frontiers | Frontiers in Medicine

TYPE Methods
PUBLISHED 17 September 2025
DOI 10.3389/fmed.2025.1591043

OPEN ACCESS

EDITED BY
Gabriel Sandblom,
Karolinska Institutet (KI), Sweden

REVIEWED BY
Safaa Albasri,
Mustansiriyah University, Iraq
Ophir Ilan,
Wolfson Medical Center, Israel
Rayan Harari,
Spaulding Rehabilitation Hospital and Harvard Medical School, United States

*CORRESPONDENCE
Jin Zhang
398448963@qq.com

RECEIVED 20 March 2025
ACCEPTED 28 August 2025
PUBLISHED 17 September 2025

CITATION
Yu Z, Liu Q and Zhang J (2025) Reliability Volume: a novel metric for surgical skill evaluation. *Front. Med.* 12:1591043. doi: 10.3389/fmed.2025.1591043

Reliability Volume: a novel metric for surgical skill evaluation

Zhipu Yu[1], Qinghua Liu[1] and Jin Zhang[2]*

[1]School of Physics and Electronic Information, Yan'an University, Yan'an, China, [2]Yan'an University Affiliated Hospital, Yan'an, China

This study introduces Reliability Volume (RV), an integrated metric combining trajectory similarity with empirical reliability estimation using threshold counts to evaluate surgical skill during repetitive training. RV quantifies both spatial precision and the probability of consistent task execution, addressing limitations of single-session metrics that neglect fatigue and performance drift. Applied to knot-tying with assistive devices, RV jointly reflects spatial accuracy and performance consistency over multiple sessions. Our results show that RV reliably tracks learning progression and is readily compatible with real-time (closed-loop) feedback systems, providing a dynamic, comprehensive, and practice-oriented assessment framework.

KEYWORDS

surgical skill, repetitive training, trajectory similarity, reliability, fatigue

1 Introduction

Surgical skill evaluation methods can be broadly categorized as subjective and objective. Subjective evaluations, including expert ratings and self-assessments, remain prevalent yet suffer from rater bias, inconsistent standards, and inefficiency (1, 2). Objective evaluations quantify surgical gestures, eye movements, or instrument trajectories (3–7), but often require specialized hardware, complex analyses, and substantial expertise, limiting practicality (8). In pursuit of more accurate assessments, quantitative metrics such as force-based (9), time-based (4, 10), and spatial indicators (e.g., path length and smoothness) (11, 12) have received considerable attention. Methods including Dynamic Time Warping (DTW), Hidden Markov Models (HMM), and kinematic feature extraction are widely used to evaluate the quality and similarity of surgical movements (13–15). Recent reviews also highlight the rapid growth of computer vision and AI for objective skill assessment and training across open, laparoscopic, and robotic platforms (16, 17).

Repetitive practice of fundamental skills is particularly important given limited operating room opportunities, duty-hour restrictions, and ethical constraints. Although repetition can improve accuracy, efficiency, and trainee confidence, most evaluation metrics focus on single sessions and do not adequately account for cumulative fatigue and performance drift during repetitive training.

Fatigue is a key external factor. Kahol et al. reported cognitive deterioration due to fatigue and sleep deprivation in virtual reality simulations that was not captured by operative time alone (18). More recent syntheses show mixed but concerning effects of surgeon fatigue on performance and outcomes and call for direct, within-task measures rather than retrospective proxies (19).

To contrast single-session (open-loop) and repeated-session (closed-loop) training, we compare traditional methods with real-time feedback systems, as illustrated in Figure 1 (20–23). Open-loop approaches provide delayed feedback only after task completion, limiting opportunities for in-task correction. Closed-loop approaches deliver

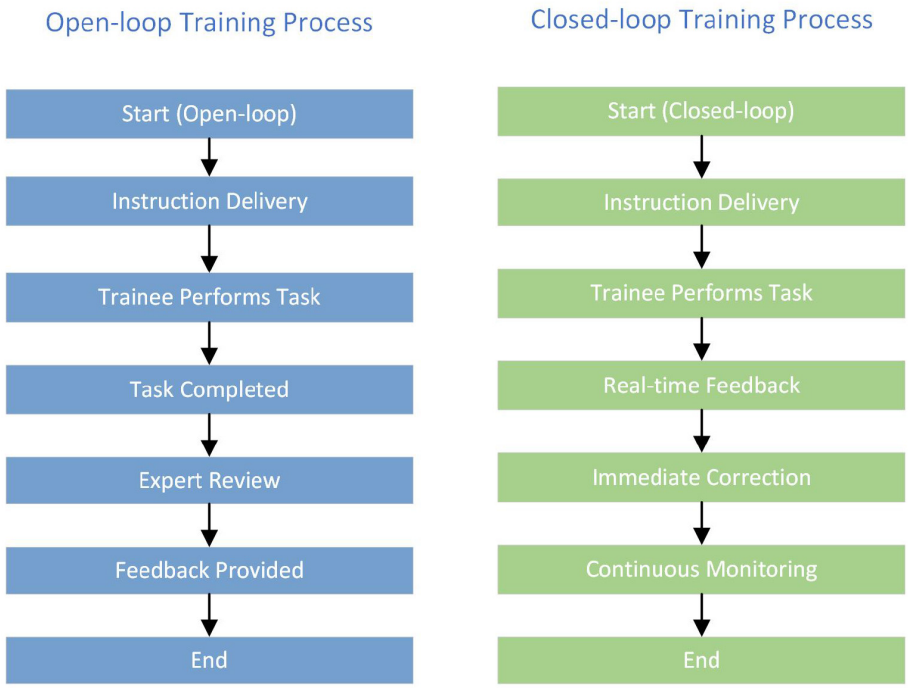

FIGURE 1
Comparison of open-loop and closed-loop training processes.

immediate feedback and continuous monitoring, enabling trainees to adjust actions promptly and mitigate negative effects from fatigue and other external factors.

Addressing these limitations, we propose *Reliability Volume* (RV), derived from Euclidean distance (13), working volume (11), and empirical reliability estimation (24, 25). Unlike traditional metrics, RV quantifies a trainee's consistency in real-time, closed-loop environments by jointly capturing short-term spatial accuracy and long-term consistency. RV thus provides a comprehensive, realistic, and practical framework that bridges theoretical modeling and real-world training.

2 Reliability volume and its calculation

RV is a bivariate descriptor reported as an ordered pair (R, V), where R is the probability of successfully completing the task within specified conditions, and V represents the corresponding working-space volume. A lower R indicates a higher probability of failure; a smaller V indicates closer alignment with the standard path. The workflow is shown in Figure 2.

$$\mathrm{RV} = (R,\ V). \tag{1}$$

Specifically, the calculation steps are as follows.

