills Laboratory infrastructure for training delivery, and OSCE logistics supported through the faculty's clinical departments. The research specifically examined three clinical skills, namely, intravenous (IV) injection, suturing/knot-tying and endotracheal intubation, which are part of the Term 2 curriculum. Term 2 students and students in later years were also not included because these skills are routinely taught to all students in subsequent terms. A total of

ARTICLE IN PRESS

113 students were assessed for eligibility; 98 consented and were randomized, while 15 declined participations (no student was excluded for failing eligibility criteria). The randomized sample included 65 women and 33 men. Participants were allocated to the routine training/standard feedback control group (n=33), face-to-face peer support group (n=33), and digital peer support group (n=32) (Table 5). All three study arms were trained in the same clinical skills laboratory, and the OSCE was conducted in the same physical setting. To support procedural standardization across groups, the same target skills, core equipment, and manikins/task trainers were used for all participants. Although final group sizes differed at the analysis stage, access to practice materials was organized using the same student-to-manikin/task trainer ratio across groups (approximately 3 students per manikin/task trainer).

The study population comprised all Term 1 medical students enrolled at ÇOMÜ Faculty of Medicine during the 2024–2025 academic year (total cohort: approximately 200 students). Because all eligible students within the cohort were invited to participate, the recruited sample (n = 98) represented approximately 49% of the total first-year class. The gender composition of the sample (66.3% female, 33.7% male) was broadly consistent with the gender distribution of the full Term 1 cohort, and sex-stratified randomization ensured proportional representation across all three study arms (Table 5). As participants were drawn from a single medical school in northwest Turkey, generalizability to institutions with different cohort sizes, curricular structures, or resource availability should be considered with caution.

ARTICLE IN PRESS

Randomization Allocation

Participants were randomly allocated in a 1:1:1 ratio using a paper-draw procedure: identical folded slips bearing group codes (control, face-to-face, digital) were placed in an opaque bag [63,64]. Randomization was stratified by sex: the 33 male participants drew first (11 per arm), followed by the 66 female participants, yielding balanced sex distributions across arms (Female/Male: 22/11, 22/11, 21/11) [65]. Academic performance and high school type were considered as additional stratification variables but not retained, as first-year students had no medical school performance record. Post hoc checks confirmed no significant between-group differences on these variables ($p > 0.05$). Allocation concealment was maintained using indistinguishable folded slips; participants were informed only that they would receive one of three

ARTICLE IN PRESS

training approaches [66,67]. Illustrative documentation of the randomization and group enrolment procedures is provided in Supplementary Figure 1f, 1g.

Due to the nature of the educational interventions, participant blinding was not feasible. However, the study operated as single-blind at the analysis level: feedback forms and output data were analyzed with masked group codes. Additionally, OSCE assessors were not informed of participants' group allocation at either the post-test or retention assessment; students attended individually scheduled time slots, and no group-identifying information was visible to raters during scoring. Randomization and reporting procedures were aligned with CONSORT guidance for randomized trials [57].

Data collection tools

In this study, data were collected using the following tools:

Clinical skill outcomes were assessed using procedure-specific checklists — the Clinical Skills Rating Scales — administered as part of a structured OSCE at two time points: immediately after the intervention (post-test) and again 7 weeks later to evaluate skill retention. The Clinical Skills Rating Scales are established checklist-based assessment instruments used within the ÇOMÜ Faculty of Medicine undergraduate curriculum; they were not newly created for this study. For this study, the existing checklists for intravenous injection, adult endotracheal intubation, and suturing/knot-tying were reviewed and confirmed through a faculty-supported quality assurance process, which involved consultation with at least three subject-matter experts per skill. These experts reviewed each checklist item against current procedural standards and participated in refinement workshops to ensure the content and sequence of steps were appropriate for the study context. Each procedural checklist item ("step") was rated using a 3-point ordinal rubric: 0 = Not adequate (the step was omitted, performed incorrectly, or performed out of sequence), 1 = Partially adequate (the step was partially performed correctly), and 2 = Adequate (the step was performed correctly, without hesitation, and without assistance). The IV injection checklist comprised 22 steps (maximum score = 44), whereas the endotracheal intubation and suturing/knot-tying checklists comprised 20 steps each (maximum score = 40). The checklists were applied to the

ARTICLE IN PRESS

study participants themselves; no separate pilot scoring of participants was conducted. Each participant was evaluated simultaneously by three expert raters who had not previously taught the students, and the mean of the three ratings was used as the student's OSCE score to minimise idiosyncratic rater effects [68]. Although ICC was not computed — as only averaged scores were retained — the use of independent triplicate ratings provides a design-level control for rater-specific bias; this is acknowledged as a limitation, and future studies should retain individual rater scores to enable formal reliability estimation. The procedure-specific Clinical Skills Rating Scales (OSCE checklists) used for outcome assessment are provided in Supplementary Note 1.

Reflective ability was assessed using the Turkish version of the Groningen Reflection Ability Scale (GRAS-TR). The original GRAS was developed to measure medical students' reflection ability [38], and the Turkish adaptation, validity, and reliability study was conducted with medical students by Şenol et al. [39]. GRAS-TR is a 19-item, 5-point Likert-type self-report scale (1 = strongly disagree to 5 = strongly agree) designed to capture reflective capacity. In the Turkish validation, factor analytic findings supported a two-factor structure—self-reflection and reflective communication—as more culturally appropriate than the original three-factor model. Four items (items 3, 4, 12, and 17) are reverse scored, and total scores range from 19 to 95, with higher scores indicating higher reflective ability. Reported internal consistency in the Turkish adaptation was $\alpha = .789$ for self-reflection, $\alpha = .654$ for reflective communication, and $\alpha = .794$ for the total scale. The GRAS-TR measures general reflective capacity as a dispositional trait, rather than reflection on a specific experience; it is therefore appropriate for use as a pre–post measure across an intervention period. GRAS-TR was administered to all participants at two time points — at baseline (during the study information meeting on, prior to any intervention) and again on the same day as the 7-week retention OSCE. The Turkish version of the Groningen Reflection Ability Scale (GRAS-TR) used in this study is provided in Supplementary Note 2.

In the qualitative strand, reflective journals were used to capture students' subjective learning experiences during and after exposure to three feedback approaches (routine instructor-led feedback, face-to-face peer support, and digital interactive peer support). Reflective journals are widely described as a structured written record that supports the reconstruction of cognitive, affective, and behavioral processes following a learning experience [20,69]. In health professions

ARTICLE IN PRESS

education, reflective writing has been linked to the development of clinical reasoning and professional growth and is commonly used to contextualize and explain quantitative learning outcomes [22,70].

A semi-structured reflective journal form was developed to elicit reflection across three sequential components. The form elicited description of the clinical skills/feedback session, feelings and thoughts related to performance and feedback, and action planning for subsequent learning goals and improvement strategies. The structure was aligned with established models of reflective journaling and was designed to facilitate progression from descriptive accounts toward interpretive and action-oriented reflection [71]. The semi-structured reflective journal form used in the qualitative strand of the study is provided in Supplementary Note 3.

Content validity was evaluated by an expert panel (medical education, psychometrics, and educational technology), yielding high indices (mean I-CVI = 0.92; S-CVI/Ave = 0.94), consistent with commonly used criteria for strong content validity [72,73]. In a pilot coding exercise, inter-rater agreement was Cohen's $\kappa = 0.81$, indicating very good agreement. Because the journal tool is open-ended, internal consistency indices have limited interpretability; however, when rubric-based scoring was applied as an auxiliary check, Cronbach's $\alpha = 0.83$ was obtained. The GRAS-TR measures general reflective capacity as a dispositional trait; the journal form, by contrast, elicits reflection on a specific learning experience through structured prompts.

All randomized participants, regardless of group allocation, were invited to complete reflective journals throughout the study period. A brief orientation to reflective writing principles, drawing on Gibbs' reflective cycle and a structured 'describe–evaluate–reflect' approach supported with exemplars, was provided to all participants at the kick-off meeting to ensure a shared understanding of the task. It is acknowledged that this orientation may have introduced a degree

ARTICLE IN PRESS

of scaffolding that could have shaped the content or style of journal entries; this is noted as a potential limitation of the qualitative strand. Participants were asked to complete a journal entry after each clinical skills session and after the retention assessment; reminders were sent via the study communication channel. An example of the study communication channel used for reminders is shown in Supplementary Figure 1d. A total of 41 students submitted at least one journal entry, and 13 provided complete journal sets covering all sessions. The same semi-structured journal form and writing prompts were used across all three study arms, with no group-specific instructions; the only between-group difference was the feedback modality experienced during training, which participants were asked to reflect upon using the identical journal structure.

