

Structuring Self-Regulated Learning in Surgical Training: A Pilot Study of Paired Forethought and Reflection Instruments in Laparoscopic Cholecystectomy

José Ignacio Rui-Wamba Barra

dr.ruiwambab@gmail.com

Pontificia Universidad Catolica de Chile <https://orcid.org/0009-0008-5253-1133>

Bernardita Becker

Pontificia Universidad Catolica de Chile

Cristóbal Vildósola

Pontificia Universidad Católica de Chile: Pontificia Universidad Catolica de Chile

Matias Aguilera

Pontificia Universidad Catolica de Chile

María Inés Gaete

Pontificia Universidad Católica de Chile: Pontificia Universidad Catolica de Chile

Lorena Isbej

Pontificia Universidad Catolica de Chile

Michelle Grunauer

Pontificia Universidad Católica de Chile: Pontificia Universidad Catolica de Chile

Gonzalo Urrejola

Pontificia Universidad Catolica de Chile

Gabriel Escalona

Pontificia Universidad Catolica de Chile

Julian Varas

Pontificia Universidad Católica de Chile: Pontificia Universidad Catolica de Chile

<https://orcid.org/0000-0002-5828-9623>

Research Article

Keywords:

Posted Date: July 1st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10036059/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Abstract

Purpose: Surgical residents operate on fewer cases than previous generations, and validated assessment tools such as OSATS, GOALS, and NOTSS remain external, observer-based instruments that capture a single moment of performance rather than pairing forethought with reflection on the same case. In routine laparoscopic training, learner-completed tools that pair preoperative forethought with postoperative self-reflection remain limited. We aimed to develop and pilot a theory-informed, paired pre- and postoperative metacognitive instrument architecture for laparoscopic cholecystectomy (LC) training and assess its feasibility and acceptability among residents. **Methods:** We conducted a single center, two phase pilot feasibility study reported following the Standards of Quality Improvement Reporting Excellence (SQUIRE 2.0). In Phase 1, an expert panel (n = 4) integrated self-regulated learning theory with design principles from validated assessment instruments to construct a six-item preoperative Yes/No checklist for forethought and an eight-item postoperative behaviourally anchored 5-point rubric for self-reflection, comprising five technical and three cognitive items. In Phase 2 (December 2025–March 2026), residents completed the paired instruments around consecutive routine LCs. Primary outcomes were instrument completion and resident-reported acceptability. **Results:** Five residents completed 14 preoperative checklists and 10 postoperative rubrics around routine LCs. In 10 of 14 initiated cases, the paired postoperative rubric was also completed, corresponding to 71% cycle completion (95% CI 45–88%). Among postoperative respondents, median ratings were 5/5 for all four acceptability items (IQRs 4–5 to 5–5). No interference with patient care was reported. The absence of a reliable registry of eligible cases precluded estimation of absolute adherence. **Conclusions:** A theory-informed, paired pre- and postoperative metacognitive architecture for LC can be completed by residents around routine cases without faculty time at the bedside and was rated favourably by postoperative respondents, although completion was modest and concentrated preoperatively. This pilot demonstrates that a complete self-regulation cycle can be operationalized in real operative workflow, but does not establish improved performance. It supports a controlled study with prospective case capture, baseline metacognitive assessment, and expert calibration to evaluate effects on learning, self-assessment accuracy, and operative transfer. **Keywords:** Surgical education; Metacognition; Self-regulated learning; Laparoscopic cholecystectomy; Feasibility study; Internship and residency.

Introduction

The conditions under which surgical residents train have changed over the past two decades. Duty-hour regulations, increasing subspecialization, competition for cases, and changes in supervisory practice have reduced residents' operative exposure, and structured approaches such as deliberate practice are increasingly required to derive educational value from the training opportunities that remain [1]. In Chile, Laparoscopic cholecystectomy (LC) is a high-volume laparoscopic procedure frequently used as a main operation for laparoscopic training [2]. It exemplifies this challenge: sufficiently frequent to constitute much of a resident's laparoscopic experience, yet sufficiently variable in difficulty that unstructured exposure may not reliably translate operative volume into competence.

