Assessing the Impact of Different Feedback Mechanisms on Suture Skills Acquisition: A Mixed Methods Study Using Phenomenological and Quantitative Analysis

Nuno Silva Gonçalves, MD,[†,§] *Carlos Collares, MD, PhD,*[†,‖,¶,‡,§§] *and José Miguel Pêgo, MD, PhD*[†,§]

[†]Life and Health Sciences Research Institute (ICVS), University of Minho, Braga, Portugal; [§]ICVS/3B's, PT Government Associate Laboratory, Braga, Portugal; [‖]Medical Education Unit, Faculty of Medicine and Biomedical Sciences, University of Algarve, Faro, Portugal; [¶]European Board of Medical Assessors, Cardiff, United Kingdom; [‡]Inspirali Educação, São Paulo, Brazil; and [§§]Faculdades Pequeno Príncipe, Curitiba, Brazil

BACKGROUND: Even though suture skills are recognized as important for medical graduates, they are often underdeveloped due to reduced surgical exposure, limited feedback opportunities, and constraints on teaching resources. While various feedback strategies exist to support skill acquisition, comparative evidence on their effectiveness from the learner's perspective is limited.

OBJECTIVE: To assess the impact of 3 feedback modalities: video review alone, video review with structured self-assessment, and video review with expert feedback, on the acquisition of basic suture skills among medical students, using both quantitative and qualitative methods.

METHODS: Sixty-eight students were randomly assigned to 3 feedback groups (A: video only, B: video + structured self-assessment, and C: video + expert feedback). Each performed a basic suture task (Part A), received group-specific feedback, and repeated the task (Part B). Performances were video-recorded and rated by blinded assessors using a 15-item checklist and a global score. Additionally, thirteen participants from group C completed semi-structured interviews on their feedback experience.

RESULTS: The mean global score rose from 3.24 to 3.52 in Group A, 3.46 to 3.63 in Group B, and 3.51 to 3.76 in Group C. ANOVA showed no significant differences between groups ($F(2,65) = 0.669$, $p = 0.516$, $\eta p^2 = 0.020$). Qualitative findings indicated that expert feedback, particularly when combined with video review, significantly enhanced motivation, confidence, and clarity. Group C participants described shifting from "fear of performing sutures" to "confidence," attributing this to the personalized, emotionally supportive feedback.

CONCLUSION: Although performance outcomes were similar across feedback types, expert feedback offered unique emotional and cognitive benefits. These findings support the integration of structured, learner-centered feedback in surgical training, combining scalability with pedagogical value. (J Surg Ed 83:103779. © 2025 The Authors. Published by Elsevier Inc. on behalf of Association of Program Directors in Surgery. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/))

KEY WORDS: surgery, feedback, qualitative analysis, suture, competence, medical education

COMPETENCIES: Technical and procedural skills, Acquisition and assessment of basic surgical suturing, Feedback literacy, Reflective practice, Development of self-assessment, Communication, Professional growth in early surgical training

INTRODUCTION

Ensuring that medical students acquire essential procedural competencies remains a persistent challenge, especially as curricular hours in classrooms expand and

Correspondence: Inquiries to Nuno Silva Gonçalves, MD, School of Medicine, University of Minho, Campus Gualtar, 4710-057 Braga, Portugal; e-mail: nunogoncalves@med.uminho.pt

 © 2025 The Authors. Published by Elsevier Inc. on behalf of Association of Program Directors in Surgery. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/) 1931-7204/$30.00 https://doi.org/10.1016/j.jsurg.2025.103779

hands-on clinical exposure diminishes.[1,2] Among these competencies, the ability to perform surgical suturing holds critical importance, not only as a technical requirement for future surgical trainees but also as a core clinical skill linked to patient safety and outcomes. Nevertheless, studies continue to show that many medical graduates lack proficiency in suturing, often due to limited access to structured practice opportunities and meaningful feedback during training.[3,4] As surgical educators seek effective, scalable ways to support skills development, the question of how best to provide feedback, both in terms of method and educational impact, remains a crucial, yet underexplored, issue.

Suture skills are one of the fundamental competencies that medical students must acquire, due to its core importance in several medical technical procedures and treatments, directly impacting patient safety and clinical outcomes.[3,5,6] However, available reports still state that these skills are underperformed by medical graduates, either by reduced surgical rotations,[6] reduced tutors, [3,7] or limited time for training.[8,9] Tallentire et al.[8] reported that 73% of UK medical graduates felt inadequately prepared to perform basic sutures, while Emmanuel et al.[3] found that only 32% of medical curricula include structured suture training.[10–12] Multiple and diverse training programs have been created, using dry labs, wet labs, hands-on, lectures, video, or computer-based sessions, with different conclusions on which 1 would be more efficient in teaching.[3,9,11,13,14] Regardless of the training methods used, 1 critical factor consistently associated with skill acquisition is the quality of feedback provided to learners. Feedback has long been recognized as a critical component of learning processes, facilitating reflection, correction of errors, and consolidation of good practices[10,15–18]; however, despite the known positive influence of feedback during training, it is not routinely used.[3,7,18]