Step 1: capture standard and training paths

We define two sets of 3D trajectories:

- **Standard path** $S = \{s_1, s_2, \ldots, s_N\}$, with $s_n = (x'_n, y'_n, z'_n)$.
- **Training path** $T_m = \{t_{m1}, t_{m2}, \ldots, t_{mN}\}$ for the m-th repetition, where $t_{mn} = (x_{mn}, y_{mn}, z_{mn})$, $m \in \{1, \ldots, M\}$.

Step 2: pointwise euclidean deviation

The deviation at index n of repetition m is

$$d_{mn} = \|t_{mn} - s_n\| = \sqrt{(x_{mn} - x'_n)^2 + (y_{mn} - y'_n)^2 + (z_{mn} - z'_n)^2}. \tag{2}$$

Step 3: order the deviations

Collect all d_{mn} and sort in descending order to obtain $D_s = \{d_{\max}, \ldots, d_j, \ldots, d_{\min}\}$, where d_j denotes a distance (radius) threshold. (Here D_s denotes the multiset of all d_{mn}).

Step 4: working-space volume

Unlike the conventional *working volume* defined as a sphere whose radius equals the average distance from a hand-centered point (11), we model a *working-space volume* as a cylindrical tube coaxial with the standard path (Figure 3). For a given threshold d_j, the working-space volume is

$$V_j = \pi d_j^2 h \tag{3}$$

where h is the arc length of the standard path and d_j is the tube radius.

Step 5: empirical reliability estimation

The Monte Carlo method is a powerful statistical tool for evaluating the ability to complete a specified surgical task within

a given time and environment (24, 25). Therefore, for each distance d_j, the state function is $Z = g(m,n) = d_{mn} - d_j$. Based on this state equation, $d_{mn} - d_j = 0$ can divide the variable space into a failure space and a reliability space, and the working-space volume, a cylinder centered on the target path, defines the reliable space. Moreover, count the number of paths n_j, which d_{mi} does not exceed d_j, and compute reliability R_j as:

$$R_j = \frac{n_j}{M} \tag{4}$$

where R_j is the reliability corresponding to distance d_j.

Step 6: Reliability Volume

Finally, the Reliability Volume at threshold d_j is

$$\mathrm{RV}_j = (R_j,\ V_j). \tag{5}$$

3 Experiment: knot-tying with assistive devices

3.1 Path data collection

Path data were collected using an optical motion-tracking system (Beijing DuLiang Technology Co.) to monitor hand movements during the experiment. The core hardware and software configurations of this system are detailed in Table 1.

As shown in Figure 4, a 12 mm reflective marker was affixed to a pair of hemostatic forceps. Trainees used the instrument to tie a suture around a needle holder, completing two full loops at a self-selected comfortable speed. Each repetition started at a prescribed start point and ended at a predefined boundary. The task was performed within a cylindrical workspace (Figure 5) with a fixed height of 2 cm and variable radius r (cm).

All trajectories were recorded at a uniform sampling frequency and saved in CSV format to ensure consistent path length for subsequent computational comparisons.

3.2 Participants

Participants included four students, three surgical residents, two attending surgeons, and one associate chief surgeon. The associate chief surgeon performed the knot-tying procedure once to define the standard path. Each trainee then imitated the task 50 times at a self-selected comfortable speed, with no time limit imposed. A total of nine trainees (five male, four female) participated, with demographic information indicated in the captions of Figures 7–15. Path data were collected via the motion-tracking system.

3.3 Standard path

Figure 6 presents the standard path generated by the associate chief surgeon, which served as the reference for trainees.

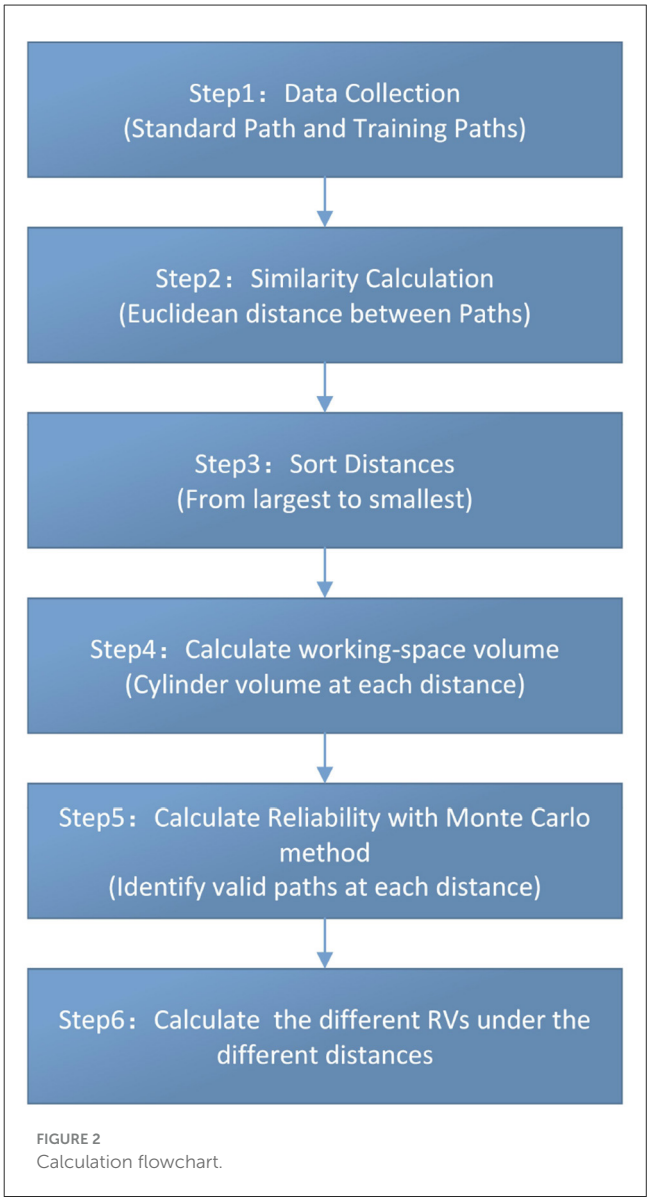

FIGURE 2
Calculation flowchart.

3.4 Reliability Volume (RV) results

Since only the horizontal displacement between the start and end points was constrained—with training paths also being influenced by trainees' experience and physical condition—the actual training paths diverge from the standard path. Accordingly, the Reliability Volume (RV) results are grouped by role: Figures 7–10 (students), Figures 11–13 (surgical residents), and Figures 14, 15 (attending surgeons). In each figure, the left panel shows how R varies with the working-space volume V at $M = 50$ repetitions; the right panel shows how the working-space volume V varies with the number of repetitions when $R = 0.95$.