Self-regulated learning (SRL) offers a useful lens for this gap. SRL is best understood not as a single framework but as an overarching process, an umbrella encompassing the cognitive, metacognitive, motivational, and behavioural processes through which learners regulate their own learning; within this tradition, Zimmerman's cyclical model provides the specific framework on which we draw, describing learning as three interdependent phases: forethought before the task (task analysis, goal-setting, strategy selection, self-efficacy beliefs), performance during the task (self-monitoring, self-control), and self-reflection after the task (self-judgment, self-reaction) [3]. SRL has been repeatedly identified as a core competency for learning in clinical workplaces, where supervision is intermittent and learning opportunities are irregular [4, 5], and recent work shows that surgeons' mindset and motivation are associated with the SRL behaviors they deploy across the forethought, performance, and reflection phases of clinical practice [6]. Operating without explicit forethought or reflection may therefore leave an important part of the learning cycle unstructured.

Metacognition, the monitoring and regulation of one's own cognition, is a central component of SRL. Moulton and colleagues' work on the "slowing down when you should" phenomenon demonstrated that experts deliberately monitor their own cognition in real time and pause at safety-relevant moments: a behavior that is metacognitive, learnable, and measurable [7]. Schraw showed that metacognitive knowledge and regulation are domain-general and teachable through explicit instructional strategies [8, 9]; Lajoie showed that learning environments can scaffold self-regulation by prompting and cueing its component processes [10]; and the reflection literature in medical education converges on the same point: unguided, generic reflection is unreliable, whereas guided, structured reflection can be embedded in routine practice [11].

Existing instruments in the surgical-education literature each address one piece of this picture. OSATS provides validated, behaviourally anchored global ratings of technical skill [12, 13]; GOALS extends this approach to laparoscopic-specific psychomotor domains [14]; and the NOTSS taxonomy operationalizes the cognitive and non-technical components of surgical performance [15]. All three, however, are designed for external observer-based rating and capture a single moment rather than pairing forethought with reflection on the same case. The Metacognitive Awareness Inventory for Surgical Educators (MAISE) operationalizes metacognition within surgery, but for educators rather than learners [16]. The deliberate-practice tradition argues that goal-setting, focused attention to specific weaknesses, and structured reflection are prerequisites for expertise in medicine [17, 18], and experimental work shows that learning-goal orientations increase both task engagement and metacognition during surgical skill acquisition in novice learners [19]. Emerging work has explored rapid-cycle deliberate practice, asynchronous tele-mentoring, and self-directed video feedback for laparoscopic skill acquisition [20–22]. We are not aware of a published learner-completed architecture in routine LC training that pairs structured forethought before the case with structured reflection after it, thereby completing a self-regulation cycle around the same operative event.

We, therefore, designed a theory-informed, paired pre- and postoperative SRL architecture for LC training. Because a learner-completed tool is only useful if residents will actually complete it in real workflow, this

pilot tested its feasibility and acceptability as a necessary first step before any evaluation of learning, calibration, or transfer.

Methods

Study design and setting

A single-center, two-phase pilot feasibility study was conducted between December 2025 and March 2026. The study is reported in accordance with the Standards for Quality Improvement Reporting Excellence (SQUIRE 2.0) [23]. Phase 1 comprised instrument development through expert consensus and literature synthesis; Phase 2 comprised prospective implementation of the paired instruments during routine LCs performed by surgical residents. A pilot feasibility design was chosen deliberately: before any controlled evaluation of learning outcomes, it is necessary to determine whether residents can complete paired instruments around real cases without disrupting clinical workflow.

Phase 1: Instrument development

An expert panel of four faculty members with combined expertise in minimally invasive surgery, simulation-based education, and surgical education research developed the paired instrument. The panel met weekly throughout the development period and followed an explicit two-step process: (a) extraction of theoretical principles from the self-regulated learning and metacognition literature, and (b) adaptation of design principles from validated surgical assessment instruments. Candidate items were drafted, mapped to the constructs they were intended to scaffold, and revised iteratively across successive meetings, with all decisions reached by consensus. The resulting instruments are best understood as theory-informed reflective scaffolds produced through a preliminary, content-informed development process.

