Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Asynchronous Video Feedback and Early Performance Variability in Resident Laparoscopic Cholecystectomy: A Multicenter Observational Study

Matías Aguilera
Pontifical Catholic University of Chile: Pontificia Universidad Catolica de Chile
Diego Sanhueza
Pontifical Catholic University of Chile: Pontificia Universidad Catolica de Chile
Jose Rui-Wamba
Pontificia Universidad Católica de Chile: Pontificia Universidad Catolica de Chile
Enzo Castiglioni
Pontifical Catholic University of Chile: Pontificia Universidad Catolica de Chile
Bernardita Becker
Pontifical Catholic University of Chile: Pontificia Universidad Catolica de Chile
María Inés Gaete
Pontifical Catholic University of Chile: Pontificia Universidad Catolica de Chile
Julian Varas
jevaras@uc.cl

Pontificia Universidad Catolica de Chile Escuela de Medicina https://orcid.org/0000-0002-5828-9623

Research Article

Keywords: surgical education, video-based feedback, laparoscopic cholecystectomy, OSATS, simulation, competency-based education

Posted Date: June 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10037614/v1

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

Abstract

Purpose

Asynchronous video-based feedback is increasingly used in competency-based surgical education, but its effects during early operative learning remain uncertain. We evaluated whether asynchronous video feedback was associated with higher technical performance or reduced early performance variability in resident-performed laparoscopic cholecystectomy.

Methods

We conducted a multicenter observational study of general surgery residents from seven residency programs across five cities. Videos were assessed by blinded raters using a five-item Objective Structured Assessment of Technical Skills (OSATS) rubric (range 5–25). Feedback exposure was defined at the resident level. Early milestones were the 3rd, 5th, and 10th recorded cholecystectomy. Analyses compared central performance, dispersion, baseline OSATS, intra-resident change, and mixed-effects models accounting for repeated observations by resident.

Results

The de-duplicated cohort included 66 residents and 531 unique videos: 281 from feedback-exposed residents and 250 from non-exposed residents. Baseline OSATS was similar between groups. Median OSATS increased with accumulated experience. At the 3rd, 5th, and 10th cases, central OSATS did not differ significantly between groups. At the 10th case, the Feedback group showed a non-significant trend toward lower dispersion (Levene $p = 0.070$). In mixed-effects modeling, case number was associated with higher OSATS (coefficient 0.93 per 10 cases; $p < 0.001$), whereas feedback exposure was not ($p = 0.905$). Simulation exposure showed a favorable descriptive trajectory but was not independently significant ($p = 0.151$).

Conclusions

Asynchronous video feedback was not associated with sustained superiority in central OSATS performance. Exploratory analyses suggest feedback may be better studied as an early standardizing intervention that reduces variability, rather than only as a direct performance booster.

Introduction

Competency-based surgical education requires longitudinal assessment of technical performance, direct observation of workplace practice, and feedback capable of guiding progression toward entrustment [1, 2]. These requirements are especially pertinent in procedural specialties, in which competence depends not

only on knowledge but also on repeated deliberate practice, graduated responsibility, and transfer of technical skills to the operating room [3–6].

Laparoscopic cholecystectomy is an informative index procedure for studying early operative skill acquisition. It is commonly performed, procedurally structured, and requires both psychomotor coordination and intraoperative judgment. Simulation-based laparoscopic training improves operating-room performance in randomized and translational studies, indicating that well-designed simulation can prepare residents for early clinical exposure [7–11]. The Objective Structured Assessment of Technical Skills (OSATS) provides a validated framework for assessing technical performance in this setting [12].

Early operative performance varies substantially across trainees. This variability is educationally consequential because technical skill is associated with clinical outcomes, and mean scores can mask low-end performance or failure to reach a minimum standard [13]. In early training, variability should therefore not be treated only as statistical noise. It may reflect heterogeneity in readiness, consistency, operative exposure, supervision, and the ability to reproduce safe technical behaviors under real operating-room conditions.

From a competency-based perspective, reducing variability may be meaningful even when average scores do not increase. A training intervention that narrows dispersion, attenuates low outliers, or helps residents converge toward a minimum acceptable performance threshold may support standardization and safer progression. This is particularly relevant in distributed residency systems, where trainees may differ in case volume, faculty feedback, and access to simulation.

Video-based review has gained prominence in surgical education because it provides a durable record of operative performance, permits blinded or repeated assessment, and supports structured feedback beyond the operating room [14–21]. Remote and asynchronous formats may be particularly useful in multicenter settings, where faculty availability, case exposure, and feedback quality differ across sites [22–27]. However, the effect of asynchronous feedback during real operations may not be uniform across the learning curve. Early feedback may function less as an immediate score booster and more as a standardizing influence that helps residents identify inconsistency, avoid repeated technical errors, and converge toward more reproducible performance.

Accordingly, we evaluated the association between asynchronous video-based feedback and early technical performance in resident-performed laparoscopic cholecystectomy, examining both central tendency and dispersion. We hypothesized that feedback would not necessarily increase central OSATS performance but might be associated with reduced early performance variability, consistent with a possible early stabilizing effect. We also examined documented simulation exposure as a training-context variable.

Methods

Study design and reporting

We conducted a multicenter comparative observational study within a national surgical education project evaluating remote asynchronous video feedback for general surgery residents. The study was reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) recommendations [28].

Setting and participants

The project included residents from seven general surgery residency programs across five cities. Residents submitted videos of laparoscopic cholecystectomy procedures performed during clinical training through a secure remote video-review workflow. For this manuscript, we included unique resident-performed laparoscopic cholecystectomy videos with available OSATS assessment, recorded case number, and documented feedback exposure status. Cohort size, video counts, and simulation-exposure subgroups are reported in the Results section.