Process

The video-based digital interactive peer support intervention was delivered via Medifeeding, a previously developed and piloted digital peer-feedback platform [74]. Representative screenshots of the Medifeeding interface, including the student feedback screen and instructor/administrator panel, are provided in Supplementary Figure 1a-c. The platform was designed using Moonen's 3-Space model and informed by self-regulated learning (Zimmerman), peer assessment/feedback principles (Falchikov), and sociocognitive/sociocultural learning perspectives (Bandura; Vygotsky) [74]. The pilot study was conducted with second year (Term 2) medical students [74]; none of the participants in the present trial took part in the pilot, and Term 2 students were explicitly excluded from the current study to prevent contamination between cohorts. Following the pilot study, minor refinements to the platform and study procedures were implemented based on student feedback before initiating the present randomized trial (Figure 6). The QR codes used to access the Medifeeding application on iOS and Android devices are shown in Supplementary Figure 1h.

ARTICLE IN PRESS

After the digital peer-feedback tool had been updated to its final version for the experimental phase, first-year medical students were invited to participate via an in-class announcement (31 January 2025), and interested volunteers were asked to attend a study briefing meeting on 3 February 2025. At the kick-off meeting (3 February 2025, 10:00), the research team introduced the study, obtained written informed consent, conducted randomisation and group registration, provided a brief reflective-writing training (20 minutes; Gibbs cycle [75] and a "Describe–Evaluate–Reflect" approach), scheduled the intervention sessions, and administered the baseline reflection measure (GRAS-TR) to all participants (Figure 7). Documentation of the kick-off meeting is provided in Supplementary Figure 1e.

Intervention sessions were scheduled within the existing skills-laboratory timetable as a supplement to — not a replacement for — the participants' regular first-year clinical skills curriculum; the target procedures (intravenous injection, suturing/knot-tying, and endotracheal intubation) are part of the Term 2 curriculum and were therefore not yet part of participants' routine instruction at the time of the study. Attendance was recorded via signature at each session. Across all arms, the instructional time (30 minutes) and practice time (60 minutes) were held constant (Table 6). Across all three study arms, the instructional context was further standardized with respect to setting, target procedures, and demonstration format. Each group received the same faculty demonstration from two instructors before group-specific feedback procedures began. Group sizes differed during follow-up analyses (routine training/instructor feedback, n=26; face-to-face peer support, n=25; digital interactive peer support, n=31), but the training environment and access to practice materials were organized using the same student-to-manikin/task trainer ratio across groups.

ARTICLE IN PRESS

The principal between-group difference was the format and timing of feedback delivery rather than the procedural content or training environment. In all three study arms, the same target procedures were addressed in the same skills-laboratory context after the same instructor-led demonstration. Thus, the intervention contrast should be understood primarily as a difference in feedback modality: routine training/instructor feedback, face-to-face peer support, or digital interactive peer support.

In the routine training/instructor feedback group, students in the control group received the standardized instructor demonstration and then practised all three target procedures (intravenous injection, suturing/knot-tying, and endotracheal intubation) on the same manikins and task trainers used across all study arms — specifically, venipuncture arms, suture pads, and adult airway management manikins available in the ÇOMÜ clinical skills laboratory. Two instructors remained present throughout the 60-minute practice period and provided ongoing, real-time verbal feedback aligned with students' observed performance. The 30-minute instructional demonstration preceded the practice period, leaving the full practice time available for instructor-guided feedback.

In the face-to-face peer support group, before the first session, students attended a structured 20-minute interactive briefing on how to deliver constructive verbal peer feedback. The briefing included an explanation of the peer-feedback principles underpinning the structured peer-feedback form, worked examples of feedback statements, and an opportunity for questions. Students did not practice delivering feedback on clinical skills during this briefing. They then received the same standardized 30-minute instructor demonstration as the other groups, using the same manikins and task trainers. After the demonstration, the instructors left the session, and students practised all three target procedures and provided verbal, improvement-oriented peer feedback to one another using the validated structured peer-feedback form — the same form subsequently embedded in the digital platform for the digital group. The peer-feedback form guided students to comment on

ARTICLE IN PRESS

specific procedural steps aligned with the OSCE checklist, focusing on corrective suggestions rather than numeric scoring. Peer feedback was exchanged in real time during the 60-minute practice period, allowing students to incorporate feedback and continue practicing within the same session.

In the digital interactive peer support group, before the first session, students attended the same structured 20-minute briefing on peer-feedback principles as the face-to-face group and additionally received a hands-on orientation to the Medifeeding platform, which they installed on their own smartphones and accessed via personal user accounts. They then received the same standardized 30-minute instructor demonstration as the other groups, using the same manikins and task trainers. After the demonstration, the instructors left the session, and students practised all three target procedures in the clinical skills laboratory during the 60-minute practice period. At the end of the session, each student recorded a video of their final practice attempt using their smartphone and uploaded it to the Medifeeding platform. Peer feedback was then exchanged asynchronously through the platform using the digital version of the same validated structured peer-feedback form used in the face-to-face condition, supplemented by multimodal interaction features (audio comments, text comments, emoji, like/dislike responses, and video comments). Because feedback was delivered asynchronously via students' own smartphones, it was generally exchanged outside the laboratory and after the session had ended — unlike the face-to-face and control conditions, in which feedback was provided in real time during practice. Students in the digital group therefore did not receive peer feedback during the practice session itself; however, they retained access to the platform and their peers' feedback and could, if they chose, return to the laboratory for additional independent practice before the OSCE.

Thus, the key distinction across conditions was not the content of feedback — all groups used the same structured peer-feedback form — but its timing and modality: the control group received

ARTICLE IN PRESS

real-time instructor feedback during practice; the face-to-face group received real-time peer feedback during practice; and the digital group received asynchronous peer feedback after the practice session, mediated through video-based platform review.

The post-intervention OSCE was conducted 3 days after the final training session. To reduce queue effects and support random arrival ordering, students booked OSCE time slots in 3-minute intervals via a shared online scheduling sheet. The OSCE was conducted in the same clinical skills laboratory used for training, with stations configured to use identical manikins, task trainers, and equipment to those used during the training sessions, ensuring that the assessment environment was as consistent as possible with the training context. Before the OSCE, all students from all three study arms were informed that the skills laboratory would be available for optional additional practice. Each station was staffed by three expert raters who had not previously taught the students and were not informed about the study details, as an attempt to support assessor blinding. Specifically, assessors were not informed of participants' group allocation (control, face-to-face, or digital) at either the post-test or retention OSCE. Students attended individually scheduled time slots, and no group-identifying information was visible to raters during scoring, supporting allocation-blind assessment. Each student completed three stations (3 minutes per station; total 9 minutes). Scores for each skill were based on the mean of three raters' evaluations to minimise idiosyncratic rater effects and better represent the overall expert judgment trend [68].

OSCE scores were not disclosed to students immediately after the post-intervention assessment. This decision was made deliberately to prevent score knowledge from influencing subsequent practice behaviour or creating differential motivation between groups ahead of the retention assessment. Students were informed only that the OSCE would be repeated 7 weeks later. The

ARTICLE IN PRESS

retention OSCE was conducted 7 weeks after the post-intervention OSCE using the same procedures, manikins, task trainers, and clinical skills laboratory setting. The retention OSCE was conducted 7 weeks after the post-intervention OSCE using the same procedures and the same clinical skills laboratory setting. On the same day, participants completed GRAS-TR again, and reflective journals were collected and closed out at study end. The 7-week interval aligns with evidence that the effectiveness of spacing depends on the retention interval and that educational schedules should be designed with test delay in mind [28].

Because the study focused on comparing processes across naturally scheduled sessions, no make-up sessions were offered for students who missed scheduled training dates. Attrition and group withdrawals were planned to be reported using the CONSORT 2025 flow diagram approach.

Data analysis

Quantitative data were analysed using jamovi (version 2.3; an open-source statistical software package with a graphical user interface). Raw scores were imported, variables were labelled, and data quality checks were performed prior to inferential analyses. Missing values were examined; because the proportion of missing data was <5% and appeared randomly distributed, missing observations were imputed using two-way mean imputation. Outliers were screened using boxplots and z-scores ($|z| > 3.29$). For each OSCE skill score and for reflective ability scores (GRAS-TR total and subscale scores), we computed mean, standard deviation, median, and minimum–maximum values. Distributional assumptions were assessed using the Shapiro–Wilk test for normality and Levene's test for homogeneity of variances. Based on these diagnostics, non-parametric methods were selected [76]. Between-group comparisons of post-intervention OSCE performance scores and 7-week retention scores across the three study arms (routine

ARTICLE IN PRESS

training/instructor feedback, face-to-face peer support, digital interactive peer support) were conducted using the Kruskal–Wallis H test. When omnibus tests were statistically significant, Dunn's post-hoc procedure with Bonferroni adjustment (Dunn–Bonferroni) was used for pairwise comparisons to control the family-wise error rate [77,78]. The significance threshold was set at $p < .05$ (two-sided), with Bonferroni-corrected α applied according to the number of planned comparisons. Within-group change over time (post-test vs. 7-week follow-up OSCE scores) was evaluated using the Wilcoxon signed-rank test. Alongside test statistics (Z), effect sizes were computed and interpreted using conventional thresholds (small ≈ 0.10, medium ≈ 0.30, large ≈ 0.50) [79].