The theoretical layer drew on Zimmerman's three-phase SRL cycle [3], Schraw's distinction between knowledge of cognition and regulation of cognition with its instructional implications [8, 9], and Lajoie's analysis of how scaffolds embedded in the learning environment can prompt and cue self-regulatory processes [10]. From these, the panel derived four operational design principles:

- Pair forethought and reflection on the same case. The instruments are not free-standing; they are explicitly designed as the pre- and post-task halves of one Zimmerman cycle for a single LC [3].
- Use criteria-referenced, task-specific prompts. Following Eva and Regehr's reformulation of self-assessment as self-monitoring, items are anchored to LC-specific behaviours rather than to global judgments, because unguided global self-assessment is systematically inaccurate [24].
- Anchor every Likert rating with a behavioural descriptor. This decision was informed directly by OSATS, in which the behaviourally anchored global rating outperformed both unanchored scales and task checklists [12].

- Keep technical and cognitive domains separate. Following the GOALS and NOTSS tradition [14, 15], the postoperative rubric segregates technical execution from cognitive process, so that a resident scoring high on manual technique but low on situational awareness can read that signal directly.

Preoperative Forethought Checklist

A six-item Yes/No checklist was completed by residents immediately before scrubbing (Table 1). The closed-response format was a deliberate design by the panel to minimize completion burden in the immediate preoperative setting (target < 2 minutes), where time and attention are limited; conceptually, the checklist operates as a scaffold that prompts and cues the forethought processes of the SRL cycle [10]. By design, an affirmative response therefore indicates only that the resident was prompted to engage with a given forethought process; it is not intended to capture the quality or depth of that forethought. Each item maps to a Zimmerman Phase 1 component: activation of prior knowledge (Item 1), self-knowledge of personal weakness (Item 2), explicit strategic planning (Item 3), directed attention to recurrent technical deviations (Item 4), and case-specific goal setting (Item 5) [3, 17]. Item 5 additionally draws on error-management training, in which deliberately orienting the learner toward a specific anticipated error improves retention and transfer of procedural skills [25]. Item 6, pre-committed criteria for help-seeking or pausing, operationalizes Moulton's "slowing down" construct on the learner's side [7].

Table 1

Preoperative Forethought Checklist. Completed by the resident immediately before scrubbing for each case. Target completion time < 2 minutes.

#	Item (Yes / No)	SRL function (Zimmerman Phase 1)
1	Reviewed imaging / clinical context for this case	Activate prior knowledge; anticipate difficulty
2	Identified the step that is personally most difficult	Self-knowledge of personal weakness
3	Clear strategy for traction / counter-traction from the start	Explicit strategic planning
4	Aware of recurrent technical deviations (hand, axis, depth)	Directed attention to common pitfalls
5	Identified the single specific error to avoid today	Case-specific goal-setting
6	Clear criteria for when to call for help or pause	Pre-commitment to safety criteria

Postoperative Self-Reflection Rubric

An eight-item rubric, with each item rated on a 5-point Likert scale with behavioral anchors (1 = novice, 2 = advanced beginner, 3 = competent, 4 = proficient, 5 = expert), was completed by residents immediately after the case (Table 2). The rubric comprises two explicit domains. Domain A, Technical Execution,

includes five items: non-dominant hand use (A1), critical-step sequencing (A2), instrument alignment (A3), tissue handling (A4), and ergonomics (A5), adapted from the GOALS dimensions of bimanual dexterity, efficiency, and tissue handling [14]. Domain B, Cognitive Processes, includes three items: situational awareness (B1), anticipation of difficulties (B2), and intraoperative decision-making (B3), aligned with the NOTSS taxonomy [15]. The rubric was intended to structure the resident’s postoperative self-reflection rather than as a graded measure of performance. It is intentionally self-administered and no parallel expert rating was collected in this pilot.

Table 2

Table 2. **Postoperative Self-Reflection Rubric.** Each item rated on a 5-point Likert scale with behavioural anchors (1 = novice, 2 = advanced beginner, 3 = competent, 4 = proficient, 5 = expert). Completed by the resident immediately after the procedure.