Various feedback modalities have been employed in medical education to promote learning and skill development. Self-assessment, for instance, encourages learners to engage in reflection, fostering self-directed learning and critical appraisal of one's own performance.[19] However, studies have shown that self-assessment can be inaccurate, especially among less experienced learners, who may either overestimate or underestimate their abilities.[20,21] Video-based self-review has emerged as another strategy, allowing learners to visually appraise their actions, identify errors, and internalize correct techniques.[22] Although video review can enhance self-awareness, its effectiveness is often limited without guided interpretation or structured feedback.[22,23] Conversely, expert feedback, delivered by experienced instructors, tends to be highly effective in improving performance by offering tailored, specific advice.[10,24] This demands significant faculty time and institutional resources, making it less scalable in certain educational settings.[25]

We identified 3 main limitations in the literature: (1) studies focus on only 1 feedback modality; (2) they assess only quantitative outcomes; and (3) they do not explore learners' subjective experiences. Therefore, we designed this study to simultaneously compare 3 feedback modalities using a mixed-methods approach, while controlling for practice time, task complexity, and participants' prior experience.

This study was designed to address these gaps by evaluating the impact of 3 distinct feedback strategies: video self-review, video self-review with structured self-assessment, and video review with expert feedback, on the acquisition of basic suture skills among 5th-year medical students. By incorporating both quantitative performance measures and qualitative analysis of students' experiences, this study aims to provide a comprehensive evaluation of the effectiveness and perceived value of each feedback strategy, informing the design of more effective and resource-efficient training programs in medical education.

METHODS

Study Design

This is an interventional study designed to evaluate the impact of different feedback mechanisms on the acquisition of suture skills among 5th-year medical students.

A mixed-methods approach was employed, combining quantitative assessments of skill improvement with qualitative analysis of participants' experiences with received feedback.

Participants sample size and selection

We calculated the sample size using GPower 3.1 for a repeated-measures ANOVA with a time-by-group interaction.[26] Assuming a medium effect size ($f = 0.25$), a correlation between measures of 0.5, 80% power, and an alpha of 0.05, a total of 42 participants (14 per group) would be required. To account for potential dropouts and increase statistical power, a convenience sample of 68 volunteer fifth-year medical students from Minho University was recruited to participate. All participants provided informed consent prior to inclusion in the study. There were no students with prior advanced suture training or formal surgical experience, so participants exclusion was not required.

Task

Participants were asked to individually perform a basic interrupted suture task for a duration of 6 minutes, while

being video recorded. This initial attempt (Part A) served as a baseline for subsequent performance comparison. After the initial recording, participants were randomly assigned into one of 3 groups:

- Group A (Control with Video Feedback): Participants reviewed their own video-recorded performance and then repeated the same suture task for another 6 minutes (Part B).
- Group B (Video + Scale Visualization): Participants reviewed their video-recorded performance and completed a self-assessment using a previously validated Suture Assessment Scale (Minho-SAS). Following this, they repeated the suture task for 6 minutes (Part B).
- Group C (Video + Expert Feedback): Participants reviewed their recorded performance and received individualized verbal feedback from an experienced surgical facilitator. The facilitators involved had feedback delivery experience and are experts on surgical suture. Additionally, the feedback was structured and based on the Suture Assessment Scale. Participants then repeated the task for 6 minutes (Part B).

The expert feedback followed a structured 15-minute protocol divided into 3 phases: (1) Rapport Phase (2 minutes): the facilitator created a safe environment by asking about prior experience and anxieties; (2) Joint Review Phase (10 minutes): the video was reviewed with the facilitator pausing at key moments identified by the Minho-SAS, always starting with correctly performed aspects before addressing areas for improvement, using specific descriptive language such as, "notice how your needle grip is at two-thirds of its length; ideally, it should be closer to the center;" (3) Consolidation Phase (3 minutes): the facilitator summarized 3 main points and checked the participant's understanding. Three surgeons with at least 5 years of teaching experience took part in a calibration workshop, practicing the protocol with 10 pilot videos until achieving over 90% agreement in identifying and prioritizing feedback points, as measured by a fidelity-to-protocol checklist.

All video recordings from both attempts (Parts A and B) were anonymized and stored in a secure database.

Assessment

All video-recorded performances were independently assessed by 3 experienced surgeons who were blinded to group allocation and assessment timing (first or second attempt). Assessments were conducted using the Minho-SAS.

We assessed inter-rater agreement on a random sample of 20 videos (29% of the total) evaluated by all 3 raters. For the global score, the ICC (2,1) was 0.87 (95% CI, 0.78-0.93), indicating excellent reliability. For individual checklist items, Fleiss' kappa ranged from 0.72 to 0.94, with a mean of 0.83. The items with the lowest agreement were "appropriate knot tension" ($\kappa = 0.72$) and "uniform spacing between stitches" ($\kappa = 0.75$), reflecting the subjective nature of these aspects.