A formal a priori power analysis was not conducted prior to the study. The sample was determined by the total number of eligible Term 1 medical students enrolled during the 2024–2025 academic year (approximately 200 students), of whom 113 were assessed for eligibility and 98 consented to participate. A post hoc power analysis for the Kruskal–Wallis test (approximated via the analogous ANOVA framework with an asymptotic relative efficiency correction of 0.955 [76]) indicated that with the analysis sample ($N = 82$; $\alpha = 0.05$), the study provided approximately 88.5% power to detect a large effect ($f = 0.40$; Cohen, 1988 [79]). Observed effect sizes across all primary outcomes ranged from $\eta^2 = 0.28$ to $\eta^2 = 0.70$, all classified as large by conventional thresholds, confirming that the study was adequately powered to detect the magnitude of effects obtained. The ability to detect medium or smaller effects may have been limited; this is acknowledged as a potential limitation, and replication with larger samples is recommended.

In the first coding cycle, an additional coder independent of the primary investigator — an Associate Professor of Medical Education with a doctoral degree in qualitative research — generated open codes. In the second cycle, both coders collaboratively consolidated codes through

ARTICLE IN PRESS

iterative comparison and theme development, consistent with mixed-methods analytic guidance [80]. Discrepancies between coders were resolved through discussion and consensus. The coding framework was considered stable when coder agreement reached 92%, after which the dataset was treated as ready for reporting. Inter-coder reliability was estimated as Cohen's $\kappa = 0.82$, indicating substantial to almost perfect agreement using commonly cited benchmarks [81].

To integrate quantitative and qualitative findings, results were brought together using joint displays, enabling side-by-side comparison of outcome patterns and journal-derived themes and supporting the generation of meta-inferences [40,41].

Abbreviations:

ÇOMÜ: Çanakkale Onsekiz Mart University

IV: Intravenous Injection

OSCE: Objective Structured Clinical Examination

UÇEP: Ulusal Çekirdek Eğitim Programı (National Core Curriculum - Türkiye)

VBPF: Video-based peer feedback

F2F: Face to Face Peer Support Group

Declarations

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available due to institutional data governance policies but are available from the corresponding author on reasonable request.

ARTICLE IN PRESS

ARTICLE IN PRESS

Code Availability

Quantitative analyses were performed using jamovi (version 2.3, open-source statistical software). No custom code or scripts were developed for this study. All analysis procedures are fully described in the Methods section and can be reproduced using jamovi with the reported parameters.

Acknowledgments

This work was supported by Çanakkale Onsekiz Mart University Scientific Research Coordination Unit, Project number: SDK-2024-4838. The authors would like to thank all the students who participated in the study.

Authors' contributions

MOY and RYK developed the digital feedback tool, collected the data, and organized the interventions. MOY and RYK analyzed the data. All authors contributed equally to writing, revising, and editing the manuscript, and have read and approved the final version.

Competing interests

The authors declare that they have no competing interests.

Authors' information

- **Mustafa Onur Yurdal:** Çanakkale Onsekiz Mart University, Faculty of Medicine, Department of Medical Education, Çanakkale, Turkey, ORCID ID: https://orcid.org/0000-0002-9632-7192
- **Remzi Y. Kıncal:** Çanakkale Onsekiz Mart University, Faculty of Education, Department of Curriculum and Instruction, Çanakkale, Turkey, ORCID ID: https://orcid.org/0000-0002-6258-393X

REFERENCES

ARTICLE IN PRESS

ARTICLE IN PRESS

1. Harden, R. M. & Gleeson, F. A. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med. Educ.* **13**, 39–54; https://doi.org/10.1111/j.1365-2923.1979.tb00918.x (1979).
2. Khan, K. Z., Ramachandran, S., Gaunt, K. & Pushkar, P. The objective structured clinical examination (OSCE): AMEE guide no. 81. Part I: An historical and theoretical perspective. *Med. Teach.* **35**, e1437–e1446; https://doi.org/10.3109/0142159X.2013.818632 (2013).
3. Grantcharov, T. P. & Reznick, R. K. Teaching procedural skills. *BMJ* **336**, 1129–1131; https://doi.org/10.1136/bmj.39517.686956.47 (2008).
4. Ramani, S. & Leinster, S. AMEE Guide no. 34: Teaching in the clinical environment. *Med. Teach.* **30**, 347–364; https://doi.org/10.1080/01421590802061613 (2008).
5. Norcini, J. et al. Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. *Med. Teach.* **33**, 206–214; https://doi.org/10.3109/0142159X.2011.551559 (2011).
6. UÇEP-2020 UCG, Ulusal Cep-2020 UYVYCG & Ulusal Cep-2020 DSBBCG. Medical faculty – national core curriculum 2020. *TED* **19**, 1–146; https://doi.org/10.25282/ted.716873 (2020).
7. Hattie, J. & Timperley, H. The power of feedback. *Rev. Educ. Res.* **77**, 81–112; https://doi.org/10.3102/003465430298487 (2007).
8. van de Ridder, J. M., Stokking, K. M., McGaghie, W. C. & ten Cate, O. T. What is feedback in clinical education? *Med. Educ.* **42**, 189–197; https://doi.org/10.1111/j.1365-2923.2007.02973.x (2008).
9. Archer, J. C. State of the science in health professional education: Effective feedback. *Med. Educ.* **44**, 101–108; https://doi.org/10.1111/j.1365-2923.2009.03546.x (2010).

ARTICLE IN PRESS

10. Henderson, M., Molloy, E., Ajjawi, R. & Boud, D. Designing feedback for impact. In *The Impact of Feedback in Higher Education* (eds Henderson, M., Ajjawi, R., Boud, D. & Molloy, E.) (Palgrave Macmillan, Cham, 2019); https://doi.org/10.1007/978-3-030-25112-3_15.

11. Topping, K. Peer assessment between students in colleges and universities. *Rev. Educ. Res.* **68**, 249–276; https://doi.org/10.3102/00346543068003249 (1998).

12. Boud, D. *Enhancing Learning through Self-Assessment* (Routledge, 2013).

13. Falchikov, N. Peer feedback marking: Developing peer assessment. *Innov. Educ. Train. Int.* **32**, 175–187; https://doi.org/10.1080/1355800950320212 (1995).

14. Field, M., Burke, J. M., McAllister, D. & Lloyd, D. M. Peer-assisted learning: A novel approach to clinical skills learning for medical students. *Med. Educ.* **41**, 411–418; https://doi.org/10.1111/j.1365-2929.2007.02713.x (2007).

15. Burgess, A., McGregor, D. & Mellis, C. Medical students as peer tutors: a systematic review. *BMC Med. Educ.* **14**, 115; https://doi.org/10.1186/1472-6920-14-115 (2014).

16. Lerchenfeldt, S. & Taylor, T. A. Best practices in peer assessment: training tomorrow's physicians to obtain and provide quality feedback. *Adv. Med. Educ. Pract.* 571–578; https://doi.org/10.2147/AMEP.S250761 (2020).

17. Lockspeiser, T. M. et al. Understanding the experience of being taught by peers: the value of social and cognitive congruence. *Adv. Health Sci. Educ.* **13**, 361–372; https://doi.org/10.1007/s10459-006-9049-8 (2008).

18. Schön, D. A. *The Reflective Practitioner: How Professionals Think in Action* (Routledge, 2017); https://doi.org/10.4324/9781315237473.

ARTICLE IN PRESS

ARTICLE IN PRESS

19. Hatton, N. & Smith, D. Reflection in teacher education: Towards definition and implementation. *Teach. Teach. Educ.* **11**, 33–49; https://doi.org/10.1016/0742-051X(94)00012-U (1995).

20. Moon, J. *Reflection in Learning and Professional Development: Theory and Practice* (Routledge, 2013); https://doi.org/10.4324/9780203822296.

21. Mamede, S. & Schmidt, H. G. The structure of reflective practice in medicine. *Med. Educ.* **38**, 1302–1308; https://doi.org/10.1111/j.1365-2929.2004.01917.x (2004).

22. Mann, K., Gordon, J. & MacLeod, A. Reflection and reflective practice in health professions education: A systematic review. *Adv. Health Sci. Educ.* **14**, 595–621; https://doi.org/10.1007/s10459-007-9090-2 (2009).

23. Wald, H. S., Borkan, J. M., Taylor, J. S., Anthony, D. & Reis, S. P. Fostering and evaluating reflective capacity in medical education: developing the REFLECT rubric for assessing reflective writing. *Acad. Med.* **87**, 41–50; https://doi.org/10.1097/ACM.0b013e31823b55fa (2012).