Item	Domain	Description	Underlying construct
A1	Technical	Non-dominant hand use	Psychomotor self-monitoring
A2	Technical	Critical-step sequencing	Procedural organisation
A3	Technical	Instrument alignment	Fine motor control
A4	Technical	Tissue handling	Gentleness and control
A5	Technical	Ergonomics	Posture and efficiency
B1	Cognitive	Situational awareness	Real-time environment scanning
B2	Cognitive	Anticipation of difficulties	Forethought–performance link
B3	Cognitive	Intraoperative decision-making	Real-time self-regulation

Phase 2: Prospective implementation

All surgical residents performing LC at our institution during the study period were invited to complete the paired instruments around each consecutive eligible case. The workflow was: (1) before surgery, the resident reviewed the case and completed the six-item preoperative checklist (< 2 minutes); (2) the LC was performed according to standard institutional technique under faculty supervision; (3) within minutes of leaving the operating room, the resident completed the eight-item postoperative rubric; and (4) instrument completion and brief resident-feedback items on usability, perceived educational value, and willingness to continue using the instruments were captured. Both instruments were administered electronically through web-based forms; the acceptability items were embedded in the postoperative form. No reliable institutional registry of all eligible LC cases per resident was available during the study period.

Outcomes

Primary outcomes (feasibility; Kirkpatrick Level 1 [26]) were instrument completion and acceptability. We distinguish three related but distinct quantities. Completion refers to the absolute number of preoperative checklists and postoperative rubrics completed, reported as counts. Cycle completion, our principal feasibility metric, is defined as the proportion of initiated cases (those in which the preoperative checklist was completed) in which both the pre and postoperative instruments were completed, that is, the proportion of started cycles that were closed. Because no reliable registry of all eligible LC cases per resident was available, absolute adherence (completed instruments divided by all eligible cases) could not be estimated. Acceptability was defined as resident-reported usability, perceived educational value, and willingness to continue using the instruments, captured in the postoperative form.

Sample size

No formal sample-size calculation was performed, consistent with the descriptive aims of a feasibility pilot. The target enrolment was a convenience sample of residents and LC cases available during the four-month implementation period. The pilot was not powered to estimate feasibility metrics with narrow precision, and confidence intervals around these estimates are reported to make that imprecision explicit.

Statistical analysis

All analyses were descriptive. Completion counts are reported as absolute numbers; cycle completion is reported as a proportion with a 95% confidence interval (Wilson score method). Acceptability items were mapped to a 1–5 scale (1 = strongly disagree to 5 = strongly agree) and summarised as medians with interquartile ranges and as the proportion of responses at or above 4 (agree). No comparative inferential statistics were planned or performed, and no psychometric analysis of internal structure (e.g., internal consistency, factor analysis) was undertaken, as the pilot sample was too small to support stable estimates. Interpretation is restricted to Kirkpatrick Level 1 and is explicitly framed as preliminary.

Ethics

Ethical approval

for this study was obtained from the Comité Ético Científico of the Facultad de Medicina, Pontificia Universidad Católica de Chile (Protocol #250103001). Residents participated voluntarily, and all data were collected and analysed in anonymised, aggregate form. No patient-identifiable information was collected.

Results

Participants and instrument completion

Five surgical residents participated, completing 14 preoperative checklists and 10 postoperative rubrics around routine LCs. In 10 of the 14 cases initiated with a preoperative checklist (71%; 95% CI 45–88%), the paired postoperative rubric was also completed. No resident requested withdrawal during the study.

Preoperative checklist response patterns

Item-level affirmative responses to the preoperative checklist ranged from 9 to 14 of 14. Item 1 (review of imaging or clinical context) was affirmed in 14 of 14 and Item 6 (clear criteria for help-seeking or pausing) in 13 of 14. Item 2 (personally most difficult step) was affirmed in 10 of 14, Item 4 (awareness of recurrent technical deviations) in 11 of 14, and Item 3 (clear traction strategy) and Item 5 (a specific error to avoid in that case) in 9 of 14 each. The two least frequently affirmed items were Item 3 (clear traction strategy) and Item 5 (a specific error to avoid), each affirmed in 9 of 14; these correspond to the more demanding, case-specific components of forethought, in contrast to the consistently affirmed items reflecting routine case review (Item 1, 14/14) and safety thresholds (Item 6, 13/14).