Qualitative Analysis

Following the second suture attempt, participants from Group C were invited to take part in a semi-structured interview conducted by a separate facilitator uninvolved in the feedback session. The interviews explored participants' perceptions of the feedback they received. The interview guide was developed based on principles from the feedback literature, particularly Archer[17] and van de Ridder et al.,[18] which emphasize that effective feedback should clarify performance gaps, be specific, timely, and delivered in a supportive climate. Accordingly, our questions addressed perceived usefulness, specificity, timing, confidence, and emotional response to the feedback.

To ensure content validity, the guide was refined through consensus among 3 authors with expertise in medical education and feedback research. It was then piloted with 2 senior medical students to check clarity and coverage of the intended domains, with minor adjustments made following their input.

The interview guide included the following questions:

- How did you feel about the feedback you received from the facilitator?
- In what ways did the feedback help you improve your performance?
- How did the feedback impact on your confidence moving forward in similar tasks?
- Was the feedback specific enough to help you understand what needed improvement?
- Was the timing of the feedback appropriate for your needs?
- Is there anything about the feedback process you would change to make it more effective?
- What type of feedback (written, verbal, video review, other) do you think would have been most useful?

All interviews were audio-recorded, transcribed verbatim, and anonymized for analysis.

A phenomenological approach was used to explore the lived experience of receiving expert feedback. This approach was based on the work developed by Collaizi,[27] including the following steps:

1. Listening the recorded interviews, read and re-read the transcripts.

2. Identification and clustering of the common themes emerging from the different recordings
3. Extract relevant statements relevant to the phenomena.
4. Synthetizing the findings into common themes and reporting.

Quantitative Analysis

Quantitative data were analyzed using JASP (JASP Team (2023). JASP (Version 0.18.1)).

Each participant completed the suture task twice (Part A and Part B) and was evaluated independently by 3 blinded assessors using a 15-item binary checklist (Yes/No) and a global performance rating on a 5-point scale. For each evaluation, the number of affirmative responses ("Yes") was summed to create a checklist score, representing task performance on a 0−15 scale. To ensure consistency and reduce inter-rater variability, the mean checklist score and mean global score were calculated for each participant in both parts of the task. These aggregated values were used for all subsequent analyses. Additionally, 2 delta variables were computed: 1 reflecting the change in checklist score (ΔSumScore = Part B − Part A) and the other representing the change in global performance score (ΔGlobalScore = Part B − Part A). This allowed both within-group improvement and between-group comparisons to be assessed.

Descriptive statistics were used to characterize the sample. Paired samples t-tests were employed to evaluate within-group differences in performance scores before and after the intervention. We conducted a 3×2 mixed ANOVA with group (A, B, C) as the between-subjects factor and time (Part A, Part B) as the within-subjects factor. Linear regression analysis was performed to identify potential predictors of improvement, including group allocation and baseline performance.

Materials and Setup

We used a standardized kit consisting of 3-0 silk suture with a 3/8 circle curved needle, a 16 cm Mayo-Hegar needle holder, a 12 cm toothed Adson forceps, and a 14 cm Metzenbaum scissors. The synthetic skin model featured a premarked 5 cm linear incision with a depth of 0.8 cm. Recordings were made with an HD camera positioned 40 cm above the workspace at a 45-degree angle. Each participant received 3 suture threads, allowing for approximately 8−10 stitches within the allocated time. Total participation time was 30 minutes for Groups A and B (6 min Part A + 15 min intervention + 6 min Part B + 3 min procedures) and 45 minutes for Group C (with an additional 15 minutes for the interview).

Ethics Commission Approval

The study was approved by the Ethics Committee for Research in Life and Health Sciences (CEICVS 072/2025). All participants provided written informed consent, and data confidentiality was maintained throughout the study.

RESULTS

A total of 68 fifth-year medical students participated in the study, divided randomly into 3 groups. The final group attribution was of 21 students for group A (Control with Video Feedback), 22 students for group B (Video + Scale Visualization), and 25 students for group C (Video + Expert Feedback). Of the 25 students in Group C, 13 were invited to participate in a formal interview for the qualitative analysis. This number was sufficient to reach thematic saturation, consistent with phenomenological research standards, while allowing for interviews to be conducted in a standardized and feasible format.

Quantitative Analysis

This analysis aimed to assess the impact of different feedback strategies on the acquisition of suture skills across the 3 study groups. As previously stated, performance scores from part A and subsequent part B suture tasks were compared within each group to determine individual improvement and between groups to evaluate the relative effectiveness of each feedback strategy.

Before the analyses, the normality of the difference scores (Part B − Part A) was assessed using the Shapiro-Wilk test, separately for each group.