In the first panels, students generally operate at smaller working-space volumes V (i.e., higher spatial precision relative to the standard path) but exhibit broader transitions in reliability from $R \approx 0$ to $R \approx 1$, indicating greater performance variability compared with experienced participants. This observation is

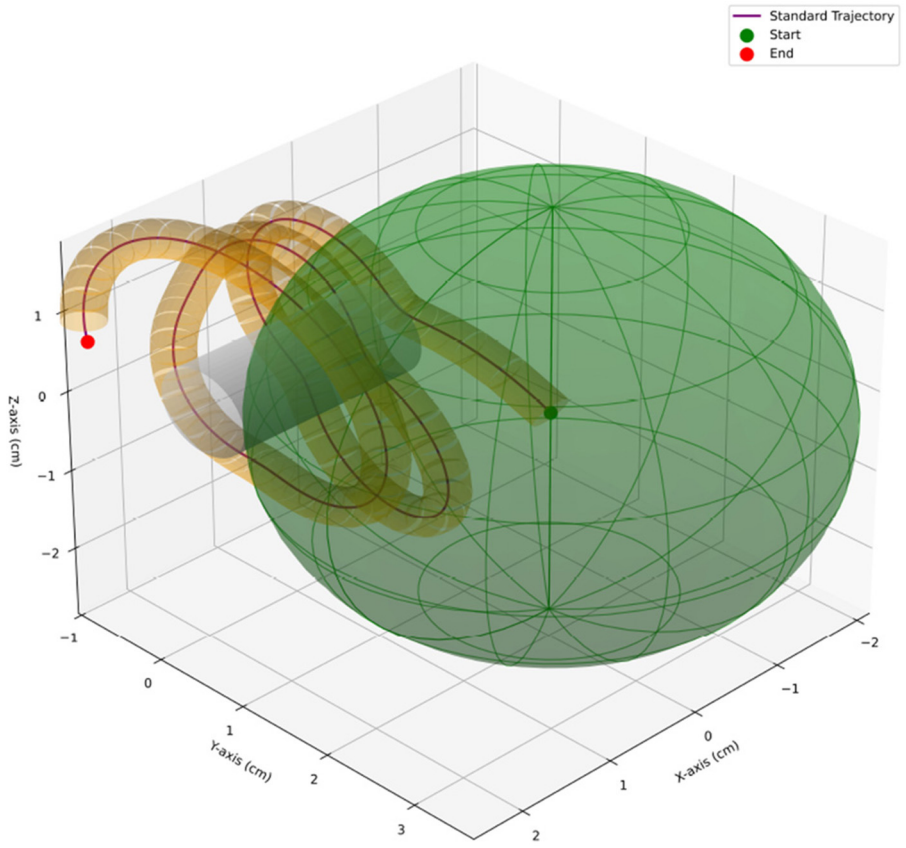

FIGURE 3
Working-space volume (cylindrical tube around the standard path) vs. conventional working volume.

TABLE 1 Optical Motion-tracking system core configuration for path data collection.

Component type	Equipment name	Brand and model	Key technical specifications
Core Hardware	Optical Motion-Tracking Camera	NOKOV Mars1.3H	Resolution: 1,280 × 1,024 (1.3 million pixels); Max acquisition frequency (full resolution): 240 Hz (adjustable); Power supply: Power over ethernet (POE); Interface: GigE/POE
	8-port POE switch (power supply)	NOKOV POE8/8-ONV1	POE power ports: 8; Data transmission port: 1; Total power output: 128 W
Core Software	Motion Tracking & Data Analysis Software	NOKOV XINGYING	Data processing: FPGA edge computing; Compatibility: Supports Windows/Linux/MATLAB/Simulink/ROS

consistent with previous reports that experienced operators tend to emphasize stability, whereas novices often trade stability for precision (11, 12).

In the second panels, RV reveals training dynamics that are often obscured by traditional single-metric summaries. When fixing $R = 0.95$, a favorable trend is a reduction in V with increasing repetitions, reflecting improved precision at a constant success probability. For example, student2, student4, and resident 2 show extended intervals of negative correlation between repetition count and V, suggesting more effective practice results.

By contrast, some participants demonstrate positive correlations or non-monotonic patterns. For instance, RV snapshots for student1 at the 10th, 25th, and 50th repetitions (Table 2) reveal that reliability at a fixed volume (e.g., 427.84 cm^3) can fluctuate (0.90 → 0.96 → 0.90). Similarly, for student2, RV snapshots at the same repetitions (Table 3) demonstrate variability at a constant volume (e.g., 613.33 cm^3), with reliability shifting from 0.90 → 0.84 → 0.88.

Clearly, this indicates that a one-size-fits-all imitation training approach may not be suitable for all trainees. While some individuals can achieve improved precision at a high success

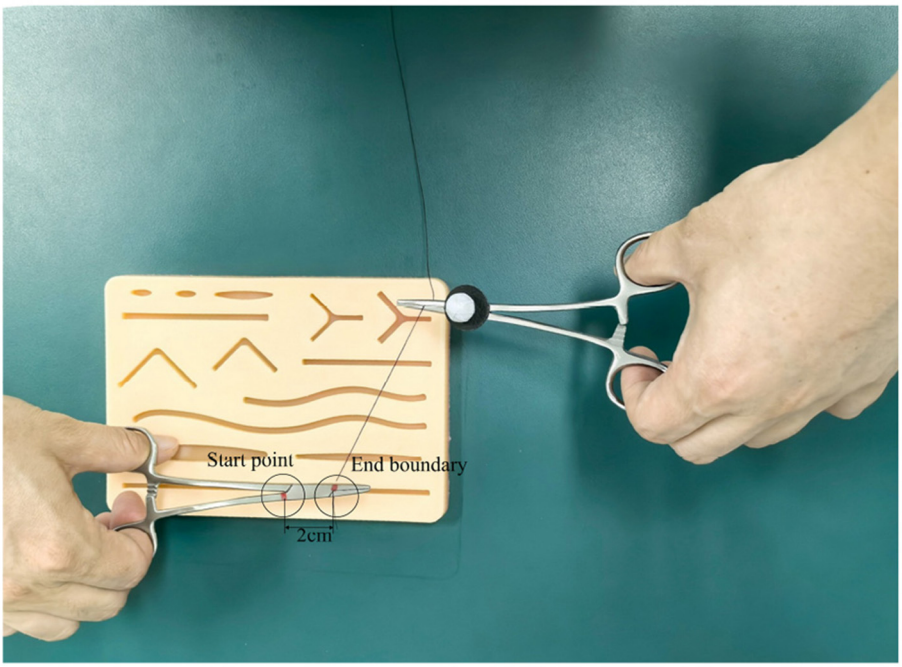

FIGURE 4
Working environment.