24. Croskerry, P., Singhal, G. & Mamede, S. Cognitive debiasing 1: Origins of bias and theory of debiasing. *BMJ Qual. Saf.* **22**, ii58–ii64; https://doi.org/10.1136/bmjqs-2012-001712 (2013).

25. Arthur, W. Jr., Bennett, W. Jr., Stanush, P. L. & McNelly, T. L. Factors that influence skill decay and retention: A quantitative review and analysis. *Hum. Perform.* **11**, 57–101; https://doi.org/10.1207/s15327043hup1101_3 (1998).

26. Bjork, R. A. Institutional impediments to effective training. In *Learning, Remembering, Believing: Enhancing Human Performance* (eds Druckman, D. & Bjork, R. A.) 295–306 (National Academy Press, Washington, DC, 1994).

ARTICLE IN PRESS

ARTICLE IN PRESS

27. Butler, A. C. & Raley, N. D. The future of medical education: Assessing the impact of interventions on long-term retention and clinical care. *J. Grad. Med. Educ.* **7**, 483–485 (2015).

28. Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T. & Rohrer, D. Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychol. Bull.* **132**, 354–380; https://doi.org/10.1037/0033-2909.132.3.354 (2006).

29. Roediger, H. L. & Butler, A. C. The critical role of retrieval practice in long-term retention. *Trends Cogn. Sci.* **15**, 20–27; https://doi.org/10.1016/j.tics.2010.09.003 (2011).

30. Custers, E. J. Long-term retention of basic science knowledge: A review study. *Adv. Health Sci. Educ.* **15**, 109–128; https://doi.org/10.1007/s10459-008-9101-y (2010).

31. Custers, E. J. & Ten Cate, O. T. Very long-term retention of basic science knowledge in doctors after graduation. *Med. Educ.* **45**, 422–430; https://doi.org/10.1111/j.1365-2923.2010.03889.x (2011).

32. Ling, Y. U., Swanson, D. B., Holtzman, K. & Bucak, S. D. Retention of basic science information by senior medical students. *Acad. Med.* **83**, S82–S85; https://doi.org/10.1097/ACM.0b013e318183e2fc (2008).

33. McGee, J., Zhao, Y. & Patel, K. The impact of digital clinical skills training on medical student performance: A systematic review. *BMC Med. Educ.* **24**, 64; https://doi.org/10.1186/s12909-024-06471-2 (2024).

34. Mitchell, O., Cotton, N., Leedham-Green, K., Elias, S. & Bartholomew, B. Video-assisted reflection: Improving OSCE feedback. *Clin. Teach.* **18**, 409–416; https://doi.org/10.1111/tct.13354 (2021).

ARTICLE IN PRESS

ARTICLE IN PRESS

35. Makrides, A. & Yeates, P. Memory, credibility and insight: How video-based feedback promotes deeper reflection and learning in objective structured clinical exams. *Med. Teach.* **44**, 664–671; https://doi.org/10.1080/0142159X.2021.2020232 (2022).

36. Vaughn, C. J. et al. Peer video review and feedback improve performance in basic surgical skills. *Am. J. Surg.* **211**, 355–360; https://doi.org/10.1016/j.amjsurg.2015.08.034 (2016).

37. Zhang, H., Liao, A. W. X., Goh, S. H. L., Yoong, S. Q., Lim, A. X. M. & Wang, W. Effectiveness and quality of peer video feedback in health professions education: A systematic review. *Nurse Educ. Today* **109**, 105203; https://doi.org/10.1016/j.nedt.2021.105203 (2022).

38. Aukes, L. C., Geertsma, J., Cohen-Schotanus, J., Zwierstra, R. P. & Slaets, J. P. J. The development of a scale to measure personal reflection in medical practice and education. *Med. Teach.* **29**, 177–182; https://doi.org/10.1080/01421590701299272 (2007).

39. Şenol, Y., Daloğlu, M., Taşdelen, G. & Demirören, M. Groningen Refleksiyon Beceri Ölçeği: Ölçek Uyarlama Çalışması. *Akd. Med. J.* **5**, 282–289; https://doi.org/10.17954/amj.2019.1374 (2019).

40. Fetters, M. D., Curry, L. A. & Creswell, J. W. Achieving integration in mixed methods designs—Principles and practices. *Health Serv. Res.* **48**, 2134–2156; https://doi.org/10.1111/1475-6773.12117 (2013).

41. Guetterman, T. C., Fetters, M. D. & Creswell, J. W. Integrating quantitative and qualitative results in health science mixed methods research through joint displays. *Ann. Fam. Med.* **13**, 554–561; https://doi.org/10.1370/afm.1865 (2015).

42. Bandura, A. *Social Foundations of Thought and Action* (Englewood Cliffs, NJ, 1986).

ARTICLE IN PRESS

43. Vygotsky, L. S. *Mind in Society: The Development of Higher Psychological Processes* (Harvard University Press, 1978).

44. Zimmerman, B. J. Becoming a self-regulated learner: An overview. *Theory Pract.* **41**, 64–70; https://doi.org/10.1207/s15430421tip4102_2 (2002).

45. Carless, D. & Boud, D. The development of student feedback literacy: Enabling uptake of feedback. *Assess. Eval. High. Educ.* **43**, 1315–1325; https://doi.org/10.1080/02602938.2018.1463354 (2018).

46. Telio, S., Ajjawi, R. & Regehr, G. The "educational alliance" as a framework for reconceptualizing feedback in medical education. *Acad. Med.* **90**, 609–614; https://doi.org/10.1097/ACM.0000000000000560 (2015).

47. Morgado, M., Botelho, J., Machado, V., Mendes, J. J., Adesope, O. & Proença, L. Video-based approaches in health education: A systematic review and meta-analysis. *Sci. Rep.* **14**, 23651; https://doi.org/10.1038/s41598-024-73671-7 (2024).

48. Al-Jundi, W., Elsharif, M., Anderson, M., Chan, P., Beard, J. & Nawaz, S. A randomized controlled trial to compare e-feedback versus "standard" face-to-face verbal feedback to improve the acquisition of procedural skill. *J. Surg. Educ.* **74**, 390–397; https://doi.org/10.1016/j.jsurg.2016.11.011 (2017).

49. Zaghal, A. et al. Face-to-face versus distance learning of basic suturing skills in novice learners: A quantitative prospective randomized trial. *BMC Med. Educ.* **22**, 290; https://doi.org/10.1186/s12909-022-03353-3 (2022).

50. Bjork, R. A. & Bjork, E. L. Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society* (eds Gernsbacher, M. A., Pew, R. W., Hough, L. M. & Pomerantz, J. R.) 56–64 (Worth, 2011).

ARTICLE IN PRESS

ARTICLE IN PRESS

51. Martinengo, L. et al. Spaced digital education for health professionals: Systematic review and meta-analysis. *J. Med. Internet Res.* **26**, e57760; https://doi.org/10.2196/57760 (2024).

52. Boud, D., Keogh, R. & Walker, D. (eds) *Reflection: Turning Experience into Learning* (Kogan Page, 1985).

53. Moon, J. A. *A Handbook of Reflective and Experiential Learning: Theory and Practice* (RoutledgeFalmer, 2004).

54. Uygur, J. et al. A Best Evidence in Medical Education systematic review to determine the most effective teaching methods that develop reflection in medical students: BEME Guide No. 51. *Med. Teach.* **41**, 3–16; https://doi.org/10.1080/0142159X.2018.1505037 (2019).

55. Lim, J. Y. et al. A systematic scoping review of reflective writing in medical education. *BMC Med. Educ.* **23**, 12; https://doi.org/10.1186/s12909-022-03924-4 (2023).

56. Franco, R. S., Franco, C. A. G. D. S., Severo, M., Ferreira, M. A. & Karnieli-Miller, O. Reflective writing in the teaching of communication skills for medical students—A systematic review. *Patient Educ. Couns.* **105**, 1842–1851; https://doi.org/10.1016/j.pec.2022.01.003 (2022).

57. Hopewell, S. et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. *Lancet* **405**, 1633–1640; https://doi.org/10.1016/S0140-6736(25)00672-5 (2025).

58. Campbell, D. T. & Stanley, J. C. *Experimental and Quasi-Experimental Designs for Research* (Ravenio Books, 2015).

ARTICLE IN PRESS

ARTICLE IN PRESS

59. Anderson-Cook, C. M. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. *J. Am. Stat. Assoc.* **100**, 708; https://doi.org/10.1198/jasa.2005.s22 (2005).

60. Roediger, H. L. III & Karpicke, J. D. Test-enhanced learning: Taking memory tests improves long-term retention. *Psychol. Sci.* **17**, 249–255; https://doi.org/10.1111/j.1467-9280.2006.01693.x (2006).