For the global performance score, results indicated that the data were not normally distributed in Group A ($p = 0.002$), whereas Groups B and C showed no significant deviation from normality ($p = 0.159$ and $p = 0.130$, respectively). Similarly, for the checklist score, Group A again failed the normality assumption ($p = 0.042$), while Groups B and C did not ($p = 0.311$ and $p = 0.113$, respectively).

Based on these findings, nonparametric Wilcoxon signed-rank tests were conducted to evaluate pre−post differences in Group A, while parametric paired samples t-tests were used for Groups 1 and 2.

Across all 3 groups, an overall increase in both global performance scores and checklist-based scores was observed between Part A and Part B of the suture task. As shown in Table 1, the mean global performance score improved in Group A from 3.24 (SD = 0.71) to 3.52 (SD = 0.71), in Group B from 3.46 (SD = 0.53) to 3.63 (SD = 0.57), and in Group C from 3.51 (SD = 0.58) to 3.76 (SD = 0.59). Similarly, checklist scores increased from 9.25 to 9.55 in Group A, from 9.61 to 9.99 in Group B, and from 9.60 to 10.21 in Group C.

As for the within-group analyses (Table 2), in Group A, the Wilcoxon signed-rank test revealed an improvement

TABLE 1. Descriptive Statistics by Group and Time Point

Group	N	Mean Global Score A	Mean Global Score B	Δ Global Score	Mean Checklist Score A	Mean Checklist Score B	Δ Checklist Score
A – Video only	21	3.24 ± 0.71	3.52 ± 0.71	0.28 ± 0.26	9.25 ± 1.95	9.55 ± 2.26	0.31 ± 0.83
B – Self-assessment	22	3.46 ± 0.53	3.63 ± 0.57	0.16 ± 0.37	9.61 ± 1.17	9.99 ± 1.33	0.39 ± 0.93
C – Expert feedback	25	3.51 ± 0.58	3.76 ± 0.59	0.26 ± 0.42	9.60 ± 1.56	10.21 ± 1.35	0.61 ± 1.16

TABLE 2. Within-Group Comparisons (Part A vs. Part B)

Group	Outcome	Test	Statistic	df	p-value	Effect Size
A	Global score	Wilcoxon	$z = -3.30$	—	0.001	$r = -1.00$
A	Checklist score	Wilcoxon	$z = -1.89$	—	0.061	$r = -0.51$
B	Global score	Paired t-test	$t(21) = -2.08$	21	0.050	$d = -0.44$
B	Checklist score	Paired t-test	$t(21) = -1.95$	21	0.065	$d = -0.42$
C	Global score	Paired t-test	$t(24) = -3.08$	24	0.005	$d = -0.62$
C	Checklist score	Paired t-test	$t(24) = -2.62$	24	0.015	$d = -0.53$

TABLE 3. Between-Group Comparisons of Improvement Scores

Outcome	Test	Statistic	df	p-Value	Effect Size
Δ global score	ANOVA	$F(2, 65) = 0.669$	2, 65	.516	$\eta^2 = 0.020$
Δ global score	Kruskal-Wallis	$H(2) = 1.178$	2	.555	—
Δ checklist score	ANOVA	$F(2, 65) = 0.576$	2, 65	.565	$\eta^2 = 0.017$
Δ checklist score	Kruskal-Wallis	$H(2) = 1.061$	2	.588	—

in global score ($z = -3.30$, $p = 0.001$, $r = -1.00$), while the improvement in checklist score approached but did not reach significance ($z = -1.89$, $p = 0.061$). In Group B, paired t-tests showed an improvement in global score ($t(21) = -2.08$, $p = 0.050$, $d = -0.44$), whereas the change in checklist score was not significant ($t(21) = -1.95$, $p = 0.065$). In Group C, statistically significant improvements were observed for both global score ($t(24) = -3.08$, $p = 0.005$, $d = -0.62$) and checklist score ($t(24) = -2.62$, $p = 0.015$, $d = -0.53$).

Between-group comparisons of improvement scores (Part B − Part A), summarized in Table 3, were conducted through a mixed ANOVA test. For the global score, there was a significant main effect of time ($F(1,65) = 12.34$, $p < 0.001$, $\eta p^2 = 0.159$), indicating overall performance improvement. No main effect of group was observed ($F(2,65) = 0.82$, $p = 0.445$, $\eta p^2 = 0.025$), nor a time × group interaction ($F(2,65) = 0.67$, $p = 0.516$, $\eta p^2 = 0.020$), suggesting similar improvement across groups. Planned contrast analyses confirmed significant improvement within all groups: Group A ($t(20) = 3.30$, $p = 0.003$, $d = 0.39$), Group B ($t(21) = 2.08$, $p = 0.050$, $d = 0.31$), and Group C ($t(24) = 3.08$, $p = 0.005$, $d = 0.43$).

Given the absence of normality distribution in several variables, nonparametric Kruskal-Wallis tests were also performed, which confirmed the absence of significant between-group differences for both global score ($H(2) = 1.178$, $p = 0.555$) and checklist score ($H(2) = 1.061$, $p = 0.588$).