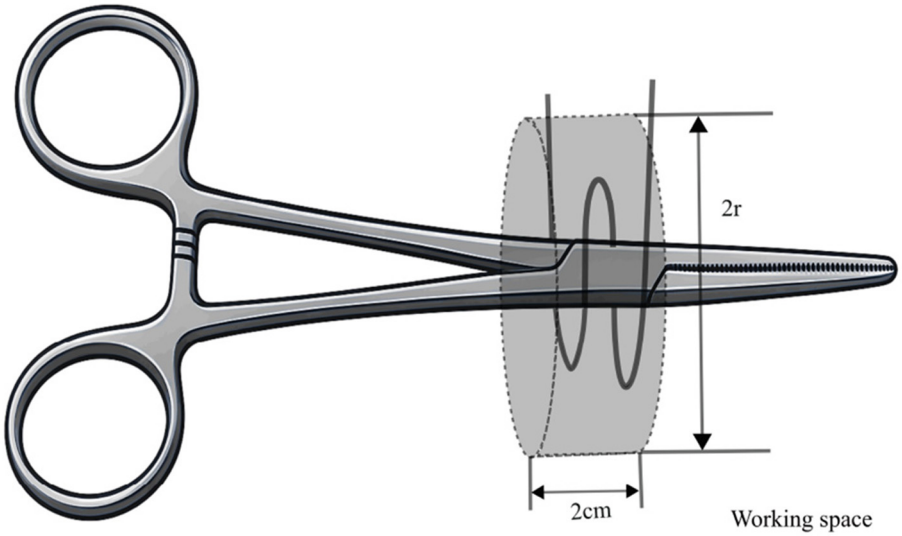

FIGURE 5
Schematic of the working-space volume.

probability after completing 50 repetitive training sessions, others may not demonstrate such progress.

3.5 Fatigue and dynamic feedback

In terms of fatigue, the ability to complete a specified surgical task under defined conditions is closely linked to fatigue accumulation with increasing repetitions. RV offers a practical means to *capture* such effects: fluctuations in R at a fixed V across repetitions are consistent with transient fatigue or distraction.

For training management, we propose a simple stopping rule compatible with closed-loop feedback: define a reliability change threshold (e.g., $|\Delta R| \geq 0.05$) at a fixed volume. When within-session reliability changes by at least this amount, the session should be paused and skill evaluated using the last stable RV point (the

measurement immediately preceding the change). For example, at a working-space volume of 427.84 cm^3, student1 should stop imitation training at the 10th repetition, as reliability declined from 0.96 to 0.90. At 613.07 cm^3, student2 should stop at the 25th repetition, as reliability shifted from 0.80 to 0.86. Notably, this stopping rule should account for gradual changes, and this will be explored in future research.

Thus, RV is not only an integrated metric for quantifying task success probability under specified conditions, but also a dynamic measure that reflects fluctuations caused by fatigue or distraction.

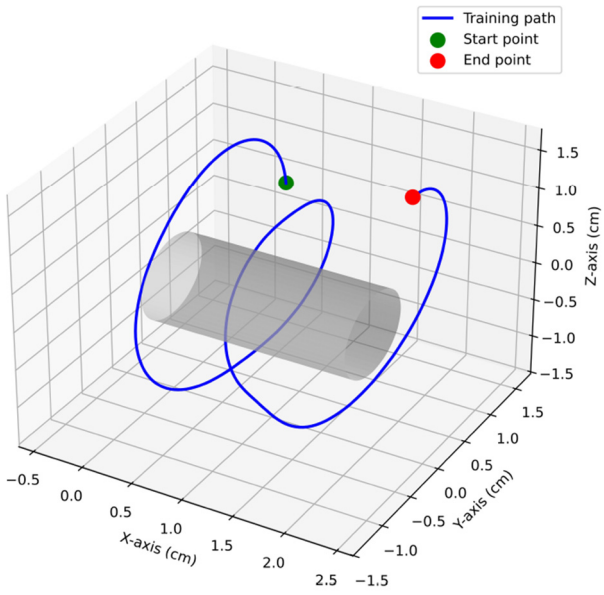

FIGURE 6
Standard path.

4 Discussion

4.1 Practicality of RV for capturing skill development

As with conventional working volume (11), RV reflects the expected gradient of spatial economy with increasing experience. In our data, the maximum of working-space volume (at $R = 1$) decreased consistently across groups: students (≈853.57 cm^3), surgical residents (≈465.62 cm^3), and attending surgeons (≈270.33 cm^3). Similarly, the average working volume was 164.50 cm^3 for students, 66.52 cm^3 for residents, and 18.30 cm^3 for attending surgeons.

Figure 16 and Table 4 highlight why RV-derived volumes may diverge from conventional working volume. The RV tube radius is defined by the worst-case deviation from the standard path (maximal d_{mn}), whereas the conventional working volume relies on the average distance from a hand-centered point. When fatigue or other uncertainties cause occasional large deviations, the RV maximum volume remains anchored to its tolerance definition and is comparatively stable. By contrast, the average-based working volume is more sensitive to fluctuations.

Thus, although both RV and conventional working volume can stratify experience, RV provides greater practical utility by integrating *all* repetitions within a closed-loop framework (Figure 3). Compared with established metrics such as path length, smoothness, and working volume, RV emphasizes consistency across repetitions rather than single-session snapshots, thereby offering complementary information for comprehensive skill assessment.

4.2 Perceived value and implications

Currently, Reliability Volume (RV) primarily focuses on spatial consistency; however, incorporating task duration represents a

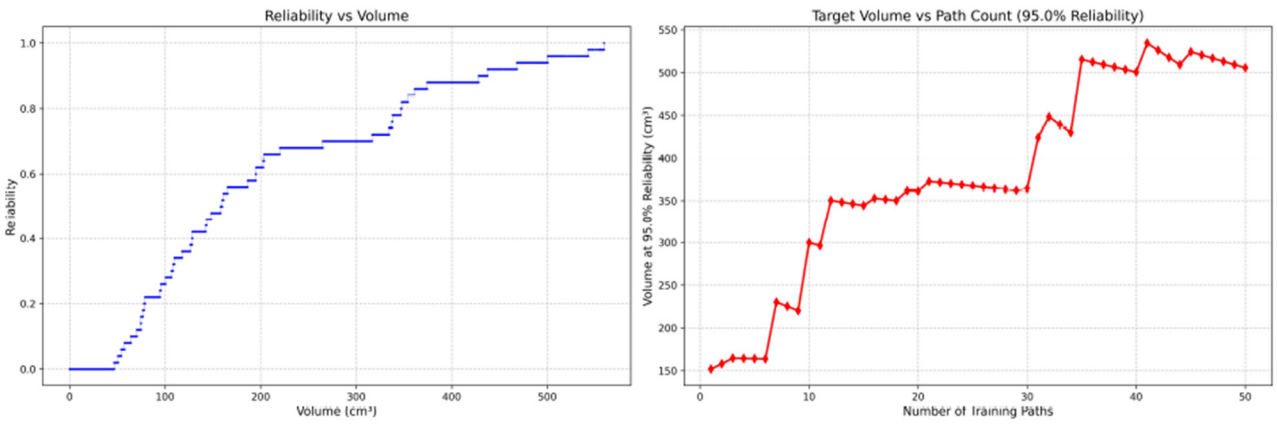

FIGURE 7
Reliability Volume (RV) results for student 1 (male). **(Left)** R vs. V; **(Right)** V vs. repetitions at $R = 0.95$.