Video-based feedback exposure

The educational exposure was remote, asynchronous feedback delivered after expert review of resident-performed laparoscopic cholecystectomy videos. Feedback consisted of structured comments on technical performance, safe dissection, exposure, operative flow, and areas for improvement. Feedback exposure was defined at the resident level: each resident was classified as belonging to either the Feedback group or the No-Feedback group, and all videos from that resident were analyzed according to that group assignment. Thus, residents did not contribute both Feedback and No-Feedback videos. Allocation followed the educational implementation workflow and was not randomized; feedback was therefore analyzed as an observational exposure.

Performance assessment

The primary performance outcome was total OSATS score. OSATS was assessed using a five-item rubric with a total score range from 5 to 25, where higher scores indicate better technical performance. Video review was performed by blinded raters according to the project workflow. OSATS was selected because it is a validated technical-skills assessment framework widely used in surgical education [12].

Simulation exposure

Simulation exposure was analyzed as a training-context variable rather than as an interaction term. The primary simulation descriptor was documented simulation exposure at the time of the recorded case, defined as any recorded basic or advanced laparoscopic simulation status other than missing or 'no training'. A secondary descriptor was completion of advanced laparoscopic simulation training before or by the resident’s tenth cholecystectomy.

Outcomes and analytic definitions

Early operative performance was assessed at predefined case-number milestones, defined as each resident's 3rd, 5th, and 10th recorded laparoscopic cholecystectomy. These milestones were selected to capture the initial phase of the operative learning curve, when performance is expected to be most variable. Longitudinal performance was also examined using cumulative case-volume bins: cases 1–5, 6–10, 11–20, 21–30, 31–40, and more than 40.

Intra-resident change was defined as ΔOSATS = mean OSATS for cases 8–10 minus mean OSATS for cases 1–3. This complete-case analysis required sufficient observations in both windows and was interpreted as exploratory. Because residents could contribute different numbers of videos, all resident-level denominators are reported separately from video-level denominators.

Baseline comparability

To evaluate baseline comparability without overweighting residents who contributed more videos, we summarized each resident's first available OSATS score by feedback group. We also summarized the resident-level mean OSATS across cases 1–3 among residents with evaluable videos in that window. These baseline summaries were used to contextualize later change-score analyses and to assess whether regression to the mean could plausibly influence ΔOSATS.

Statistical analysis

Analyses were designed to evaluate both central performance and performance variability. Continuous OSATS outcomes were summarized using means, medians, interquartile ranges, and measures of dispersion. Welch's t-test was used to compare mean OSATS or ΔOSATS between groups while allowing for unequal variances [29]. Mann–Whitney U was used as a non-parametric sensitivity analysis for distributional differences [30]. Levene's test was used to compare between-group variance, because reduced variability was considered a potential marker of learning-curve stabilization [31]. Statistical tests were two-sided with $\alpha=0.05$.

Because multiple videos were contributed by the same residents, videos were not statistically independent. As a complementary analysis, we fitted a linear mixed-effects model with total OSATS as the dependent variable. Feedback exposure, cumulative case number scaled per 10 cases, and documented simulation exposure were included as fixed effects. Resident was included as a random intercept to account for repeated videos contributed by the same trainee [32,33]. Center-level random effects could not be included because the analytic Excel file did not contain a center identifier. A random-slope model allowing resident-specific case-number trajectories was explored when feasible. These models were intended to assess whether feedback was associated with OSATS after accounting for repeated observations and accumulated operative experience, while recognizing that mixed-effects modeling does not remove confounding inherent to non-randomized exposure assignment.

Ethical considerations

Ethics approval, consent, funding, competing interests, and author contribution statements are provided in the separate title page/declarations file to preserve blinded peer review. All recordings were de-identified before analysis.

Results

Cohort and denominator structure

The de-duplicated analytic cohort included 66 residents and 531 unique laparoscopic cholecystectomy videos (Table 1). Feedback-exposed residents contributed 281 videos, whereas residents in the No-Feedback group contributed 250 videos. Forty-one residents were in the Feedback group and 25 were in the No-Feedback group. Documented simulation exposure was present in 321 videos (60.5%). Thirteen residents completed advanced laparoscopic simulation training before or by their tenth cholecystectomy; this subgroup contributed 69 videos.

Table 1
Cohort and analysis denominators

Characteristic	Overall	Feedback	No-Feedback
Residents	66	41	25
Unique operative videos	531	281	250
Documented simulation exposure, videos	321 (60.5%)	192 (68.3%)	129 (51.6%)
No documented simulation exposure, videos	210 (39.5%)	89 (31.7%)	121 (48.4%)
Advanced simulation completed by or before case 10, residents	13	—	—
Advanced simulation completed by or before case 10, videos	69	—	—
ΔOSATS complete-case analysis, residents	18	10	8

OSATS = Objective Structured Assessment of Technical Skills. Documented simulation exposure was defined at the video level as any recorded basic or advanced laparoscopic simulation status other than missing or no training.

Baseline OSATS comparability

Resident-level baseline OSATS was similar between groups when defined as each resident's first available OSATS score (Table 3). The Feedback group had a mean baseline OSATS of 17.0 (median 17.0), compared with 16.8 (median 18.0) in the No-Feedback group (Welch p = 0.920; Mann–Whitney p = 0.979). Among residents with evaluable cases 1–3, mean baseline-window OSATS was 17.9 in the Feedback group and 17.0 in the No-Feedback group. These resident-level summaries suggest broadly comparable early performance, while still supporting cautious interpretation because allocation was not randomized.