Qualitative Analysis

To complement the quantitative evaluation, a qualitative analysis was conducted with participants from Group C, who received video-based feedback from an expert facilitator. The aim was to explore students' perceptions of the feedback process and its influence on their learning, confidence, and motivation.

After the transcription and thematic analysis, the authors aimed to identify recurring clusters that reflect how students experienced and internalized the feedback. The emerging themes provide insights into the emotional, cognitive, and practical impact of the

feedback intervention on skill acquisition. From this analysis, 5 clusters were extracted and are discussed below with relevant quotations from the participants.

We rigorously followed Colaizzi's steps for phenomenological analysis.[27] Two researchers independently transcribed and coded all 13 interviews. Initial inter-coder agreement was 78%, calculated using percentage agreement. Discrepancies were resolved through discussion until consensus was reached, with a third researcher acting as an arbitrator in 3 cases. An external audit was conducted by an experienced qualitative researcher, who reviewed 30% of the codings and the theme derivation. The 5 final themes were presented to 5 randomly selected participants (member checking), all of whom confirmed that the themes accurately reflected their experiences. A reflective journal was maintained to document analytical decisions and potential researcher biases.

Perceived Value and Utility

This theme emerged across all 13 interviews, with participants consistently identifying expert feedback as an essential catalyst for understanding technical nuances. P2 expressed this perception, stating, "only when we have these small moments of feedback do we understand the little things that need improvement," a sentiment echoed by P13, who valued having someone point out specific errors that would otherwise go unnoticed during self-observation. P7 highlighted the uniqueness of this experience within the educational context: "It is often not possible to have someone observing us in real time," underscoring how structured feedback fills a critical gap in traditional medical training. Together, these perspectives underscore how structured feedback fills a gap left by traditional training.

Participants felt that feedback helped them refine their technique, especially by drawing attention to details (e.g., tension on the knot, hand positioning) they would not have noticed independently. This enhanced their belief in feedback as a critical step in skill acquisition, not merely evaluation.

Emotional Impact: From Anxiety to Confidence

Participants described how initial anxiety or insecurity about their performance shifted into confidence and motivation following constructive feedback. P3 reflected, "In the beginning, I was very nervous… but the feedback was very calm and that helped me," while P13 simply stated, "I no longer feel nervous." P1 noted, "Now I know what I can do," and P7 emphasized, "I was afraid to do a suture in the hospital rotations, but now I'm confident."

Analysis revealed that the consistent use of positive reinforcement at the start of feedback sessions was a key driver of this transformation. Feedback that first acknowledged correct actions before addressing areas for improvement fostered psychological safety and optimism: "Feedback was very good because it focused on the positive parts first, and that did not put me down. That was very important." (P1) This sentiment was reiterated by P4, who appreciated "receiving compliments about minor goals that I achieved, which was very valuable," and by P5, who described "the peaceful and calm format of giving feedback, with positive reinforcement first in a tension-free environment" as highly enjoyable.

Importance of simultaneous video review with a facilitator as a feedback amplifier

Co-visualization with the facilitator was described as a transformative element of expert feedback. Students reported that their own observations, the facilitator's comments, and the video evidence converged to create a clearer understanding. As P4 stated, "watching simultaneously allowed me to receive feedback while seeing exactly the moment." Beyond timing, this process also provided cognitive alignment. P8 added, "there was room to share with the facilitator what I was feeling," highlighting how co-visualization created a dialogic space for testing self-perceptions against expert input.

The co-review of the video created bidirectional communication, allowing learners to ask questions and reflect in real time. Synchronizing visual and verbal cues improved understanding of errors, and the shared video helped bridge the gap between what students intended to do and what they performed.

Participants also highlighted that this co-review process reduced misunderstandings that can occur with verbal-only feedback, where there might be a disconnect between what the facilitator sees and what the learner remembers or perceives. This finding suggests that the value of expert feedback may lie as much in the interactive process as in the technical content conveyed.

Specificity and Individualization

Participants emphasized the value of feedback that was personalized and detailed, focusing on their technique rather than generic mistakes. This specificity enhanced both applicability and retention of learning. Students described feeling seen, heard, and coached rather than merely evaluated, which fostered trust in the learning process and in the facilitator. P9 noted, "it was objective, which led me to understand well what I needed to improve," while P6 highlighted, "I could ask all my doubts and questions; it was directed to me."

Feedback Timing and Retention

Participants consistently described the immediacy of feedback—delivered shortly after the first task—as

crucial for learning. Receiving feedback while the experience was still fresh helped them anchor key points and apply improvements effectively in the second round. The proximity of feedback to upcoming high-stakes assessments, such as the OSCE, further enhanced its relevance and appreciation, aligning with participants' emotional and cognitive readiness. P12 remarked, "if it was much later, I would not have recalled it," while P10 reflected, "since it was immediately after, I could see and remember all my changes in the second part," and P8 added, "I still remembered all my difficulties."