Resident acceptability and perceived educational value

Acceptability items were captured only among the 10 postoperative completers and rated comparably for both instruments. Among these respondents, median ratings were 5/5 (IQR 4–5) for ease of sustained use (8 of 10 responses at agree or strongly agree), 5/5 (IQR 5–5) for clarity of instructions (10 of 10), 5/5 (IQR 4–5) for perceived usefulness for learning and surgical performance (10 of 10), and 5/5 (IQR 4–5) for ease of completion within the routine clinical workflow (10 of 10).

Unintended consequences

No adverse events related to educational intervention were reported. No resident reported that the instruments interfered with patient care, and no deviations from standard LC technique or supervision attributable to the instruments were observed.

Discussion

This pilot shows that a paired pre- and postoperative metacognitive instrument architecture for LC, grounded in SRL theory and informed by validated surgical assessment frameworks, can be completed by residents around routine cases without dedicated faculty time at the bedside. The most defensible feasibility signal, however, is not the overall completion rate but the asymmetry between the two components: residents completed more preoperative checklists than paired postoperative rubrics (14 versus 10), indicating that the preoperative forethought component was easier to embed in real workflow than the postoperative reflection component, and identifying the postoperative phase as the more fragile point. The full forethought reflection cycle was closed in 71% of initiated cases; a figure to be read cautiously given the small, self-selected sample; acceptability ratings among postoperative respondents clustered at the ceiling of the scale, and no resident withdrew; we nonetheless regard the pre/post asymmetry as more informative than this headline percentage. Consistent with this asymmetry, the item-level response pattern showed that residents arrived prepared and safety-aware, whereas deliberate,

case-specific components of forethought naming a personal weakness, defining a traction strategy, and identifying a specific error to avoid, were less consistently affirmed, consistent with the gap the architecture was designed to scaffold.

Our findings connect three strands of evidence. First, Moulton and colleagues' description of "slowing down when you should" established intraoperative metacognition as an observable expert behavior rather than an educational abstraction [7]; checklist Item 6 and rubric Item B1 operationalize this construct from the resident's perspective. Second, MAISE demonstrated that planning, monitoring, and reflection are recognizable dimensions of metacognition within surgery, although from the educator's perspective [16]; our architecture shifts that lens to the learner. Third, deliberate-practice theory emphasizes goal setting, focused attention to weaknesses, and structured reflection as prerequisites for expertise [17, 18], while experimental work shows that learning-goal orientations increase metacognition and engagement during surgical skill acquisition [19], and that error-focused training can improve retention and transfer of procedural skills [25]. Checklist Items 2 and 5 directly translate these principles into case-specific prompts.

Compared with extended-reality, eye-tracking, asynchronous tele-mentoring, self-directed video feedback, and structured non-technical-skills curricula [21, 22, 27–31], this intervention is intentionally minimalist. It requires no additional hardware, no simulator time outside clinical workflow, and no expert reviewer at the bedside. Rather than competing with these approaches, the paired instruments function as a self-administered metacognitive scaffold that could be layered onto existing LC curricula, simulation models, or workplace-based training. Visual-scanning training addresses situational awareness from the outside in, by training the perceptual front end [31]; the present rubric targets the same domain from the inside out, by prompting structured self-monitoring.

Clinical and educational interpretation

The architecture proposed here is best understood as a low-cost scaffolding layer intended to convert routine clinical exposure into structured deliberate practice. In a training environment where operative opportunities are constrained and deliberate practice informed approaches are increasingly emphasized [1], a tool that takes under two minutes before each case and a few minutes after it, without requiring faculty time at the bedside, has practical appeal. Whether this low burden translates into sustained use, however, is not established by this pilot. The pre/post pairing asks the same resident, on the same case, to make forethought explicit and then compare it with structured post-hoc self-assessment, operationalizing Zimmerman's self-regulation cycle around a real operative event [3].