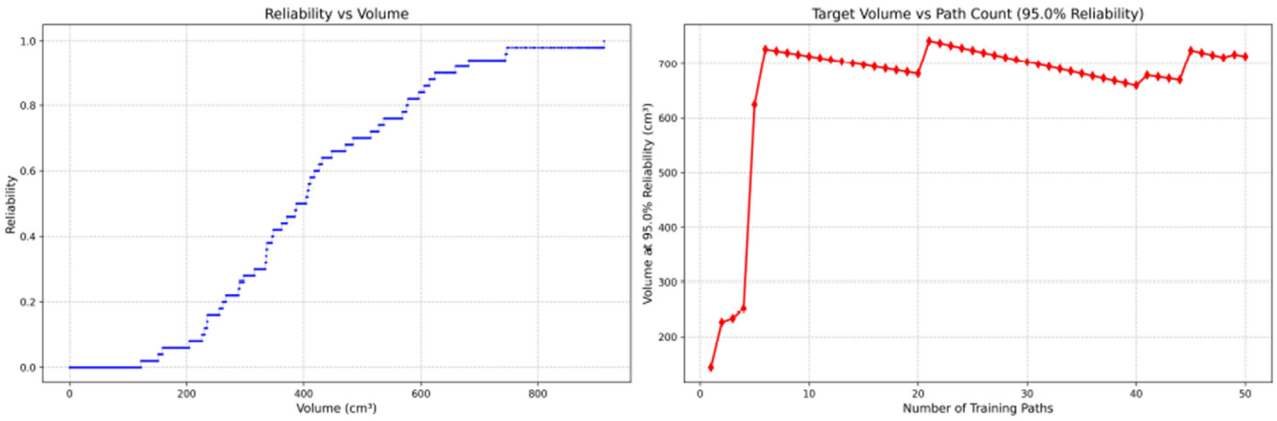

FIGURE 8
Reliability Volume (RV) results for student 2 (male). **(Left)** *R* vs. *V*; **(Right)** *V* vs. repetitions at $R = 0.95$.

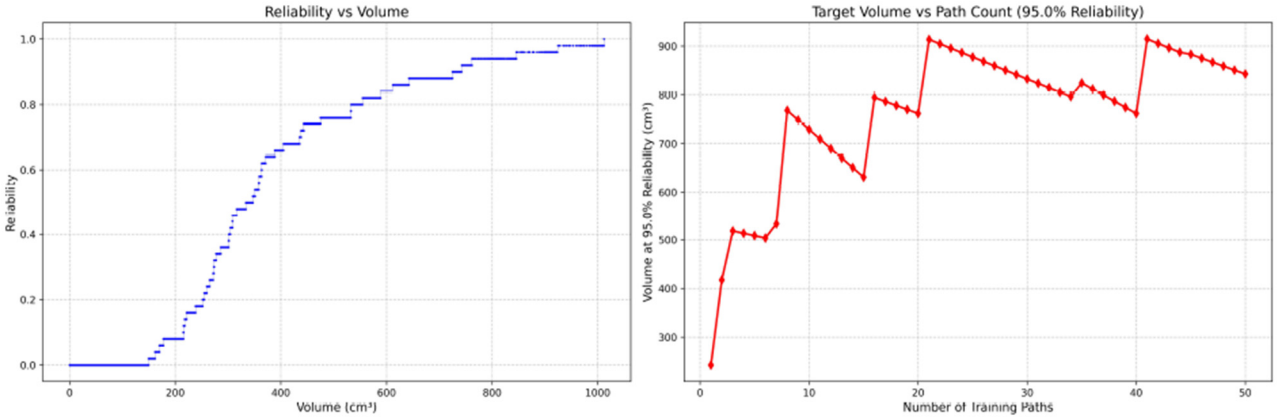

FIGURE 9
Reliability Volume (RV) results for student 3 (male). **(Left)** *R* vs. *V*; **(Right)** *V* vs. repetitions at $R = 0.95$.

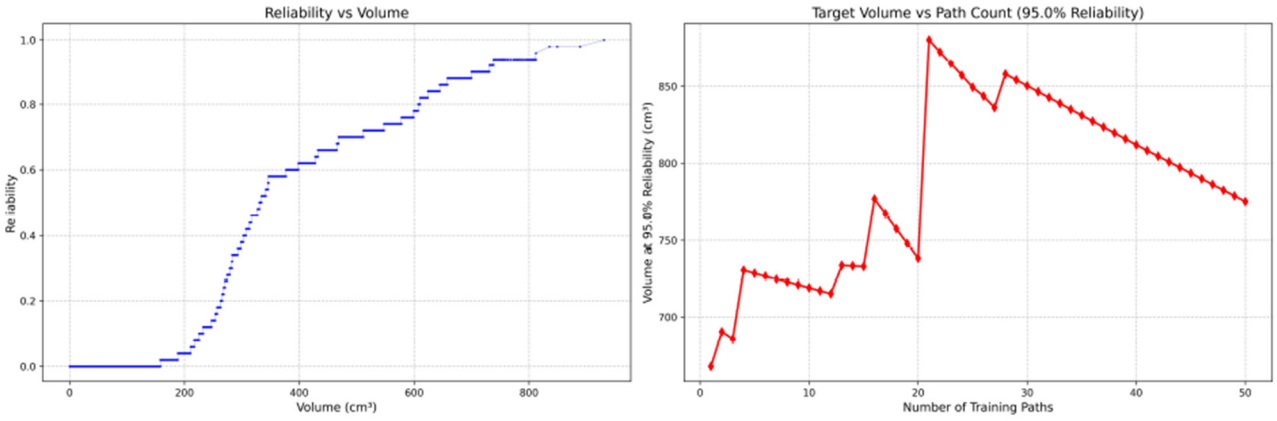

FIGURE 10
Reliability Volume (RV) results for student 4 (female). **(Left)** *R* vs. *V*; **(Right)** *V* vs. repetitions at $R = 0.95$.