Table 3
Baseline comparability and intra-resident ΔOSATS complete-case analysis

Measure	Feedback	No-Feedback	Between-group test
First available OSATS per resident	n = 41; mean 17.0; median 17.0	n = 25; mean 16.8; median 18.0	Welch p = 0.920; Mann–Whitney p = 0.979
Mean OSATS across cases 1–3 per resident	n = 22; mean 17.9; median 19.2	n = 12; mean 17.0; median 17.5	Descriptive baseline window
ΔOSATS = mean cases 8–10 − mean cases 1–3	n = 10; mean − 2.02; median − 2.17	n = 8; mean 2.27; median 2.58	Welch p = 0.006; Mann–Whitney p = 0.013; Levene p = 0.009

ΔOSATS was defined as mean OSATS for cases 8–10 minus mean OSATS for cases 1–3. The complete-case analysis included only residents with evaluable videos in both windows and was interpreted as exploratory.

Learning curve by feedback exposure

Across the overall cohort, OSATS increased with accumulated operative experience, consistent with a learning curve (Fig. 1). In longitudinal case bins, median OSATS increased from 18 in the earliest bin to 22 in the No-Feedback group and 19 in the Feedback group in the latest bin (> 40 cases), although the later bins contained fewer residents and were not used as primary between-group tests. The Feedback and No-Feedback curves did not show a robust or sustained separation in central performance.

Early milestone and dispersion analyses

At predefined early milestones, central OSATS performance was similar between groups (Table 2; Fig. 2). At the 3rd case, median OSATS was 18.0 in both groups (Welch p = 0.803; Mann–Whitney p = 0.826). At the 5th case, median OSATS was 18.5 in the Feedback group and 18.0 in the No-Feedback group (Welch p = 0.913; Mann–Whitney p = 1.000). At the 10th case, median OSATS was 18.0 versus 17.5, respectively (Welch p = 0.691; Mann–Whitney p = 0.718). Dispersion did not differ significantly at any milestone. The lowest Levene p value was observed at the 10th case (p = 0.070), suggesting a non-significant trend toward lower variability in the Feedback group.

Table 2
Early milestone OSATS comparisons by feedback exposure

Case milestone	Total n	Feedback n	Feedback median (IQR)	No-Feedback n	No-Feedback median (IQR)	Welch p	Mann–Whitney p	Levene p
3rd	20	15	18.0 (15.0–21.0)	5	18.0 (17.0–18.0)	0.803	0.826	0.830
5th	24	14	18.5 (15.5–21.5)	10	18.0 (16.2–21.2)	0.913	1.000	1.000
10th	28	18	18.0 (13.2–19.0)	10	17.5 (13.2–22.8)	0.691	0.718	0.070

Each milestone represents the resident's recorded case number. Because each resident could contribute at most one video to a given case-number milestone after de-duplication, the number of videos equals the number of residents in each milestone row.

Intra-resident change

The ΔOSATS complete-case analysis included 18 residents: 10 in the Feedback group and 8 in the No-Feedback group (Table 3). Mean ΔOSATS was − 2.02 in the Feedback group and + 2.27 in the No-Feedback group (Welch p = 0.006; Mann–Whitney p = 0.013; Levene p = 0.009). Because this analysis required observations in both early and later windows, included a small resident-level sample, and was sensitive to case-order definitions, it was treated as exploratory rather than as evidence of a causal effect in either direction.

Simulation exposure and training context

Documented simulation exposure was associated descriptively with a more favorable early trajectory (Fig. 3). In video-level longitudinal bins, mean OSATS among videos with documented simulation exposure increased from 17.7 in cases 1–5 to 18.7 in cases 11–20. Among videos without documented simulation exposure, mean OSATS changed from 17.4 to 15.4 over the same bins. These findings describe performance patterns within a training pathway and were not interpreted as causal effects of simulation.

Exploratory mixed-effects model

In an exploratory mixed-effects model with resident random intercepts (Table 4; Fig. 4), case number was independently associated with OSATS (coefficient 0.93 points per 10 cases; 95% CI 0.62 to 1.25; $p < 0.001$), supporting the presence of a learning curve. Feedback exposure was not independently associated with higher OSATS (coefficient − 0.09; 95% CI − 1.55 to 1.38; p = 0.905). Documented simulation exposure had a positive but non-significant association with OSATS (coefficient 0.95; 95% CI − 0.35 to 2.24; p = 0.151).

Table 4
Exploratory mixed-effects model for total OSATS

Term	Coefficient	SE	95% CI	p value
Intercept	15.96	0.75	14.48 to 17.43	< 0.001
Feedback exposure	-0.09	0.75	-1.55 to 1.38	0.905
Case number, per 10 cases	0.93	0.16	0.62 to 1.25	< 0.001
Documented simulation exposure	0.95	0.66	-0.35 to 2.24	0.151

Model included resident as a random intercept. Case number was centered and scaled per 10 cases. Center-level effects could not be included because center identifiers were unavailable in the analytic file.

Discussion

Principal findings

In this multicenter observational study of resident-performed laparoscopic cholecystectomy, asynchronous video feedback was not associated with sustained superiority in central OSATS performance across the learning curve. Central-tendency comparisons were non-significant at every early milestone, and mixed-effects modeling that accounted for repeated observations by resident similarly did not identify an independent association between feedback exposure and higher OSATS. The most relevant feedback-related signal was a non-significant trend toward lower dispersion at the 10th case. These results are exploratory, but they clarify the type of effect that early video feedback may plausibly produce and the type of analysis required to detect it.