The gap between acceptability and use

An apparent paradox deserves direct discussion: residents rated the instruments highly, yet the volume of completed instruments was modest and preoperative checklists outnumbered postoperative rubrics. Several explanations are plausible. The minutes before scrubbing are relatively protected and the preoperative checklist is brief, whereas the immediate postoperative period concentrates competing

obligations, including operative documentation, patient handover, ward tasks, and on-call responsibilities, while the more demanding eight-item rubric is due. Instrument design may also have contributed: the architecture was deployed as two free-standing instruments rather than as a single linked workflow, so completing the checklist did not automatically prompt completion of the postoperative rubric. An integrated, single-flow design in which the preoperative entry opens its paired postoperative component could reduce this gap. Finally, the discordance between high acceptability and modest use is consistent with an intention, behaviour gap, possibly amplified by social desirability toward the implementing team. Critically, these acceptability ratings were provided only by residents who completed the postoperative form; the subgroup least likely to have found the instruments burdensome; and therefore cannot be read as evidence that the architecture is sustainably implementable across all residents or across all eligible cases. Favourable perception among completers appears necessary but insufficient for sustained use; because uptake among non-completers and across the unknown denominator of eligible cases remains unmeasured, future studies should measure use objectively rather than infer it from acceptability ratings.

Methodological scope and limitations

The principal strengths of this study are its explicit theoretical grounding, the design of rubric items informed by validated assessment frameworks, the separation of technical and cognitive domains, the low burden of the instruments in real workflow, and the explicit framing of the pilot as a Kirkpatrick Level 1 evaluation. Several limitations define what this pilot can and cannot claim.

First, no reliable registry of all eligible LC cases per resident was available. Case logging was inconsistent and operative assignments shifted in ways the study team could not audit. Absolute adherence therefore could not be estimated, and completion-based metrics are vulnerable to self-selection: residents may have initiated instruments preferentially in cases they expected to complete. The number of LCs in which instruments were not initiated is unknown and may be substantial. A prospective, auditable registry of eligible cases is therefore essential for the subsequent controlled study.

Second, participants were residents at a stage in which LC is not invariably released to them as primary operator. In our program, residents progress through a supervised learning curve before being entrusted with critical steps of the procedure. The flow of eligible cases per resident was therefore irregular and partly outside their control, constraining instrument volume independently of motivation and further complicating the denominator.

Third, the sample was small, single-center, and recruited by convenience within one residency program. The cycle completion estimate carries a wide confidence interval (95% CI 45–88%) and may not generalize to programs with different case volumes, supervision models, or assessment cultures. This pilot should therefore be interpreted as testing whether the architecture is implementable at all, not as estimating how implementable it will be across settings.

Fourth, the study provides content-oriented validity evidence only, through expert panel development, theoretical mapping, and derivation from validated instruments. Evidence regarding response process, internal structure, relations to other variables, and consequences remains to be collected. The small sample precluded psychometric analysis of the rubric, and item-level self-ratings from the postoperative rubric are therefore reserved for the controlled study.

Fifth, the rubric was self-administered and was not paired with external expert rating; the calibration and accuracy of resident self-assessment cannot be estimated, which is a known limitation of unguided self-evaluation in the health professions [24]. Sixth, evaluation was restricted to Kirkpatrick Level 1; no inference about learning, behavior change, or patient outcomes is warranted [26]. Seventh, no baseline measure of metacognitive awareness was obtained, so pre-existing metacognitive capacity could not be separated from the intervention itself.

Finally, completion does not capture the depth or quality of reflection. High completion may coexist with superficial engagement. Acceptability was also captured only within the postoperative form and therefore reflects only residents who completed that step; those who did not, precisely the cases where burden may have been greatest, contributed no acceptability data, likely biasing acceptability estimates upward. Reported acceptability may also reflect social-desirability or novelty effects over a short implementation window, and sustained use beyond that period remains untested.

Implications and future directions

This pilot supports a defined next step, a controlled study evaluating whether the paired architecture produces measurable changes beyond acceptance. That study should incorporate a prospective registry of all eligible LC cases, baseline measurement of metacognitive awareness, external expert ratings using OSATS/GOALS items, and Kirkpatrick Level 2 and Level 3 outcomes, including repeat-case performance and, where feasible, visual-scanning patterns [12, 14, 31, 32]. A longer implementation window and, ideally, a multicenter design will be needed to test durability of use and generalizability.

Thus, the contribution of this pilot is not evidence of improved operative performance, but evidence that a complete SRL cycle can be operationalized around real operative cases. Its future evaluation will require objective denominators, expert calibration, assessment of reflection quality, and longer follow-up.