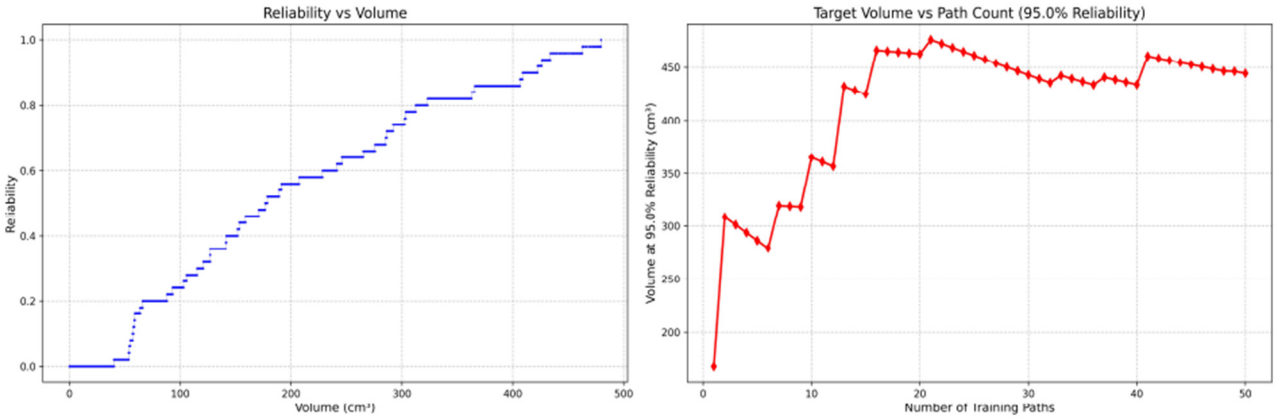

FIGURE 11
Reliability Volume (RV) results for surgical resident 1 (male). **(Left)** R vs. V; **(Right)** V vs. repetitions at $R = 0.95$.

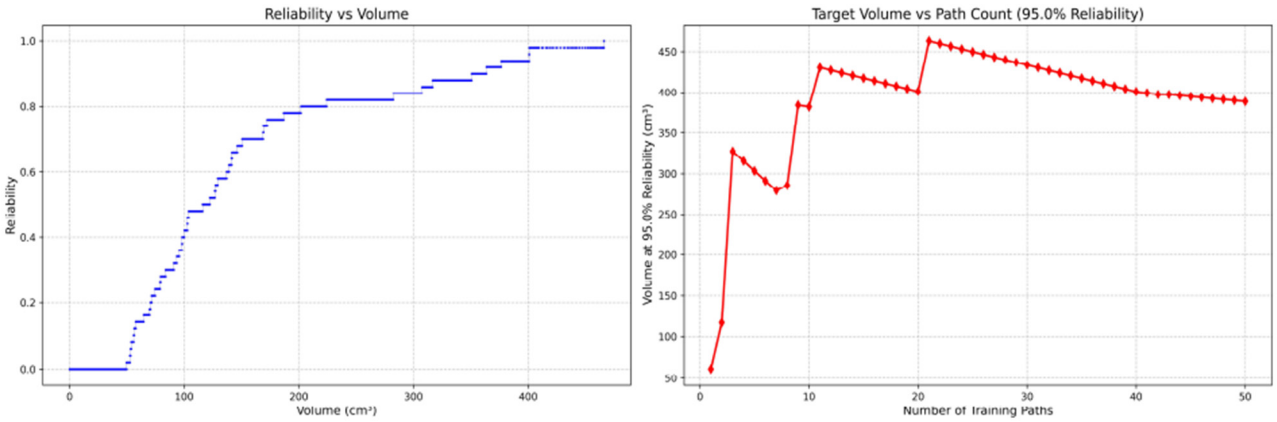

FIGURE 12
Reliability Volume (RV) results for surgical resident 2 (female). **(Left)** R vs. V; **(Right)** V vs. repetitions at $R = 0.95$.

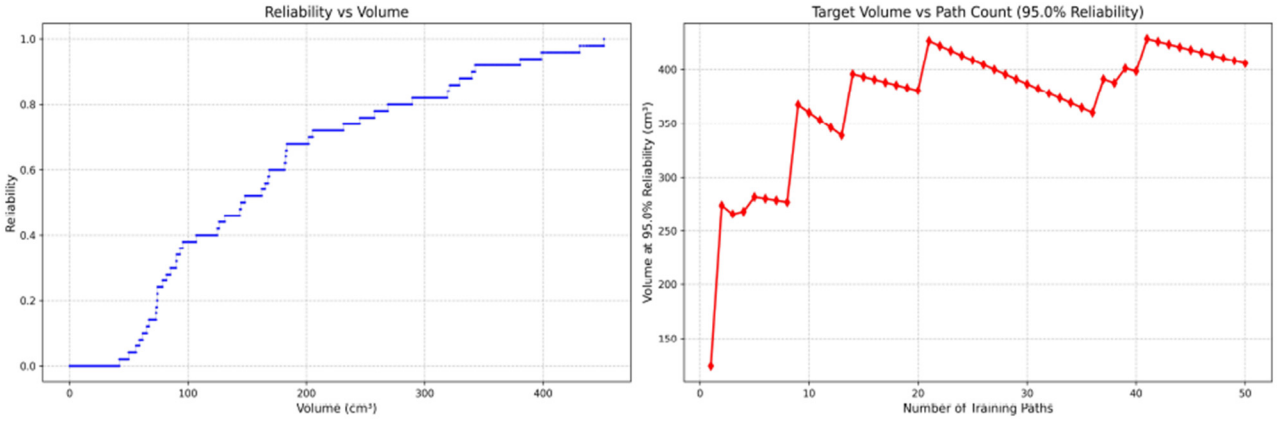

FIGURE 13
Reliability Volume (RV) results for surgical resident 3 (female). **(Left)** R vs. V; **(Right)** V vs. repetitions at $R = 0.95$.

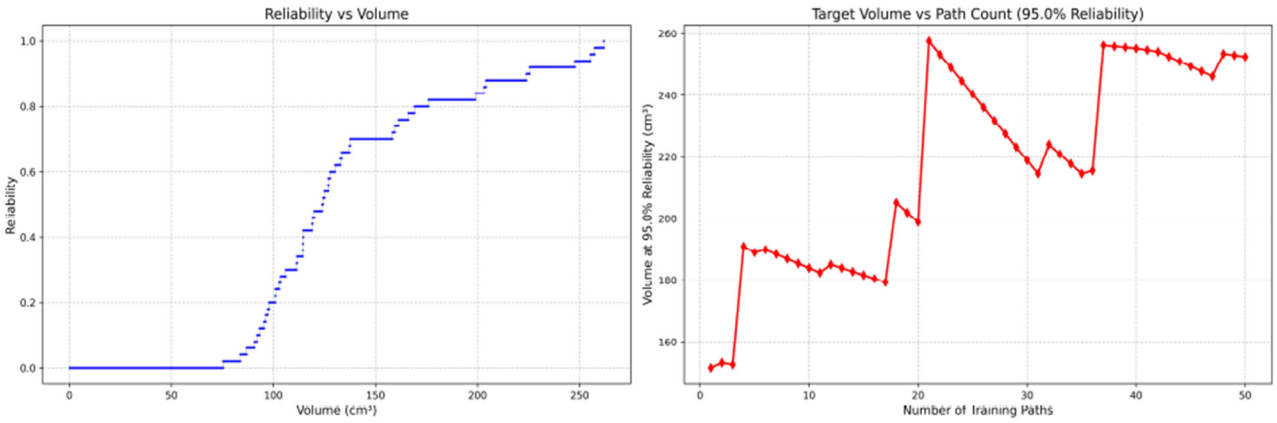

FIGURE 14
Reliability Volume (RV) results for attending surgeon 1 (female). **(Left)** *R* vs. *V*; **(Right)** *V* vs. repetitions at $R = 0.95$.