These observations refine how asynchronous video feedback should be evaluated in competency-based surgical education. Prior work supports video-based coaching and assessment as feasible and potentially useful for surgical learning [14–21], but the effect of feedback during real operations may be more complex than a simple increase in mean scores. Residents differ in baseline skill, case exposure, readiness for feedback, attending supervision, and ability to translate comments into subsequent operative behavior. A feedback intervention delivered in this context may not shift group averages uniformly, especially when case complexity and clinical supervision vary across procedures.

The distinction between central tendency and variability is central to the contribution of this study. If a program-level intervention fails to increase the mean OSATS score but reduces the frequency of low-performing observations, narrows inter-resident heterogeneity, or helps trainees converge toward reproducible technical behaviors, that effect may still be educationally important. In competency-based education, progression decisions depend not only on high average performance but also on reliability and consistency. Therefore, dispersion-sensitive outcomes may better capture early standardization than mean comparisons alone.

This interpretation should be made carefully. The Levene result at the 10th case was not statistically significant, and the study was not powered primarily to detect differences in variance. However, the distributional figure illustrates why variability deserves attention: similar medians can coexist with different spreads. A narrower spread may indicate that fewer residents are performing at the low end, whereas a wider spread can signal inconsistent readiness or uneven access to preparatory experiences. Future studies should explicitly power analyses for both central performance and variability, and should predefine whether reduced dispersion represents a meaningful educational endpoint.

The intra-resident ΔOSATS analysis illustrates the risk of overinterpreting simple change scores in small non-randomized samples. The complete-case subgroup showed a more favorable change in the No-Feedback group than in the Feedback group, which runs counter to a simple feedback-benefit narrative. This should not be interpreted as evidence that feedback harmed performance. Change scores are sensitive to baseline differences, sparse follow-up, case-order definitions, and regression to the mean. In this study, ΔOSATS was intentionally treated as exploratory and primarily served to highlight the need for denominator transparency and resident-level modeling.

The simulation findings are best interpreted as contextual rather than causal. Documented simulation exposure was associated descriptively with a more favorable early OSATS trajectory, and residents completing advanced simulation by case 10 formed a clinically meaningful subgroup. These observations are consistent with the broader literature on simulation-based transfer to the operating room [7–11]. Nevertheless, simulation status was not randomized, varied over time, and was not independently significant in the mixed-effects model. It should therefore be considered a training-context variable that may influence early performance patterns rather than a confirmed modifier of feedback effectiveness.

From an educational perspective, the findings support a staged approach to early operative learning: front-load structured simulation, use early asynchronous video review to identify inconsistency and low-end performance, and transition to targeted coaching as residents gain experience. This reframes video feedback as part of a longitudinal educational system rather than a stand-alone intervention. Remote and asynchronous workflows may be particularly valuable in geographically distributed programs because they permit expert review, auditability, and feedback delivery outside the constraints of local faculty availability [22–27].

Several limitations should be acknowledged. First, feedback exposure was observational and assigned at the resident level rather than randomized; therefore, residual confounding cannot be excluded. Second, although the mixed-effects model accounted for repeated videos by resident, the analytic file did not contain center identifiers, so center-level clustering could not be modeled directly. Third, the number of residents available at specific early milestones and in the complete-case ΔOSATS analysis was small, and estimates from those analyses should be interpreted cautiously. Fourth, feedback quality, faculty style, attending involvement, operative complexity, and institutional culture may have varied across cases and sites. Finally, operative outcomes such as complications, conversion, and postoperative events were not consistently available for linkage to video-derived performance data.

Future studies should prospectively evaluate asynchronous video feedback in larger cohorts using repeated-measures and multilevel analytic approaches. A candidate model would include feedback exposure, cumulative case number, and simulation exposure as fixed effects, with resident- and center-level random intercepts. When sample size permits, random slopes for case number by resident could account for the fact that trainees may start at different baseline skill levels and improve at different rates. Such models would better estimate whether feedback is associated with higher OSATS, lower variability, or a different learning trajectory after accounting for resident baseline performance, center-level effects, and accumulated operative experience.

Finally, this study underscores the importance of matching statistical methods to educational theory. If the theory of change is that feedback improves average technical skill, then central-tendency outcomes are appropriate. If the theory is that early feedback standardizes performance, reduces low-end outliers, or helps trainees become more consistent, then variance, distributional shape, and threshold-based outcomes are also required. The present results do not confirm a stabilizing effect, but they provide a framework for designing studies that can test this possibility more rigorously.

Another implication concerns the scale at which performance data should be interpreted. Video-level analyses are useful for describing the clinical learning environment because each recording represents an actual operative encounter. However, educational decisions are made about residents, and residents are trained within programs. For this reason, denominators were reported at both the video and resident levels, and the mixed-effects model was included to reduce the risk of interpreting repeated videos from one resident as independent trainees. This distinction should be explicit in future video-based surgical education studies.

A further practical implication is that feedback programs should define success metrics before implementation. If the objective is remediation, the key endpoint may be reduction in low-end observations; if the objective is acceleration, the endpoint may be slope of improvement; and if the objective is standardization, dispersion and threshold attainment may be more informative than average OSATS. Predefining these outcomes would prevent post hoc interpretation of weak or discordant signals and would align analysis with the intended educational mechanism.

Conclusions

Asynchronous video feedback was not associated with sustained superiority in central OSATS performance in resident-performed laparoscopic cholecystectomy. Exploratory analyses suggested a possible early reduction in performance variability, raising the hypothesis that video feedback may play a standardizing role early in technical skill acquisition. These findings support evaluating feedback interventions using both central-tendency and dispersion-sensitive metrics in adequately powered studies with repeated-measures and multilevel analytic approaches.