61. Patrício, M., Julião, M., Fareleira, F., Young, M., Norman, G. & Vaz Carneiro, A. A comprehensive checklist for reporting the use of OSCEs. *Med. Teach.* **31**, 112–124; https://doi.org/10.1080/01421590802578277 (2009).

62. Creswell, J. W. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* 4th edn (SAGE Publications, 2014).

63. Altman, D. G. & Bland, J. M. Statistics notes: Treatment allocation in controlled trials: Why randomise? *BMJ* **318**, 1209; https://doi.org/10.1136/bmj.318.7192.1209 (1999).

64. Torgerson, D. J. & Roberts, C. Understanding controlled trials: Randomisation methods—Concealment. *BMJ* **319**, 375–376; https://doi.org/10.1136/bmj.319.7206.375 (1999).

65. Kernan, W. N., Viscoli, C. M., Makuch, R. W., Brass, L. M. & Horwitz, R. I. Stratified randomization for clinical trials. *J. Clin. Epidemiol.* **52**, 19–26; https://doi.org/10.1016/S0895-4356(98)00138-3 (1999).

66. Moher, D. et al. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ* **340**, c869; https://doi.org/10.1136/bmj.c869 (2010).

ARTICLE IN PRESS

ARTICLE IN PRESS

67. Schulz, K. F. & Grimes, D. A. Allocation concealment in randomised trials: Defending against deciphering. *Lancet* **359**, 614–618; https://doi.org/10.1016/S0140-6736(02)07750-4 (2002).

68. Gwet, K. L. *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters* 4th edn (Advanced Analytics, LLC, 2014).

69. Boud, D. Using journal writing to enhance reflective practice. *New Dir. Adult Contin. Educ.* **90**, 9–18; https://doi.org/10.1002/ace.16 (2001).

70. Chaffey, L. J., de Leeuw, E. J. & Finnigan, G. A. Facilitating students' reflective practice in a medical course: Literature review. *Educ. Health (Abingdon)* **25**, 198–203; https://doi.org/10.4103/1357-6283.109787 (2012).

71. Dyment, J. E. & O'Connell, T. S. Assessing the quality of reflection in student journals: A review of the research. *Teach. High. Educ.* **16**, 81–97; https://doi.org/10.1080/13562517.2010.507308 (2011).

72. Polit, D. F. & Beck, C. T. The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Res. Nurs. Health* **29**, 489–497; https://doi.org/10.1002/nur.20147 (2006).

73. Polit, D. F., Beck, C. T. & Owen, S. V. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res. Nurs. Health* **30**, 459–467; https://doi.org/10.1002/nur.20199 (2007).

74. Yurdal, M. O. & Kıncal, R. Y. Development of a digital peer-feedback tool for clinical skills training: a pilot study with second-year medical students. *Sci. Rep.* **15**, 40807; https://doi.org/10.1038/s41598-025-24650-z (2025).

75. Gibbs, G. *Learning by Doing: A Guide to Teaching and Learning Methods* (Further Education Unit, 1988).

ARTICLE IN PRESS

76. Gibbons, J. *Nonparametric Statistical Inference* 5th edn (Chapman & Hall, 2010).

77. McKnight, P. E. & Najab, J. Kruskal–Wallis test. In *The Corsini Encyclopedia of Psychology* 4th edn (eds Weiner, I. B. & Craighead, W. E.) (Wiley, 2010).

78. Field, A. *Discovering Statistics Using IBM SPSS Statistics* 5th edn (SAGE Publications, 2018).

79. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* 2nd edn (Lawrence Erlbaum Associates, 1988).

80. Creswell, J. W. & Plano Clark, V. L. *Designing and Conducting Mixed Methods Research* 4th edn (SAGE Publications, 2023).

81. Landis, J. R. & Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174; https://doi.org/10.2307/2529310 (1977).

ARTICLE IN PRESS

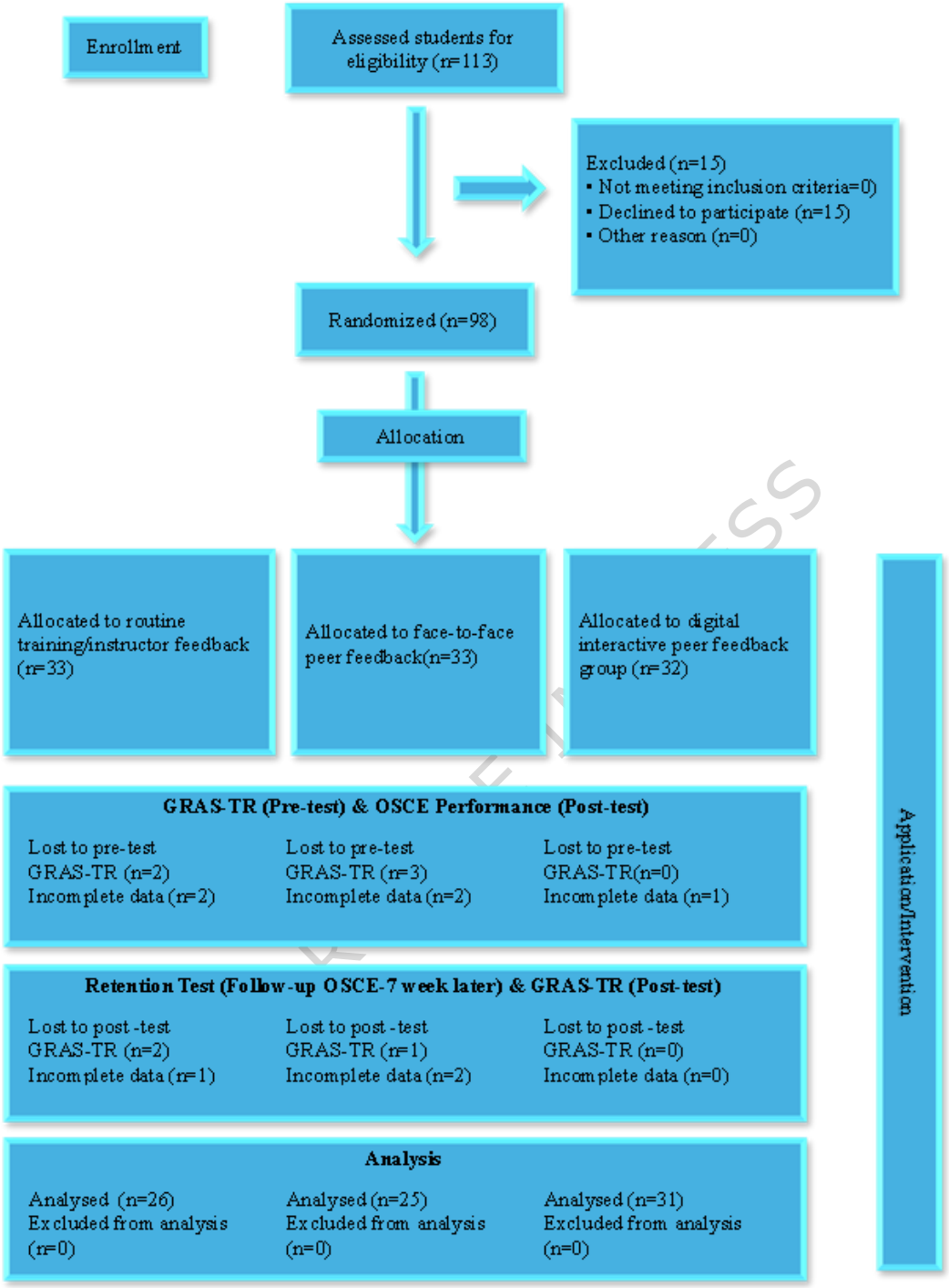


Figure 1. CONSORT flow diagram of participant enrolment, randomisation, follow-up, and analysis.

ARTICLE IN PRESS

CONSORT flow diagram of participant enrolment, randomisation, follow-up, and analysis. Of 113 students assessed for eligibility, 98 consented and were randomised to routine training/instructor feedback (n=33), face-to-face peer support (n=33), or digital interactive peer support (n=32); complete outcome data were available for 82 participants (control n=26, face-to-face n=25, digital n=31), who were included in the primary analyses.