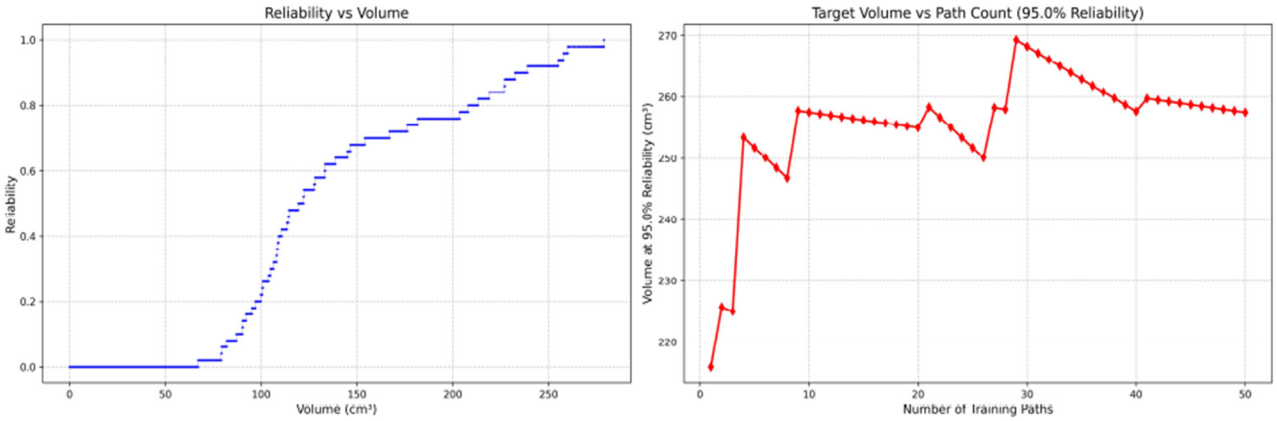

FIGURE 15
Reliability Volume (RV) results for attending surgeon 2 (male). **(Left)** *R* vs. *V*; **(Right)** *V* vs. repetitions at $R = 0.95$.

TABLE 2 Reliability Volume (RV) calculations for student 1 at different repetitions.

10 repetitions		25 repetitions		50 repetitions	
Reliability	Volume (cm^3)	Reliability	Volume (cm^3)	Reliability	Volume (cm^3)
0.90	427.84	0.96	427.84	0.90	427.84
0.90	427.26	0.96	427.26	0.90	427.26
0.80	427.18	0.92	427.18	0.88	427.18
0.80	427.12	0.92	427.12	0.88	427.12

TABLE 3 Reliability Volume (RV) calculations for student 2 at different repetitions.

10 repetitions		25 repetitions		50 repetitions	
Reliability	Volume (cm^3)	Reliability	Volume (cm^3)	Reliability	Volume (cm^3)
0.90	613.33	0.84	613.33	0.88	613.33
0.90	613.17	0.84	613.17	0.88	613.17
0.80	613.07	0.80	613.07	0.86	613.07
0.80	613.02	0.80	613.02	0.86	613.02

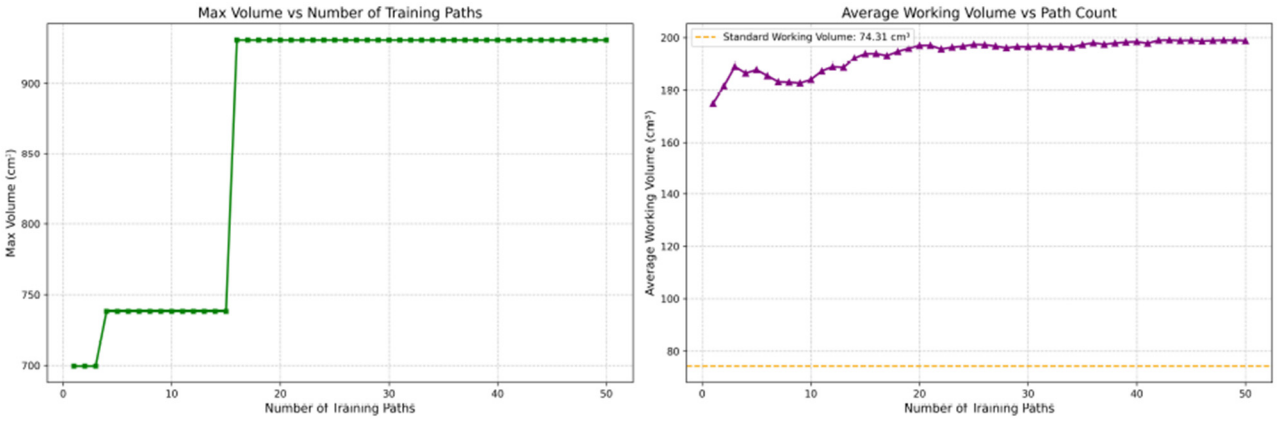

FIGURE 16
Comparison of Reliability Volume (RV) and conventional working volumes for student 2.

TABLE 4 Comparison of RV and conventional working volumes for student 2.

Number of training paths	Max RV volume (cm^3)	Working volume (cm^3)
10	738.23	196.13
20	930.56	216.59
30	930.56	195.13
40	930.56	204.63
50	930.56	190.08

critical future extension, as prolonged execution may also serve as an indicator of skill variability. While this study demonstrates the feasibility and practicality of the RV metric, the potential impacts of fatigue and other confounding factors require further investigation. Notably, moderating variables such as gender and prior health status were not included in the current analysis. Future research should therefore enroll larger and more diverse cohorts, integrate direct fatigue assessments, and evaluate additional clinical tasks. Furthermore, given that the number of repetitions was used as a proxy for actual training time in this study, integrating RV into automated real-time feedback systems could enhance training efficiency and skill retention by delivering immediate, actionable guidance (16, 17).