References

1. Frank JR, Snell LS, Cate OT, Holmboe ES, Carraccio C, Swing SR, et al. Competency-based medical education: theory to practice. Med Teach. 2010;32(8):638-645. https://doi.org/10.3109/0142159X.2010.501190.
2. ten Cate O. Entrustability of professional activities and competency-based training. Med Educ. 2005;39(12):1176-1177. https://doi.org/10.1111/j.1365-2929.2005.02341.x.
3. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. Psychol Rev. 1993;100(3):363-406. https://doi.org/10.1037/0033-295X.100.3.363.
4. Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach. 2005;27(1):10-28. https://doi.org/10.1080/01421590500046924.
5. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? Acad Med. 2011;86(6):706-711. https://doi.org/10.1097/ACM.0b013e318217e119.
6. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. JAMA. 2011;306(9):978-988. https://doi.org/10.1001/jama.2011.1234.
7. Dawe SR, Pena GN, Windsor JA, Broeders JAJL, Cregan PC, Hewett PJ, Maddern GJ. Systematic review of skills transfer after surgical simulation-based training. Ann Surg. 2014;259(2):236-248. https://doi.org/10.1097/SLA.0000000000000245.
8. Zendejas B, Brydges R, Hamstra SJ, Cook DA. Simulation-based laparoscopic surgical training for improving surgical skills. Cochrane Database Syst Rev. 2013;(8):CD010478. https://doi.org/10.1002/14651858.CD010478.pub2.
9. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, Satava RM. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg. 2002;236(4):458-463. https://doi.org/10.1097/00000658-200210000-00008.
10. Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. Br J Surg. 2004;91(2):146-150. https://doi.org/10.1002/bjs.4407.
11. Larsen CR, Soerensen JL, Grantcharov TP, Dalsgaard T, Schouenborg L, Ottosen C, et al. Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. BMJ. 2009;338:b1802. https://doi.org/10.1136/bmj.b1802.
12. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. 1997;84(2):273-278. https://doi.org/10.1046/j.1365-2168.1997.02502.x.
13. Birkmeyer JD, Finks JF, O'Reilly A, Oerline M, Carlin AM, Nunn AR, et al. Surgical skill and complication rates after bariatric surgery. N Engl J Med. 2013;369(15):1434-1442. https://doi.org/10.1056/NEJMsa1300625.
14. Augestad KM, Butt K, Ignjatovic D, Keller DS, Kiran R. Video-based coaching in surgical education: a systematic review and meta-analysis. Surg Endosc. 2020;34(2):521-535.

https://doi.org/10.1007/s00464-019-07265-0.

15. Daniel R, McKechnie T, Kruse CC, Levin M, Lee Y, Doumouras AG, Hong D, Eskicioglu C. Video-based coaching for surgical residents: a systematic review and meta-analysis. Surg Endosc. 2023;37(2):1429-1439. https://doi.org/10.1007/s00464-022-09379-4.
16. Hu YY, Mazer LM, Yule SJ, Arriaga AF, Greenberg CC, Lipsitz SR, Gawande AA, Smink DS. Complementing operating room teaching with video-based coaching. JAMA Surg. 2017;152(4):318-325. https://doi.org/10.1001/jamasurg.2016.4619.
17. Bonrath EM, Dedy NJ, Gordon LE, Grantcharov TP. Comprehensive surgical coaching enhances surgical skill in the operating room: a randomized controlled trial. Ann Surg. 2015;262(2):205-212. https://doi.org/10.1097/SLA.0000000000001214.
18. Herrera-Almario GE, Kirk K, Guerrero VT, Jeong K, Kim S, Hamad GG. The effect of video review of resident laparoscopic surgical skills measured by self- and external assessment. Am J Surg. 2016;211(2):315-320. https://doi.org/10.1016/j.amjsurg.2015.05.039.
19. Green JL, Suresh V, Bittar P, Ledbetter L, Mithani SK, Allori AC. The utilization of video technology in surgical education: a systematic review. J Surg Res. 2019;235:171-180. https://doi.org/10.1016/j.jss.2018.09.015.
20. Pugh CM, Hashimoto DA, Korndorffer JR Jr. The what? how? and who? of video-based assessment. Am J Surg. 2021;221(1):13-18. https://doi.org/10.1016/j.amjsurg.2020.06.027.
21. Youssef SC, Aydin A, Canning A, Khan N, Ahmed K, Dasgupta P. Learning surgical skills through video-based education: a systematic review. Surg Innov. 2023;30(2):220-238. https://doi.org/10.1177/15533506221120146.
22. Sanhueza RD, Jarry TC, Varas CJ. HOW I DO IT: Breaking boundaries in surgical education by delivering expert feedback to residents anytime and anywhere. The LAPPCLINIC project. Surg Open Sci. 2025;28:7-10. https://doi.org/10.1016/j.sopen.2025.09.007.
23. Jarry C, Varas J. Distance simulation in surgical education. Surgery. 2025;180:109097. https://doi.org/10.1016/j.surg.2024.109097.
24. Miguieles Schilling M, Montero Jaras I, Valencia Coronel B, Duran Espinoza V, Gaete Dañobeitia MI, Belmar Riveros F, et al. Home-based training for first-year surgery residents: learning to perform an open manual intestinal anastomosis with remote and asynchronous feedback. Glob Surg Educ. 2024;3:63. https://doi.org/10.1007/s44186-024-00259-0.
25. Montero Jaras I, Duran Espinoza V, Miguieles Schilling M, Riveros F, Fernández Ú, Valencia Coronel B, et al. Effectiveness of a Train the Trainers course for digital feedback in healthcare simulation via a remote and asynchronous learning program. Glob Surg Educ. 2024;3:64. https://doi.org/10.1007/s44186-024-00258-1.
26. Jarry Trujillo C, Vela Ulloa J, Escalona Vivas G, Grasset Escobar E, Villagrán Gutiérrez I, Achurra Tirado P, et al. Surgeons vs ChatGPT: assessment and feedback performance based on real surgical scenarios. J Surg Educ. 2024;81(7):960-966. https://doi.org/10.1016/j.jsurg.2024.03.012.
27. Cruz E, Selman R, Figueroa Ú, Belmar F, Jarry C, Sanhueza D, et al. A scalable solution: effective AI implementation in laparoscopic simulation training assessments. Glob Surg Educ. 2025;4:46.