ARTICLE IN PRESS

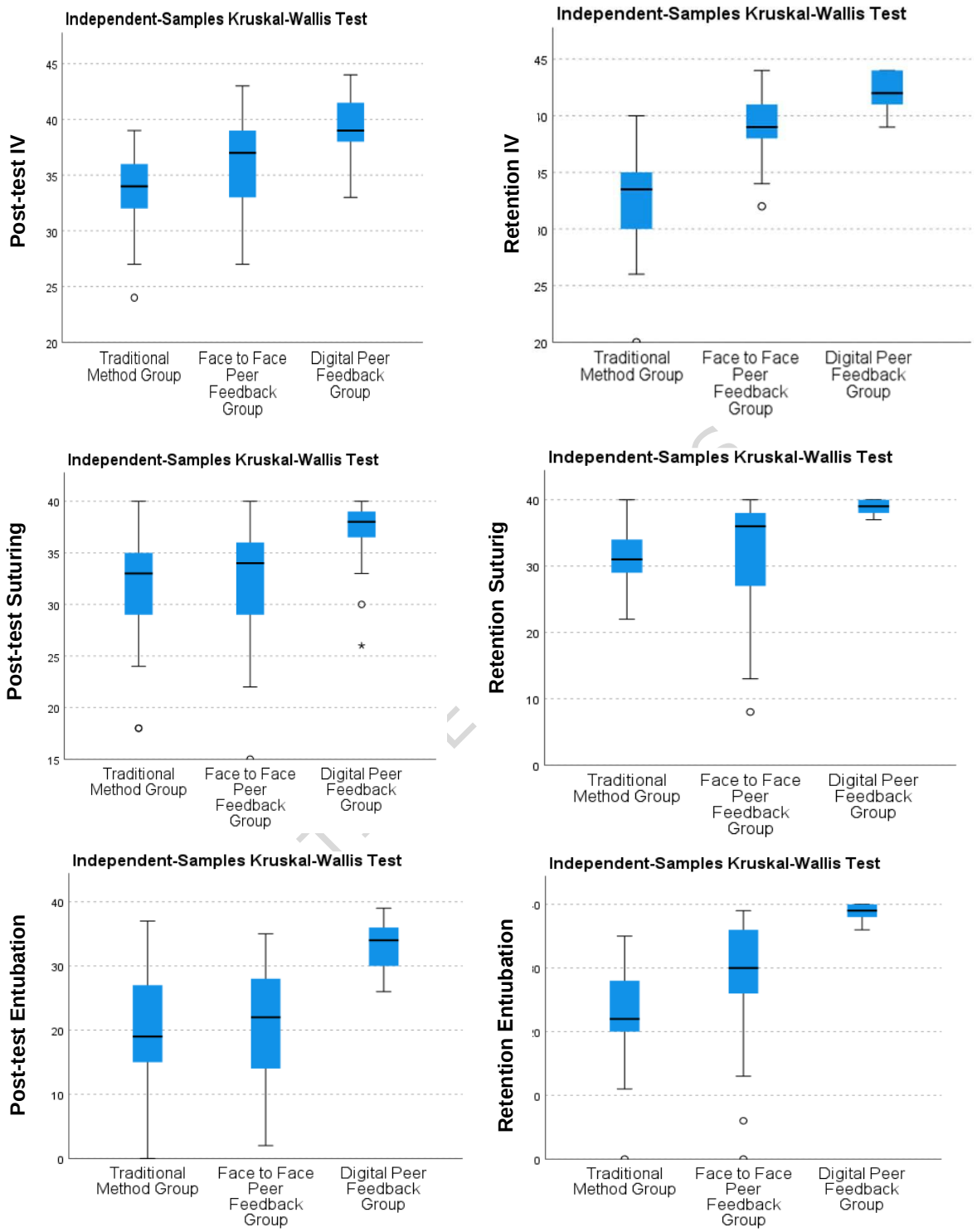
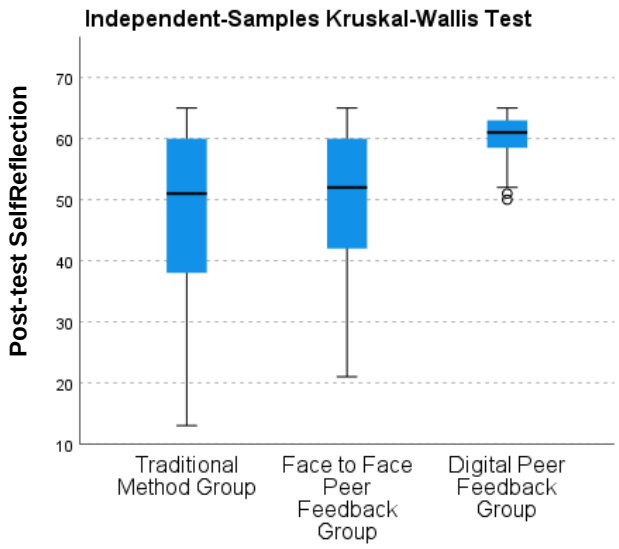


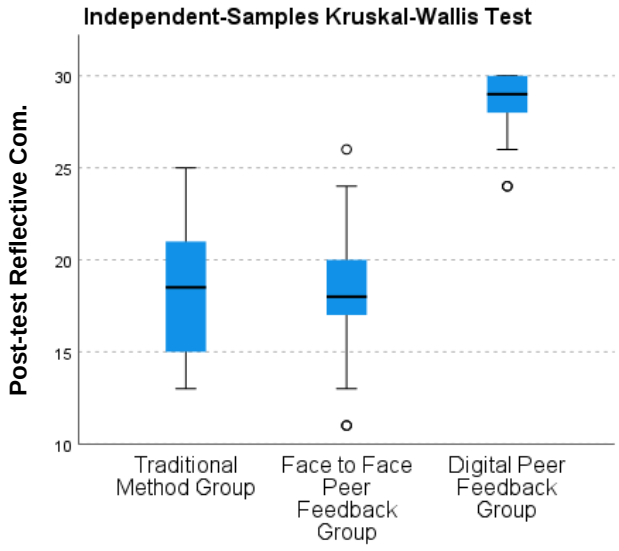
Figure 2. OSCE performance and retention by study arm across three clinical skills.

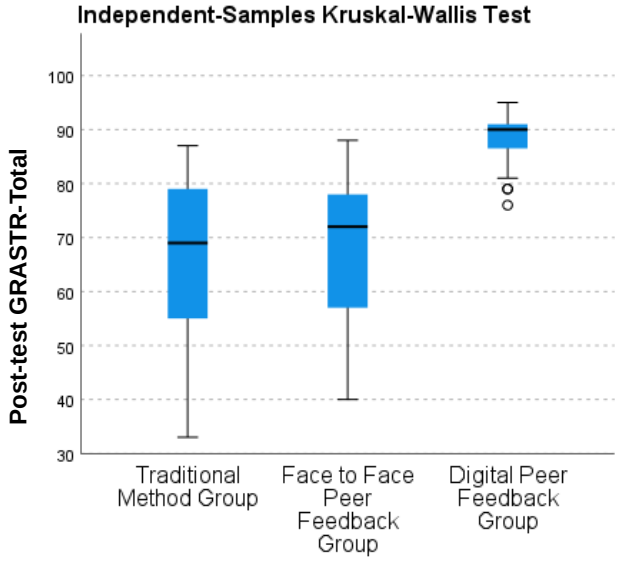
ARTICLE IN PRESS

OSCE performance and retention scores by study arm across three clinical skills. Boxplots display post-test and 7-week retention scores for intravenous injection (0–44), suturing/knot-tying (0–40), and endotracheal intubation (0–40) in the routine training/instructor feedback (control), face-to-face peer support, and digital interactive peer support arms; boxes represent the interquartile range (IQR), horizontal lines indicate the median, whiskers indicate dispersion, and points denote outliers.

ARTICLE IN PRESS


Independent-Samples Kruskal-Wallis Test
Post-test SelfReflection
70
60
50
40
30
20
10
Traditional Method Group
Face to Face Peer Feedback Group
Digital Peer Feedback Group


Independent-Samples Kruskal-Wallis Test
Post-test Reflective Com.
30
25
20
15
10
Traditional Method Group
Face to Face Peer Feedback Group
Digital Peer Feedback Group


Independent-Samples Kruskal-Wallis Test
Post-test GRASTR-Total
100
90
80
70
60
50
40
30
Traditional Method Group
Face to Face Peer Feedback Group
Digital Peer Feedback Group

ARTICLE IN PRESS

Figure 3. Reflective ability outcomes (GRAS-TR) at follow-up by study arm.

Reflective ability outcomes (GRAS-TR) at 7-week follow-up by study arm. Boxplots show self-reflection, reflective communication, and total GRAS-TR scores across the routine training/instructor feedback, face-to-face peer support, and digital interactive peer support arms; boxes represent the IQR, horizontal lines indicate the median, whiskers indicate dispersion, and points denote outliers.



Figure 4. Thematic network derived from reflective journals.

Thematic network derived from qualitative content analysis of reflective journals. The network takes a star-shaped form in which Educational Gain is the central node directly linked to five peripheral themes — Specialized Feedback, Self-Efficacy and Confidence, Learning Strategy and Sustainability, Emotional Climate and Motivation, and Technical-Structural Barriers and Facilitators — indicating Educational Gain as the primary linkage point across the thematic structure.