Interestingly, 1 participant expressed a preference for an even more immediate approach, stating, "in my case, I would like to receive immediate feedback during the suture" (P11).

DISCUSSION

This study employed a mixed-methods approach, combining a quantitative assessment of performance improvements with a qualitative exploration of students' experiences, which provided a comprehensive understanding of how feedback influences suture skill development and learners' perceptions with different feedback strategies.

Our quantitative findings, showing equivalent improvement across groups, appear to contradict the meta-analysis by Hatala et al.,[28] which consistently demonstrated the superiority of expert feedback. However, closer examination reveals important methodological differences. First, studies included in the meta-analysis employed longer training periods (a mean of 4.5 sessions versus our single session), suggesting that differences between feedback modalities may only emerge with extended practice. Second, our posthoc power analysis revealed only 23% power to detect the observed differences ($\eta p^2 = 0.020$), indicating that a sample size of 175 participants per group would be required to draw definitive conclusions about equivalence. Third, the basic interrupted suture task may represent a skill with a low "complexity ceiling," where initial improvements are easily achieved regardless of feedback modality. Porte et al.[10] found expert feedback to be superior when using a more complex task (end-to-end anastomosis), suggesting that task complexity may moderate the differential effectiveness of feedback modalities.

However, no significant differences were found between the groups regarding the degree of improvement. Several factors may explain this outcome. First, although the sample size was adequate for an exploratory study, it may have been underpowered to detect small inter-group differences. Additionally, the short interval between tasks does not allow for tracking long-term differences. The task used in this study—a basic suture—represents a foundational technique that, once initially understood, may not exhibit dramatic variation across attempts in such a brief training window. While participants in all groups improved, the small differences in delta scores between groups suggest that short-term exposure to any form of feedback may be sufficient to elicit modest gains. Future studies incorporating repeated training sessions or more complex procedures may be better suited to detect nuanced differences between feedback strategies.

The inclusion of a qualitative component using a phenomenological approach revealed relevant insights. By exploring the lived experiences of learners, we probed into emotional, cognitive, and motivational aspects that could be overlooked in performance-based quantitative assessments. The emotional impact of the feedback process was perceived by students as of particular importance. Several students reported transitioning from anxiety and uncertainty before the feedback to increased confidence and motivation afterward. This emotional trajectory was often linked to the tone and format of the feedback: positive reinforcement preceding constructive criticism, a calm and supportive environment, and the perception that facilitators were genuinely invested in their learning. Such findings align with educational theories that emphasize psychological safety and learner-centered feedback as essential to optimal performance and engagement.[29] Reinforcing what students did well before pointing out areas for improvement helps maintain self-esteem and openness to learning, a technique well-established in health professions education but not always consistently applied in practice.[17,25]

Another significant aspect identified by participants was the value of reviewing the performance video together with the expert facilitator. This co-viewing approach allowed for real-time clarification, alignment of perceptions, and meaningful dialogue between the learner and the teacher. Students emphasized that this synchronous format helped them better understand what needed improvement and why, avoiding miscommunication or misinterpretation that might occur with feedback provided without visual context. This supports the notion that video, when paired with expert guidance, serves not only as a reflective tool but as a shared cognitive framework that enhances engagement, precision, clarity, and retention of feedback. These findings align with prior research showing that verbal-only feedback can be misunderstood or forgotten if not specific and clearly linked to performance.[7,25] By synchronizing commentary with visual evidence, video-assisted feedback enhances clarity and retention helping to explain why our participants perceived co-review as particularly effective.[23]

The interviews also emphasized the importance of feedback being specific, individualized, and tailored to the learner's unique performance, not being generic, but focused on their strengths and weaknesses. This perceived personalization helped foster a sense of being coached rather than evaluated, encouraging greater trust and receptiveness to learning. In our framework, coaching is conceived as an iterative, learner-centered dialogue that guides individuals through progressive skill development, whereas evaluation represents a point-in-time, summative judgment against predefined competency standards. Coaching fosters ongoing reflection and autonomy by providing tailored feedback, goal-setting, and opportunities for self-assessment over time.[25,30] In contrast, evaluation delivers a definitive verdict on performance without necessarily supplying the developmental support learners need to bridge gaps.[17] The structure provided by the Minho-SAS may have also contributed to the clarity and utility of the expert feedback.

The current study, however, presents some limitations. First, the relatively small sample size may limit generalizability and statistical power, particularly for detecting subtle differences between groups. Second, despite being seen as positive by the students, the study design involved only a single repetition of the suture task; therefore, it is unclear whether the observed improvements are retained over time. Third, qualitative interviews were conducted only with participants in the expert feedback group, limiting the ability to compare subjective experiences across all feedback modalities.