https://doi.org/10.1007/s44186-025-00355-9.

28. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370(9596):1453-1457. https://doi.org/10.1016/S0140-6736(07)61602-X.
29. Welch BL. The generalization of Student's problem when several different population variances are involved. Biometrika. 1947;34(1-2):28-35. https://doi.org/10.1093/biomet/34.1-2.28.
30. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat. 1947;18(1):50-60. https://doi.org/10.1214/aoms/1177730491.
31. Levene H. Robust tests for equality of variances. In: Olkin I, Ghurye SG, Hoeffding W, Madow WG, Mann HB, editors. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. Stanford: Stanford University Press; 1960. p. 278-292.
32. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982;38(4):963-974. https://doi.org/10.2307/2529876.
33. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1):1-48. https://doi.org/10.18637/jss.v067.i01.

Figures

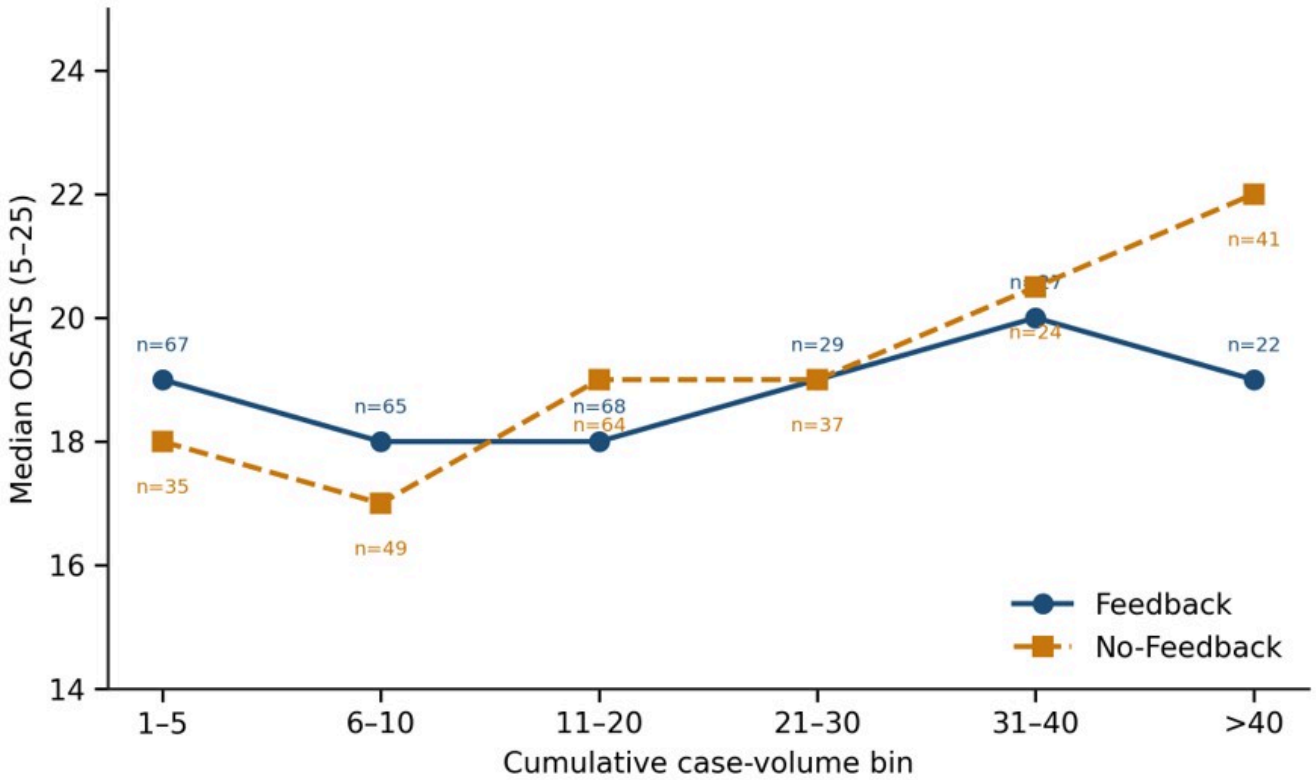


Figure 1

Median OSATS by cumulative case bin and feedback exposure. Point labels show the number of videos contributing to each bin