5 Conclusion

We propose Reliability Volume (RV), an integrated metric that combines trajectory similarity with an empirical reliability-based framework to assess surgical skill in repetitive, realistic training settings. RV quantifies both spatial precision and the probability of consistent task execution, addressing limitations of single-session metrics that overlook fatigue and performance drift. Evidence from knot-tying tasks demonstrates that RV effectively captures consistency over repetitions and reveals trade-offs between precision and reliability. Future work will broaden participant diversity, evaluate additional training scenarios, and investigate the integration of RV into automated real-time feedback systems.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

ZY: Writing – original draft. QL: Software, Investigation, Writing – review & editing. JZ: Data curation, Investigation, Resources, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was

supported by Yan'an University's Doctoral Startup Project (Grant No.: YDBK2020-14).

Acknowledgments

The authors thank Professor Jin Zhang for invaluable guidance and the reviewers for constructive feedback that improved this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Gen AI was used in the creation of this manuscript. To assist in language refinement and grammatical corrections.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Sargeant J, Armson H, Chesluk B, Dornan T, Eva K, Holmboe E, et al. The processes and dimensions of informed self-assessment: a conceptual model. *Acad Med*. (2010) 85:1212–20. doi: 10.1097/ACM.0b013e3181d85a4e

2. van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ*. (2005) 39:309–17. doi: 10.1111/j.1365-2929.2005.02094.x

3. Zia A, Essa I. Automated surgical skill assessment in RMIS training. *Int J Comput Assist Radiol Surg*. (2018) 13731–739. doi: 10.1007/s11548-018-1735-5

4. D'Angelo AL, Rutherford DN, Ray RD, Laufer S, Kwan C, Cohen ER, et al. Idle time: an underdeveloped performance metric for assessing surgical skill. *Am J Surg*. (2015) 209:645–51. doi: 10.1016/j.amjsurg.2014.12.013

5. Ahmidi N, Ishii M, Fichtinger G, Gallia GL, Hager G. An objective and automated method for assessing surgical skill in endoscopic sinus surgery using eye-tracking and tool-motion data. *Int Forum Allergy Rhinol*. (2012) 29:507–15. doi: 10.1002/alr.21053

6. Yamauchi Y, Yamashita J, Morikawa O, Hashimoto R, Mochimaru M, Fukui Y, et al. Surgical skill evaluation by force data for endoscopic sinus surgery training system. In: *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002*. Berlin; Heidelberg: Springer (2002). p. 44–51. doi: 10.1007/3-540-45786-0_6

7. Franklin GF. *Feedback Control of Dynamic Systems*. Menlo Park, CA: Addison-Wesley Longman Publishing Co, Inc. (1993).

8. Hamza H, Shabir D, Aboumarzouk O, Al-Ansari A, Shaban K, Navkar NV. Automated skills assessment in open surgery: a scoping review. *Eng Appl Artif Intell*. (2025) 153:110893. doi: 10.1016/j.engappai.2025.110893

9. Laufer S, D'Angelo AD, Kwan C, Ray RD, Yudkowsky R, Boulet JR, et al. Rescuing the clinical breast examination: advances in classifying technique and assessing physician competency. *Ann Surg*. (2017) 266:1069–74. doi: 10.1097/SLA.0000000000002024

10. Boyajian GP, Zulbaran-Rojas A, Najafi B, Atique MMU, Loor G, Gilani R, et al. Development of a sensor technology to objectively measure dexterity for cardiac surgical proficiency. *Ann Thorac Surg*. (2024) 117:635–43. doi: 10.1016/j.athoracsur.2023.07.013

11. D'Angelo AL, Rutherford DN, Ray RD, Laufer S, Mason A, Pugh CM. Working volume: validity evidence for a motion-based metric of surgical efficiency. *Am J Surg*. (2016) 211:445–50. doi: 10.1016/j.amjsurg.2015.10.005

12. Azari DP, Miller BL, Le BV, Greenberg CC, Radwin RG. Quantifying surgeon maneuevers across experience levels through marker-less hand motion kinematics of simulated surgical tasks. *Appl Ergon*. (2020) 87:103:136. doi: 10.1016/j.apergo.2020.103136

13. Keogh EJ, Palpanas T, Zordan VB, Gunopulos D, Cardle M. Indexing Large Human-Motion Databases. In: *Very Large Data Bases Conference*. Berlin, Heidelberg: Springer.

14. Ahmidi N, Tao L, Sefati S, Gao Y, Lea C, Haro BB, et al. A dataset and benchmarks for segmentation and recognition of gestures in robotic surgery. *IEEE Trans Biomed Eng*. (2017) 64:2025–41. doi: 10.1109/TBME.2016.2647680

15. Lam K, Chen J, Wang Z, Iqbal FM, Darzi A, Lo B, et al. Machine learning for technical skill assessment in surgery: a systematic review. *npj Digit Med*. (2022) 5:24. doi: 10.1038/s41746-022-00566-0

16. Boal MWE, Anastasiou D, Tesfai F, Ghamrawi W, Mazomenos E, Curtis N, et al. Evaluation of objective tools and artificial intelligence in robotic surgery technical skills assessment: a systematic review. *Br J Surg*. (2024) 111:znad331. doi: 10.1093/bjs/znad331

17. Rahimi AM, Uluç E, Hardon SF, Bonjer HJ, van der Peet DL, Daams F. Training in robotic-assisted surgery: a systematic review of training modalities and objective and subjective assessment methods. *Surg Endosc*. (2024) 38:3547–55. doi: 10.1007/s00464-024-10915-7

18. Kahol K, Leyba MJ, Deka M, Deka V, Mayes S, Smith M, et al. Effect of fatigue on psychomotor and cognitive skills. *Am J Surg*. (2008) 195:195–204. doi: 10.1016/j.amjsurg.2007.10.004

19. Reijmerink IM, van der Laan MJ, Wietasch JKG, Hooft L, Cnossen F. Impact of fatigue in surgeons on performance and patient outcome: systematic review. *Br J Surg*. (2023) 111:znad397. doi: 10.1093/bjs/znad397

20. Kolb D. *Experiential Learning: Experience As the Source of Learning and Development, Vol 1 of Journal of Business Ethics*. Englewood Cliffs, NJ: Prentice Hall. (1984).

21. Hatano G, Inagaki K. *Child Development and Education in Japan Child Development and Education in Japan*. New York, NY, US: W H Freeman Times Books Henry Holt & Co. (1986).

22. Kluger A, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol Bull*. (1996) 119:254–84. doi: 10.1037/0033-2909.119.2.254

23. Schön DA. *The Reflective Practitioner: How Professionals Think in Action*. New York, NY: Routledge (1992).

24. Wei P, Lu Z, Yuan X. Monte Carlo simulation for moment-independent sensitivity analysis. *Reliab Eng Syst Safety*. (2013) 110:60–7. doi: 10.1016/j.ress.2012.09.005

25. Zhang H, Dai H, Beer M, Wang W. Structural reliability analysis on the basis of small samples: an interval quasi-Monte Carlo method. *Mech Syst Signal Process*. (2013) 37:137–51. doi: 10.1016/j.ymssp.2012.03.001