Based on the integration of quantitative and qualitative findings, we propose a 3-tiered implementation model designed to optimize resources while maximizing pedagogical benefits.[16]

- Level 1 (Initiation): a mandatory initial session with expert feedback for all learners to establish technical standards and a mental framework for self-assessment (investment: 15 minutes of faculty time per student).
- Level 2 (Deliberate Practice): 3 to 5 sessions using video with structured self-assessment through the Minho-SAS, enabling autonomous yet guided practice (investment: initial development of recording infrastructure).
- Level 3 (Consolidation): a final expert feedback session for learners who have not achieved proficiency in self-assessment or before summative evaluations (investment: 15 minutes of faculty time per student, for approximately 30% of the cohort).

Compared to an ideal model of continuous expert feedback (rarely implemented in practice), this hybrid model would reduce faculty workload by approximately 65% while preserving the motivational benefits identified in expert feedback at critical learning moments. For a class of 100 students, the estimated faculty time required would be 35 hours, compared to 100 hours under the traditional model.

We recommend 5 priority directions for future research. First, longitudinal studies should track skill retention at 1, 3, and 6 months, analyzing forgetting curves across feedback modalities. Second, investigations using tasks of increasing complexity—such as continuous suturing, intradermal suturing, and anastomosis—are needed to determine whether task complexity moderates the differential effectiveness of feedback types. Third, formal economic analyses should compare the cost per competency achieved between modalities, including infrastructure expenses and faculty time. Fourth, translational studies should assess whether the motivational benefits of expert feedback lead to greater autonomous practice and improved OSCE performance. Finally, the development and validation of an artificial intelligence (AI) algorithm trained on a video database annotated with expert feedback could provide a scalable approach to replicating the identified benefits. AI-based systems could support the technical aspects of feedback, such as automated scoring of checklists, detection of common errors, or generation of preliminary performance reports. However, the affective and motivational dimensions of feedback rely on human interaction and cannot currently be replicated by AI. A blended approach, where AI handles routine technical feedback and educators focus on individualized guidance and emotional support, may represent a pragmatic balance between scalability and pedagogical effectiveness. Future studies should incorporate a priori power calculations using G*Power, based on realistic effect sizes from our findings ($\eta p^2 \approx 0.02-0.03$ for between-group differences).

CONCLUSION

This study demonstrates that all 3 feedback modalities investigated can facilitate measurable short-term improvement in basic suturing performance, with no statistically significant differences observed in the context of a single session. However, phenomenological analysis revealed that participants receiving expert feedback experienced unique cognitive and affective benefits, including greater clarity regarding specific errors, increased confidence, and the development of a mental framework for future self-assessment. Future studies should evaluate whether the cognitive and affective benefits identified in expert feedback translate to better skill retention and actual clinical performance, aspects not assessed in the present study. The integration of

quantitative and qualitative methodologies provided a more nuanced understanding of how different forms of feedback influence procedural skill acquisition, underscoring the importance of considering both objective outcomes and subjective experiences when designing educational interventions.