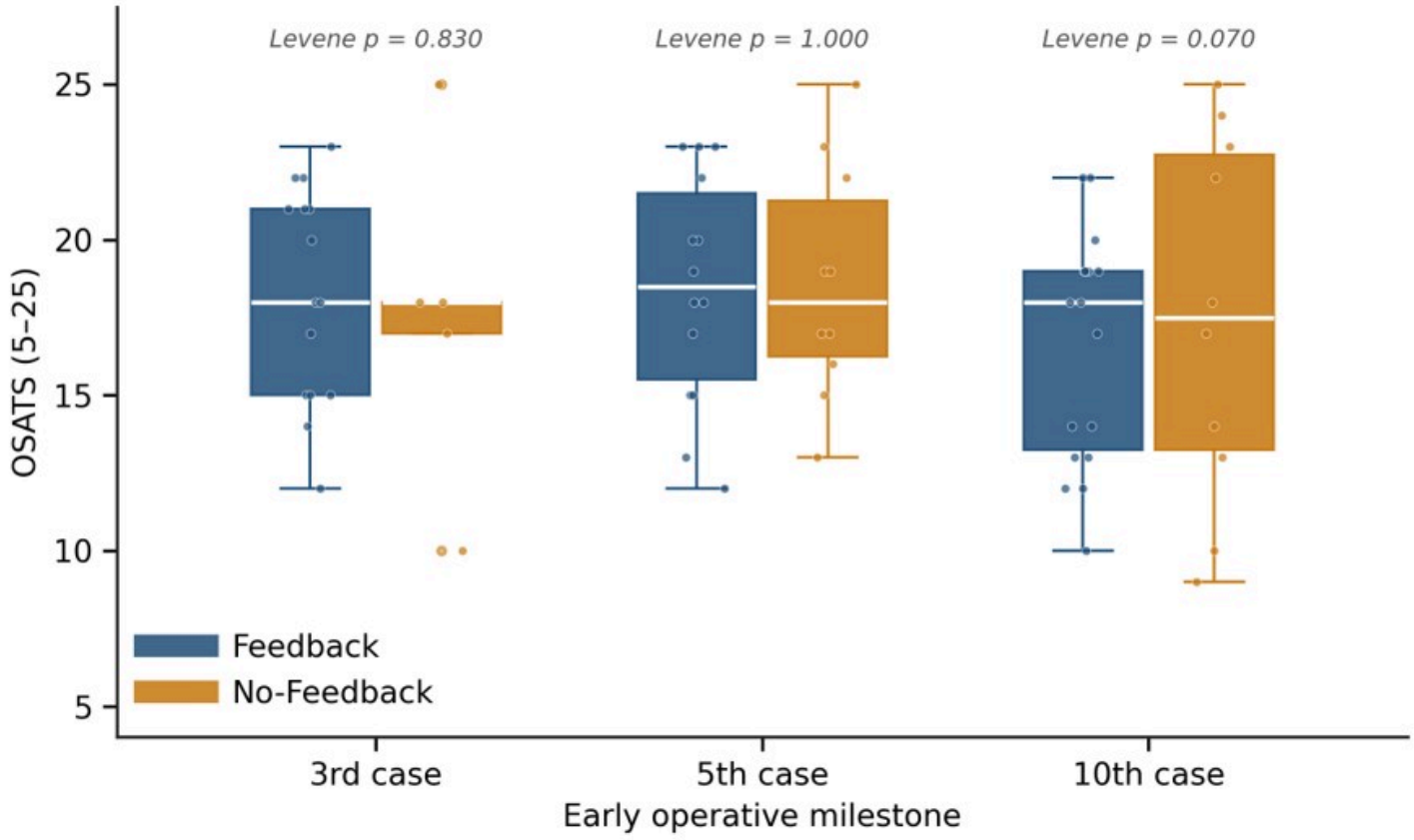


Figure 2

Early OSATS distributions at the 3rd, 5th, and 10th recorded laparoscopic cholecystectomy by feedback exposure. Boxplots show median and interquartile range; points show individual milestone observations. Median-centered Levene p values are shown above each milestone