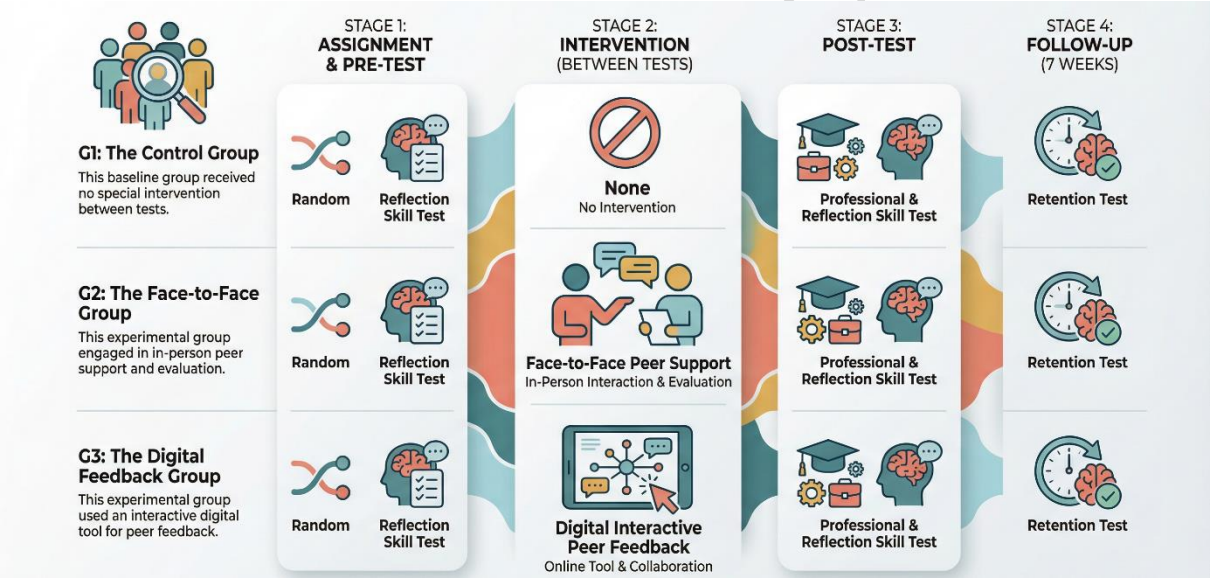


Figure 5. Interventions delivered to the intervention and control groups.

Overview of interventions delivered to the three study arms. All groups received the same standardised instructor demonstration and practised the same target clinical skills; the principal between-group difference was feedback modality — routine instructor feedback (control), real-time verbal peer feedback (face-to-face), or asynchronous video-based peer feedback via the Medifeeding platform (digital).

ARTICLE IN PRESS



Figure 6. Developed&implemented video-based digital peer feedback tool login screen.

ARTICLE IN PRESS

Login screen of the Medifeeding video-based digital peer feedback platform developed and implemented for this study. The platform was designed to allow students to upload practice videos, exchange structured peer feedback using embedded forms, and interact through multimodal features including audio comments, text comments, emoji responses, and video comments, accessible via personal user accounts on iOS and Android devices.

ARTICLE IN PRESS

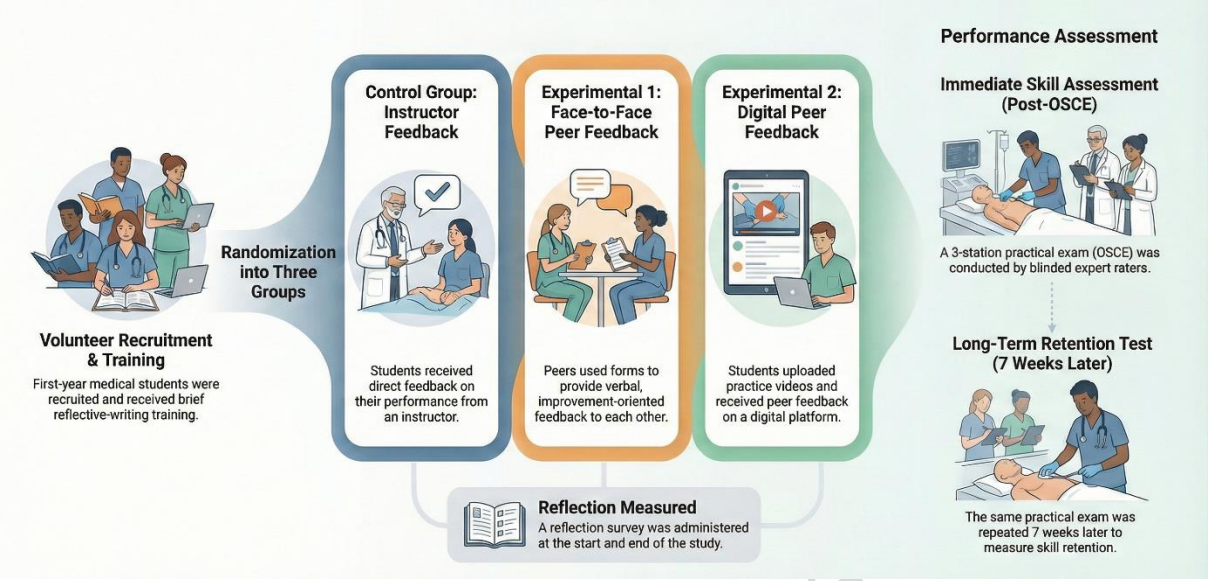


Figure 7. Anatomy of the intervention processs and data collection procedures.

Timeline and structure of the intervention process and data collection procedures. The study comprised a kick-off meeting (baseline GRAS-TR administration and randomisation), three intervention sessions per group (intravenous injection, endotracheal intubation, suturing/knot-tying), a post-intervention OSCE conducted 3 days after the final session, and a 7-week retention OSCE with follow-up GRAS-TR administration and reflective journal collection.

Table 1. Descriptive statistics for OSCE performance (post-test) and retention (7-week follow-up) by study arm.

Outcome (score range)	Group (N)	Post-test Mean (SD)	Post-test Median (Min–Max)	7-week Mean (SD)	7-week Median (Min–Max)
IV injection (0–44)	Routine training / instructor feedback (26)	33.77 (3.46)	34 (24–39)	32.50 (4.38)	33.5 (20–40)
	Face-to-face peer support (25)	35.88 (4.19)	37 (27–43)	39.08 (2.89)	39 (32–44)
	Digital interactive peer support (31)	39.58 (3.03)	39 (33–44)	42.19 (1.68)	42 (39–44)
Suturing / knot-tying (0–40)	Routine training / instructor feedback (26)	31.58 (5.54)	33 (18–40)	31.19 (4.09)	31 (22–40)
	Face-to-face peer support (25)	32.64 (6.17)	34 (15–40)	32.04 (8.61)	36 (8–40)
	Digital interactive peer support (31)	37.13 (3.02)	38 (26–40)	39.06 (1.03)	39 (37–40)

ARTICLE IN PRESS

Endotracheal intubation (0–40)	Routine training / instructor feedback (26)	20.58 (9.29)	19 (0–37)	23.12 (7.19)	22 (0–35)
	Face-to-face peer support (25)	20.76 (8.87)	22 (2–35)	28.12 (10.22)	30 (0–39)
	Digital interactive peer support (31)	32.97 (3.63)	34 (26–39)	38.84 (1.32)	39 (35–40)

Descriptive statistics for OSCE performance and 7-week retention scores by study arm. Post-test and retention mean scores (SD) and medians (min–max) are presented for intravenous injection (0–44), suturing/knot-tying (0–40), and endotracheal intubation (0–40) across the routine training/instructor feedback (n=26), face-to-face peer support (n=25), and digital interactive peer support (n=31) groups.

Table 2. Between-group comparisons of OSCE outcomes and reflection (Kruskal–Wallis; Dunn–Bonferroni post-hoc).