Conclusion

A theory-informed, paired pre- and postoperative metacognitive architecture for LC can be completed by surgical residents around routine cases without faculty time at the bedside and was rated favourably by those who completed it, although completion was modest and concentrated in the preoperative phase. This pilot provides only preliminary evidence that a paired forethought–reflection process can be embedded in routine operative workflow; it does not demonstrate effectiveness, improved learning, self-assessment accuracy, or operative performance, and postoperative completion remains the principal implementation challenge. A controlled study with prospective case capture, baseline metacognitive

assessment, and expert calibration is needed to establish whether the architecture affects learning, self-assessment accuracy, and operative transfer.

References

1. Wickramasinghe D, Vincent J. The use of deliberate practice in simulation-based surgical training for laparoscopic surgery: a systematic review. *BMC Med Educ.* 2025;25(1):1047. <https://doi.org/10.1186/s12909-025-07613-w>.
2. Marinho R, Novo de Matos J, Castelo-Branco Sousa M, Forgione A, Gonçalves MR. SafeChole4all: pilot validation of a novel curriculum for safe laparoscopic cholecystectomy training. *Surg Endosc.* 2025;40(2):974–83. <https://doi.org/10.1007/s00464-025-12342-8>.
3. Zimmerman BJ. Becoming a self-regulated learner: an overview. *Theory Pract.* 2002;41(2):64–70. https://doi.org/10.1207/s15430421tip4102_2.
4. Brydges R, Butler D. A reflective analysis of medical education research on self-regulation in learning and practice. *Med Educ.* 2012;46(1):71–9. <https://doi.org/10.1111/j.1365-2923.2011.04100.x>.
5. Isbej L, Waterval D, Fuentes-Cimma J, Riquelme A, de Bruin ABH. Strengthening teachers' dual role in selfregulated learning during diagnostic reasoning supervision. *Med Teach.* 2026. <https://doi.org/10.1080/0142159X.2026.2684561>.
6. Dabekaussen KFAA, Veenstra GL, Vollmann M, Lombarts KMJMH, Jaarsma ADC, Heineman E, et al. Exploring workplace learning in surgical practice: how mindset and motivation are associated with self-regulated learning behaviors. *Perspect Med Educ.* 2025;14(1):699–709. <https://doi.org/10.5334/pme.2144>.
7. Moulton CA, Regehr G, Mylopoulos M, MacRae HM. Slowing down when you should: a new model of expert judgment. *Acad Med.* 2007;82(10 Suppl). <https://doi.org/10.1097/ACM.0b013e3181405a76>. S109-16.
8. Schraw G. Promoting general metacognitive awareness. *Instr Sci.* 1998;26(1–2):113–25. <https://doi.org/10.1023/A:1003044231033>.
9. Schraw G, Moshman D. Metacognitive theories. *Educ Psychol Rev.* 1995;7(4):351–71. <https://doi.org/10.1007/BF02212307>.
10. Lajoie SP. Metacognition, self regulation, and self-regulated learning: a rose by any other name? *Educ Psychol Rev.* 2008;20(4):469–75. <https://doi.org/10.1007/s10648-008-9088-1>.
11. Sandars J. The use of reflection in medical education: AMEE Guide 44. *Med Teach.* 2009;31(8):685–95. <https://doi.org/10.1080/01421590903050374>.
12. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg.* 1997;84(2):273–8. <https://doi.org/10.1046/j.1365-2168.1997.02502.x>.
13. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative bench station examination. *Am J Surg.* 1997;173(3):226–30. <https://doi.org/10.1016/S0002->

9610(97)89597-9.

14. Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondré K, Stanbridge D, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg*. 2005;190(1):107–13. <https://doi.org/10.1016/j.amjsurg.2005.04.004>.
15. Yule S, Flin R, Paterson-Brown S, Maran N, Rowley D. Development of a rating system for surgeons' non-technical skills. *Med Educ*. 2006;40(11):1098–104. <https://doi.org/10.1111/j.1365-2929.2006.02610.x>.
16. Jung S, Carnahan S, Krecko LK, Godfrey M. A metacognitive awareness inventory for surgical educators: creation of the MAISE. *J Surg Educ*. 2022;79(3):569–73. <https://doi.org/10.1016/j.jsurg.2021.12.001>.
17. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med*. 2004;79(10 Suppl). <https://doi.org/10.1097/00001888-200410001-00022>. S70-81.
18. Bond WF, Kuhn G, Binstadt E, Quirk M, Wu T, Tews M, et al. The use of simulation in the development of individual cognitive expertise in emergency medicine. *Acad Emerg Med*. 2008;15(11):1037–45. <https://doi.org/10.1111/j.1553-2712.2008.00229.x>.
19. Gardner AK, Jabbour IJ, Williams BH, Huerta S. Different goals, different pathways: the role of metacognition and task engagement in surgical skill acquisition. *J Surg Educ*. 2016;73(1):61–5. <https://doi.org/10.1016/j.jsurg.2015.08.007>.
20. Briones-Zamora KH, Briones-Claudett KH, Serrano Torres WE, et al. Rapid cycle deliberate practice with spaced training achieves comparable results in achieving simple laparoscopic skills to traditional weekly training and saves time. *J Surg Educ*. 2025;82(11):103686. <https://doi.org/10.1016/j.jsurg.2025.103686>.
21. Abdalgadir M, Soliman D, Shaher Y, Nelson A, Worrell S, Ditillo M, et al. Asynchronous tele-mentoring in proficiency-based training for the fundamentals of laparoscopic surgery: a use of smartphones and deliberate practice. *Surg Endosc*. 2025;40(2):1674–9. <https://doi.org/10.1007/s00464-025-12488-5>.
22. Liu B, Gong B, Zhang X, Lu C, Zheng B, Yang J. Self-directed teaching system driven by video analysis feedback improves surgical skill training for clinical interns: a randomized controlled trial. *BMC Med Educ*. 2026. <https://doi.org/10.1186/s12909-026-09104-y>. Epub ahead of print.
23. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QQuality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. *BMJ Qual Saf*. 2016;25(12):986–92. <https://doi.org/10.1136/bmjqs-2015-004411>.
24. Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. *Acad Med*. 2005;80(10 Suppl):S46–54. <https://doi.org/10.1097/00001888-200510001-00015>.
25. Gardner AK, Abdelfattah K, Wiersch J, Ahmed RA, Willis RE. Embracing errors in simulation-based training: the effect of error training on retention and transfer of central venous catheter skills. *J Surg Educ*. 2015;72(6):e158–62. <https://doi.org/10.1016/j.jsurg.2015.08.002>.

26. Kirkpatrick DL. Evaluating training programs: the four levels. 1st ed. San Francisco: Berrett-Koehler; 1994.
27. Tariq S, Cainelli F, Sahraoui AZ, Achar A. Extended reality in laparoscopic cholecystectomy training: a systematic review and meta-analysis. *J Surg Res.* 2025;316:143–50. <https://doi.org/10.1016/j.jss.2025.10.047>.
28. Oussi N, Sandblom G, Enochsson L. Enhancing laparoscopic simulator performance with eye-tracking video feedback: a mixed-methods pilot study. *Front Surg.* 2026;13:1780289. <https://doi.org/10.3389/fsurg.2026.1780289>.
29. Happ MN, Howell TC, Pollak KI, et al. Building surgical character: a dynamic simulation curriculum for nontechnical skills. *J Surg Educ.* 2025;82(4):103416. <https://doi.org/10.1016/j.jsurg.2024.103416>.
30. Ngam BN, Tchinde NMJ, Kim E, et al. The impact of a self-directed, low-cost laparoscopic simulation-based training model among surgical trainees in Cameroon. *World J Surg.* 2026;50(2):298–306. <https://doi.org/10.1002/wjs.70219>.
31. Zheng B, Zheng Y, Zhang Y, Yao YY. Training surgeons' visual scanning pattern in laparoscopic surgery to enhance patient safety. *J Surg Educ.* 2025;83(2):103821. <https://doi.org/10.1016/j.jsurg.2025.103821>.
32. Schraw G, Dennison RS. Assessing metacognitive awareness. *Contemp Educ Psychol.* 1994;19(4):460–75. <https://doi.org/10.1006/ceps.1994.1033>.