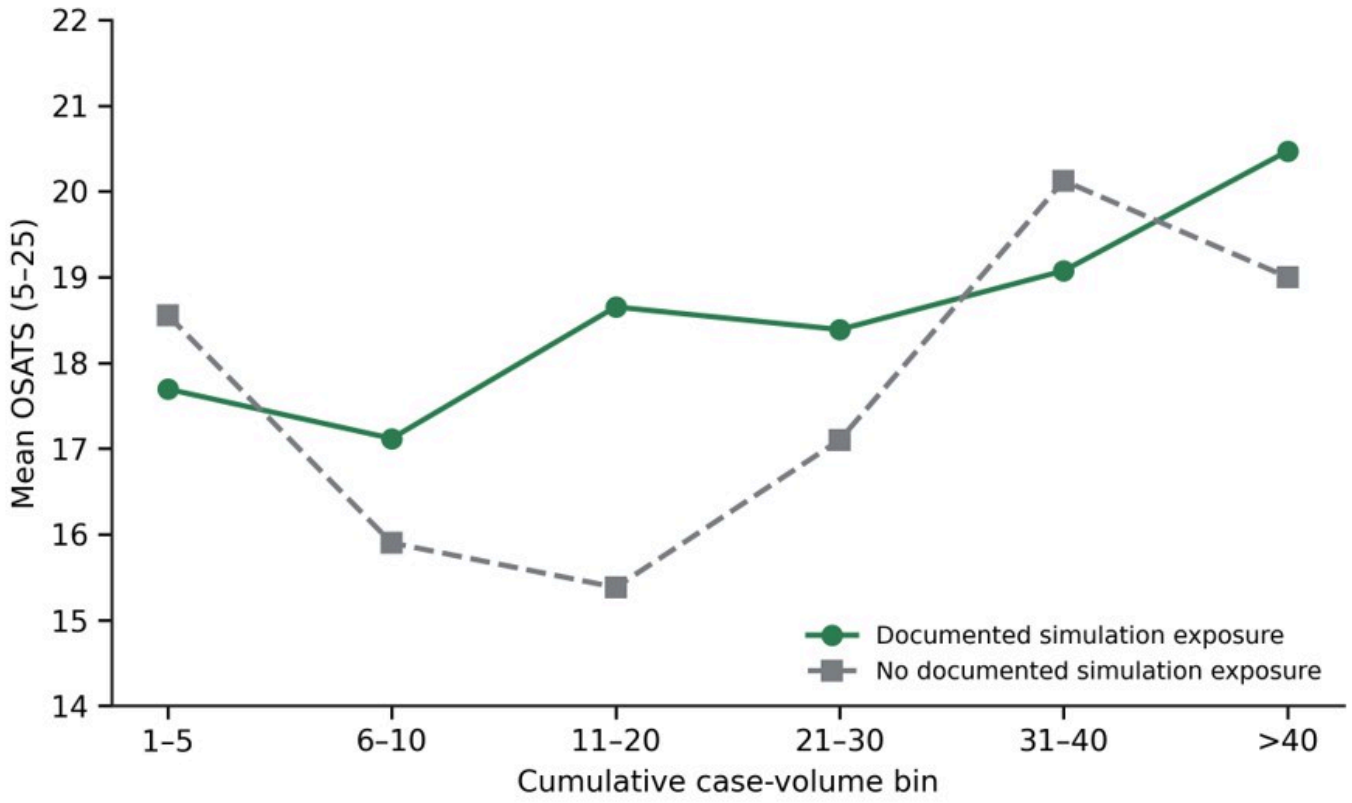


Figure 3

Mean OSATS by cumulative case bin and documented simulation exposure. Simulation exposure was defined at the video level

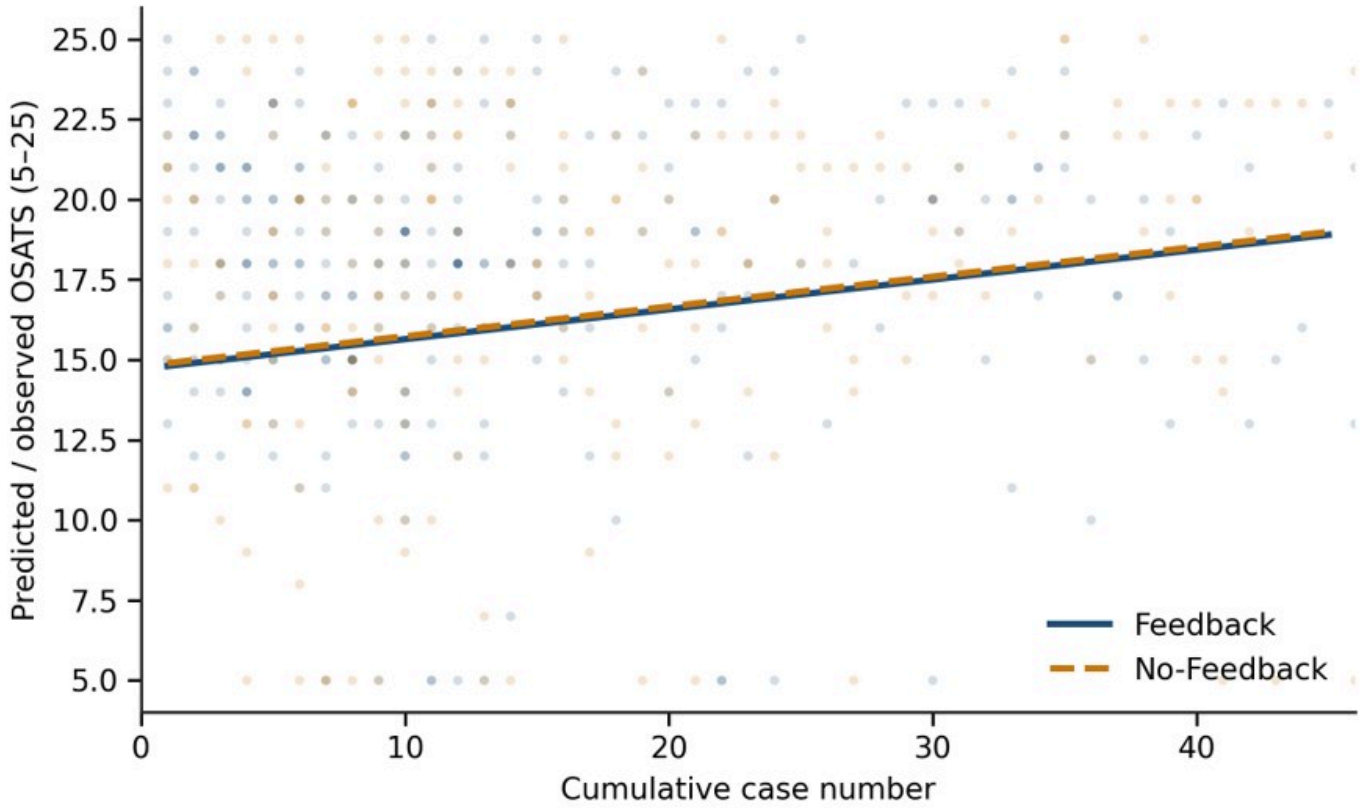


Figure 4

Exploratory mixed-effects model predictions for OSATS by feedback exposure across cumulative case number, with observed video-level OSATS values shown as faint points. Model included resident random intercepts and adjusted for documented simulation exposure