Outcome	Control (n=26)	F2F (n=25)	Digital (n=31)	H (df=2)	p	Significant post-hoc (Dunn–Bonferroni)
IV injection – post-test	33.77 (3.46); 34 (24–39)	35.88 (4.19); 37 (27–43)	39.58 (3.03); 39 (33–44)	29.59	<.001	Control < F2F; Control < Digital; F2F < Digital
IV injection – retention (7 weeks)	32.50 (4.38); 33.5 (20–40)	39.08 (2.89); 39 (32–44)	42.19 (1.68); 42 (39–44)	54.91	<.001	Control < F2F; Control < Digital; F2F < Digital
Suturing/knot-tying – post-test	31.58 (5.54); 33 (18–40)	32.64 (6.17); 34 (15–40)	37.13 (3.02); 38 (26–40)	22.79	<.001	Control < Digital; F2F < Digital
Suturing/knot-tying – retention (7 weeks)	31.19 (4.09); 31 (22–40)	32.04 (8.61); 36 (8–40)	39.06 (1.03); 39 (37–40)	39.23	<.001	Control < Digital; F2F < Digital
Intubation – post-test	20.58 (9.29); 19 (0–37)	20.76 (8.87); 22 (2–35)	32.97 (3.63); 34 (26–39)	36.79	<.001	Control < Digital; F2F < Digital
Intubation – retention (7 weeks)	23.12 (7.19); 22 (0–35)	28.12 (10.22); 30 (0–39)	38.84 (1.32); 39 (35–40)	56.20	<.001	Control < F2F; Control < Digital; F2F < Digital
Self-reflection (GRAS-TR) – baseline	47.50 (13.49); 48 (13–65)	44.76 (14.63); 48 (13–64)	46.74 (14.09); 50 (13–65)	0.49	.785	—
Self-reflection (GRAS-TR) – follow-up	49.00 (13.32); 51 (13–65)	49.84 (11.49); 52 (21–65)	60.03 (3.99); 61 (50–65)	16.35	<.001	Control < Digital; F2F < Digital

ARTICLE IN PRESS

Reflective communication (GRAS-TR) – baseline	17.00 (3.25); 16.5 (13–24)	17.64 (2.98); 18 (11–24)	18.10 (2.77); 18 (14–26)	2.07	.356	—
Reflective communication (GRAS-TR) – follow-up	18.15 (3.15); 18.5 (13–25)	18.40 (3.69); 18 (11–26)	28.35 (1.62); 29 (24–30)	56.68	<.001	Control < Digital; F2F < Digital
GRAS-TR total – baseline	64.50 (13.79); 67.5 (27–87)	62.40 (16.15); 68 (27–84)	64.84 (15.20); 68 (29–86)	0.30	.859	—
GRAS-TR total – follow-up	67.15 (14.29); 69 (33–87)	68.24 (12.69); 72 (40–88)	88.39 (4.53); 90 (76–95)	48.81	<.001	Control < Digital; F2F < Digital

Groups: Control = routine training/instructor feedback (n=26); F2F = face-to-face peer support (n=25); Digital = digital interactive peer support (n=31). Values are mean (SD); median (min–max). Reflection was assessed at baseline and follow-up (administered alongside the follow-up/retention assessment).

Between-group comparisons of OSCE performance, retention, and reflective ability outcomes. Kruskal–Wallis H statistics, p values, and significant pairwise comparisons (Dunn–Bonferroni post-hoc) are reported for all OSCE outcomes at post-test and 7-week retention, and for GRAS-TR subscale and total scores at baseline and follow-up; Control = routine training/instructor feedback (n=26), F2F = face-to-face peer support (n=25), Digital = digital interactive peer support (n=31); values are mean (SD); median (min–max).

Table 3. Mapping of themes to categories and related codes derived from reflective journals.

Theme	Categories	Related codes
Educational gain (core theme)	Skill improvement, error correction, skill speed, transfer	Error awareness; corrective action step; skill acceleration; step-by-step learning; transfer to clinical practice
Specialized feedback	Visual, millimetric, verbal detail	Visual demonstration; millimetric guidance; verbal detail
Self-efficacy & confidence	Belief, reduced anxiety, validation	Mastery belief; reduced anxiety; confidence via peer/confirmatory feedback
Learning strategy & sustainability	Repetition, self-testing, teaching peers, learning from peers	Repetition plan; self-testing; teaching peers; learning from peers
Emotional climate & motivation	Positive affect, challenge, stressor	Positive affect; challenge motivation; stressor awareness
Technical-Structural Barriers and Facilitators	Speed, interface, access to materials	Access speed; interface usability; access to materials/infrastructure

Thematic structure derived from qualitative content analysis of reflective journals. Six themes, their constituent categories, and related codes are presented as identified through two-cycle content analysis of 41 reflective journals, with Educational Gain functioning as the core theme linking the remaining five peripheral themes.

Table 4. Joint display integrating quantitative and qualitative findings across research questions.

ARTICLE IN PRESS

Outcome / Research question	Key quantitative findings (Kruskal–Wallis; post-hoc) (n=82)	Qualitative corroboration (themes; exemplar quotations) (n=41)	Integration (convergence / complementarity / dissonance)	Meta-inference (what the integrated evidence suggests)
RQ1. Clinical skill performance (OSCE post-test: IV injection, suturing/knot-tying, intubation)	Digital peer support outperformed both face-to-face and routine training/instructor feedback practice across all three skills (all p=.001). Face-to-face exceeded routine training/instructor feedback practice only for IV injection.	Educational gain was linked to more specific feedback and error awareness (e.g., "2 mm deviation… very valuable" [P5]); students emphasized replayable, multimodal cues and confirmatory signals (e.g., "green confirmation").	Convergence	Persistently available, specific, and confirmatory feedback (with replay/visual cues) likely promoted rapid error detection and correction, translating into higher OSCE performance.
RQ2. Retention of clinical skills (7-week OSCE re-test)	Digital peer support showed the highest retention across skills (all p=.001); for intubation, the pattern was routine training/instructor feedback < face-to-face < digital.	Retention was described through learning strategy & sustainability (planned repetition, self-testing, peer practice) and was facilitated by persistent records; constraints were described under technical–structural barriers (e.g., slow loading, limited filtering).	Complementarity	Retention gains appear supported by structured self-regulation (plan–practice–monitor–adjust) enabled by persistent resources; technical friction may attenuate engagement for some learners.
RQ3. Reflection skills (GRAS-TR: self-reflection, reflective communication; pre–post)	Groups were comparable at baseline. At post-test, digital peer support scored higher than both comparators on self-reflection, reflective communication, and total reflection (all p=.001).	Themes highlighted self-efficacy & confidence and motivational climate supported by specific, verifiable feedback (e.g., "my hand no longer trembles… after watching the correct angle" [P3]).	Convergence	A psychologically safe, confirmatory feedback environment may reduce anxiety and strengthen self-efficacy, encouraging deeper self-assessment and reflective communication.
RQ4. Overall pattern across outcomes (integrated interpretation)	Across outcomes, the digital arm showed consistent advantages (performance, retention, reflection), with the largest separation observed in retention and reflective communication.	Thematic map positioned educational gain as the central theme fed by specific feedback, self-efficacy, and sustainability; technical and moderation concerns were recurrent boundary conditions.	Meta-inference generation	Taken together, the intervention's benefit profile is consistent with a 'feedback → self-regulation → sustained practice' pathway; implementation quality (technical performance, moderation) is likely critical for scalability.

ARTICLE IN PRESS

ARTICLE IN PRESS

Joint display integrating quantitative and qualitative findings across the four research questions. For each research question, key quantitative findings (Kruskal–Wallis; n=82), qualitative corroboration from reflective journal themes and exemplar quotations (n=41), the type of integration (convergence, complementarity, or dissonance), and the resulting meta-inference are presented side by side.

Table 5. Participant sex distribution and allocation to study arms

	Consented & Randomized	Control (Routine training / instructor feedback)	Face-to-Face Peer Support	Digital Interactive Peer Support	Qualitative Subset*	Quantitative Subset†
Female, n (%)	65 (66.3)	22 (66.7)	22 (66.7)	21 (65.6)	22 (53.7)	55 (67.1)
Male, n (%)	33 (33.7)	11 (33.3)	11 (33.3)	11 (34.4)	19 (46.3)	27 (32.9)
Total, n	**98**	**33**	**33**	**32**	**41**	**82**

*Qualitative subset refers to participants who provided qualitative data (e.g., reflective journals).

† Quantitative subset refers to participants who provided quantitative study data (e.g., questionnaire-based measures).

Participant sex distribution and allocation across study arms and analytical subsets. Female and male counts and percentages are shown for the full randomized sample (n=98) and for each study arm — routine training/instructor feedback (n=33), face-to-face peer support (n=33), digital interactive peer support (n=32) — as well as for the qualitative (n=41) and quantitative (n=82) analytical subsets; *qualitative subset = participants who submitted at least one reflective journal; †quantitative subset = participants with complete outcome data.

Table 6. Timing of intervention sessions by group and clinical skill (2025).

Group	Intravenous injection	Endotracheal intubation	Suturing / knot-tying
Control (routine training/instructor feedback)	03 Feb 2025	11 Feb 2025	21 Feb 2025
Experimental Group 1 (face-to-face peer feedback)	03 Feb 2025	12 Feb 2025	21 Feb 2025
Experimental Group 2 (digital peer feedback)	05 Feb 2025	14 Feb 2025	21 Feb 2025

ARTICLE IN PRESS

Dates of intervention sessions for each study arm and clinical skill. Training sessions for intravenous injection, endotracheal intubation, and suturing/knot-tying were delivered across three separate dates in February 2025, with the control group (routine training/instructor feedback) and experimental groups (face-to-face and digital peer support) attending on partially overlapping but distinct schedules.

ARTICLE IN PRESS