REFERENCES

1. Klumpp S., Rocha I., Amirfazli C., et al. Impact of early procedural exposure on pre-clinical medical students' Confidence. Cureus n.d.;17:e77411. https://doi.org/10.7759/cureus.77411
2. Saad S, Richmond C, King D, Jones C, Malau-Aduli B. The impact of pandemic disruptions on clinical skills learning for pre-clinical medical students: implications for future educational designs. *BMC Med Educ*. 2023;23:364. https://doi.org/10.1186/s12909-023-04351-9.
3. Emmanuel T, Nicolaides M, Theodoulou I, Yoong W, Lymperopoulos N, Sideris M. Suturing skills for medical students: a systematic review. *In Vivo*. 2021;35:1–12. https://doi.org/10.21873/invivo.12226.
4. Alam S, McGuinness E. 1124 Enhancing suturing skills in medical students and junior doctors. *BJS*. 2025;112. https://doi.org/10.1093/bjs/znaf128.641. znaf128.641.
5. Bitar ER, Hassanieh J, Rahhal S, Zaiter B, Zaghal A. Handedness in surgical education: evaluating suturing proficiency among left- and right-handed trainees. *Surg Open Sci*. 2025;24:51–57. https://doi.org/10.1016/j.sopen.2025.02.006.
6. Davis CR, Toll EC, Bates AS, Cole MD, Smith FCT. Surgical and procedural skills training at medical school: a national review. *Int J Surg*. 2014;12:877–882. https://doi.org/10.1016/j.ijsu.2014.05.069.
7. Sender Liberman A, Liberman M, Steinert Y, McLeod P, Meterissian S. Surgery residents and attending surgeons have different perceptions of feedback. *Med Teach*. 2005;27:470–472. https://doi.org/10.1080/01425590500129183.
8. Tallentire VR, Smith SE, Wylde K, Cameron HS. Are medical graduates ready to face the challenges of Foundation training? *Postgrad Med J*. 2011;87:590–595. https://doi.org/10.1136/pgmj.2010.115659.
9. Shaharan S, Neary P. Evaluation of surgical training in the era of simulation. *World J Gastrointest Endosc*. 2014;6:436–447. https://doi.org/10.4253/wjge.v6.i9.436.
10. Porte MC, Xeroulis G, Reznick RK, Dubrowski A. Verbal feedback from an expert is more effective than self-accessed feedback about motion efficiency in learning new surgical skills. *Am J Surg*. 2007;193:105–110. https://doi.org/10.1016/j.amjsurg.2006.03.016.
11. Bauer F, Rommel N, Kreutzer K, et al. A novel approach to teaching surgical skills to medical students using an ex vivo animal training model. *J Surg Educ*. 2014;71:459–465. https://doi.org/10.1016/j.jsurg.2014.01.017.
12. Brunt ML, Halpin VJ, Klingensmith ME, et al. Accelerated skills preparation and assessment for senior medical students entering surgical internship. *J Am Coll Surg*. 2008;206:897–904. https://doi.org/10.1016/j.jamcollsurg.2007.12.018.
13. Routt E, Mansouri Y, de Moll EH, Bernstein DM, Bernardo SG, Levitt J. Teaching the simple suture to medical students for long-term retention of skill. *JAMA Dermatol*. 2015;151:761–765. https://doi.org/10.1001/jamadermatol.2015.118.
14. Bennett SR, Morris SR, Mirza S. Medical students teaching Medical students surgical skills: the benefits of peer-assisted learning. *J Surg Educ*. 2018;75:1471–1474. https://doi.org/10.1016/j.jsurg.2018.03.011.
15. Burgess A, van Diggele C, Roberts C, Mellis C. Tips for teaching procedural skills. *BMC Med Educ*. 2020;20:458. https://doi.org/10.1186/s12909-020-02284-1.
16. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. *Psychol Rev*. 1993;100:363–406. https://doi.org/10.1037/0033-295X.100.3.363.
17. Archer JC. State of the science in health professional education: effective feedback. *Med Educ*. 2010;44:101–108. https://doi.org/10.1111/j.1365-2923.2009.03546.x.
18. van de Ridder JMM, Stokking KM, McGaghie WC, ten Cate OTJ. What is feedback in clinical education? *Med Educ*. 2008;42:189–197. https://doi.org/10.1111/j.1365-2923.2007.02973.x.
19. Yan Z, Carless D. Self-assessment is about more than self: the enabling role of feedback literacy. *Assess Eval Higher Educ*. 2022;47(8):1050–1061. https://doi.org/10.1080/02602938.2021.2001431.
20. Raaijmakers SF, Baars M, Paas F, van Merriënboer JJG, van Gog T. Effects of self-assessment feedback on self-assessment and task-selection accuracy. *Metacognition Learning*. 2019;14:21–42. https://doi.org/10.1007/s11409-019-09189-5.

21. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol*. 1999;77:1121–1134. https://doi.org/10.1037//0022-3514.77.6.1121.

22. Scaffidi MA, Walsh CM, Khan R, et al. Influence of video-based feedback on self-assessment accuracy of endoscopic skills: a randomized controlled trial. *Endosc Int Open*. 2019;7:E678–E684. https://doi.org/10.1055/a-0867-9626.

23. Brame CJ. Effective educational Videos: principles and guidelines for maximizing student learning from video content. *CBE Life Sci Educ*. 2016;15:es6. https://doi.org/10.1187/cbe.16-03-0125.

24. Tieken KR, Kelly G, Maxwell J, Visenio MR, Reynolds J, Fingeret AL. Feedback versus compliments versus both in suturing and knot tying simulation: a randomized controlled trial. *J Surg Res*. 2024; 294:99–105. https://doi.org/10.1016/j.jss.2023.09.037.

25. Burgess A, van Diggele C, Roberts C, Mellis C. Feedback in the clinical setting. *BMC Med Educ*. 2020;20:460. https://doi.org/10.1186/s12909-020-02280-5.

26. Faul F, Erdfelder E, Buchner A, Lang A-G. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav Res Meth*. 2009;41:1149–1160. https://doi.org/10.3758/BRM.41.4.1149.

27. Colaizzi PF. Psychological research as a phenomenologist views it. Valle RS, King M, editors. Existential-Phenomenological Alternatives for Psychology, New York: Oxford University Press; 1978:48–71.

28. Hatala R, Cook DA, Zendejas B, Hamstra SJ, Brydges R. Feedback for simulation-based procedural skills training: a meta-analysis and critical narrative synthesis. *Adv Health Sci Educ Theory Pract*. 2014;19:251–272. https://doi.org/10.1007/s10459-013-9462-8.

29. Newman A, Donohue R, Eva N. Psychological safety: a systematic review of the literature. *Human Res Manage Rev*. 2017;27:521–535. https://doi.org/10.1016/j.hrmr.2017.01.001.

30. Toh RQE, Koh KK, Lua JK, et al. The role of mentoring, supervision, coaching, teaching and instruction on professional identity formation: a systematic scoping review. *BMC Med Educ*. 2022;22:531. https://doi.org/10.1186/s12909-022-03589-z.