

ORIGINAL PAPER

Initial performance of novices across two-dimensional, three-dimensional, and robot-assisted surgery

Piotr Kania¹⁻³, Szymon P. Kania⁴, Aleksandra Ciżyńska⁴, Zofia J. Kania⁴, Marcin Jarzemski⁵, Maria Baida¹, Paweł Marczuk¹, Markijan Kubis¹, Jakub Biedrzycki¹, Grzegorz Szewczyk^{2,6,7}, Maciej Salagierski³, Przemysław Mitura⁸

¹Department of Urology and Urological Oncology, St. John Paul II Masovian Regional Hospital, Siedlce, Poland

²Institute of Health Sciences, University of Siedlce, Poland

³Department of Urology and Urologic Oncology, Institute of Medical Sciences, Collegium Medicum, University of Zielona Gora, Poland

⁴Faculty of Medicine, Medical University of Lublin, Poland

⁵Department of Urology, Jan Bizieli University Hospital in Bydgoszcz, Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University, Torun, Poland

⁶Department of Obstetrics, Gynecology and Gynecologic Oncology, St. John Paul II Masovian Regional Hospital, Siedlce, Poland

⁷Department of Biophysics, Physiology and Pathophysiology, Medical University of Warsaw, Poland

⁸Department of Urology and Urological Oncology, Faculty of Medicine, Medical University of Lublin, Poland

Citation: Kania P, Kania SP, Ciżyńska A, et al. Initial performance of novices across two-dimensional, three-dimensional, and robot-assisted surgery. Cent European J Urol. 2026; doi: 10.5173/cej.2026.0371

Article history

Submitted: Sep. 16, 2026

Accepted: Sep. 27, 2026

Published online: Sep. 30, 2026

Corresponding author

Piotr Kania
Department of Urology
and Urological Oncology
St. John Paul II Masovian
Regional Hospital
26 Poniatowskiego St.,
08-110 Siedlce, Poland
pkania@szpital.siedlce.pl

Introduction Initial performance in minimally invasive surgery may depend on the platform used by complete novices. We compared two-dimensional laparoscopy (2DL), three-dimensional laparoscopy (3DL), and robot-assisted surgery (RAS) using selected European Basic Laparoscopic Urological Skills (E-BLUS) tasks. **Material and methods** In this prospective multicentre comparative simulation study, 116 adults without prior practical laparoscopic or robotic experience were allocated to 2DL (n = 39), 3DL (n = 34), or RAS (n = 43). Allocation was mainly determined by equipment availability and was only partially randomised. Participants performed Peg Transfer followed by Needle Guidance. Outcomes included completion time, Peg Transfer errors, E-BLUS-derived performance criteria, Raw NASA-TLX workload, and ergonomic discomfort.

Results Of 116 participants, 104 completed the protocol. RAS was faster than both laparoscopic platforms for both tasks (p < 0.001), whereas 2DL–3DL differences were inconsistent. Among protocol completers, 41.9% of RAS participants met both E-BLUS performance criteria, compared with 3.8% with 3DL and none with 2DL. Peg Transfer errors were more frequent with 2DL than with 3DL or RAS. The Needle Guidance-to-Peg Transfer time ratio was 2.32 with 2DL, 1.70 with 3DL, and 1.69 with RAS (interaction p = 0.002). Overall workload and ergonomic discomfort were lowest with RAS (p < 0.001 and p = 0.002, respectively).

Conclusions RAS provided the most favourable initial performance profile and lowest subjective burden, while 3DL–2DL differences were less consistent and task-dependent. Longitudinal studies are needed to determine whether these early differences persist during training.

Key Words: laparoscopy ↔ robot-assisted surgery ↔ minimally invasive surgery
↔ laparoscopic simulation ↔ surgical education ↔ novice training

INTRODUCTION

Minimally invasive surgery has become an integral component of contemporary surgical practice. Over the past two decades, numerous surgical specialties have undergone a profound transition from open

surgery to conventional two-dimensional laparoscopy (2DL) and, more recently, to robot-assisted surgery (RAS) [1]. A major intermediate step in this evolution was the introduction of three-dimensional laparoscopy (3DL), which improved depth perception, spatial orientation, and hand-eye coordination

compared with conventional 2DL [2, 3]. However, adoption of 3DL technology has not been universal. While some centres embraced stereoscopic vision, others favoured advancements in image quality and resolution, such as ultra-high-definition imaging, while continuing to use planar 2DL visualisation [4]. The introduction of RAS represented a further technological advancement. In addition to three-dimensional vision, robotic systems offer articulated wristed instruments with seven degrees of freedom, tremor filtration, motion scaling, enhanced instrument dexterity, and improved surgeon ergonomics. Collectively, these features aim to overcome several limitations of conventional laparoscopy, facilitate the performance of complex minimally invasive procedures, and potentially accelerate the acquisition of basic minimally invasive surgical skills [5, 6].

As minimally invasive technologies continue to evolve, surgical training should adapt accordingly. The development of future surgeons increasingly begins during undergraduate medical education, where simulation-based training provides a safe environment for the acquisition of fundamental psychomotor skills before exposure to clinical practice. Consequently, two related questions arise: on which platform should novice trainees begin developing minimally invasive surgical skills, and how do these platforms differ in the physical and cognitive workload imposed during initial training? Previous studies suggest that robotic assistance may reduce some physical and cognitive demands compared with conventional laparoscopy, whereas direct comparisons with 3DL remain scarce; however, these effects may depend on operator experience and task complexity [7, 8].

The increasing adoption of minimally invasive and robot-assisted procedures also reshapes surgical training. As selected open and conventional laparoscopic operations become less frequently performed, opportunities to acquire and maintain these skills – and access to sufficiently experienced mentors – may gradually diminish. Surveys of urology trainees and educators have demonstrated substantial variation in hands-on exposure and expectations regarding surgical training [9, 10]. Moreover, declining exposure has been documented for selected open procedures [11]. At the same time, training curricula should prepare future surgeons for a wide range of clinical scenarios, including the management of intraoperative complications and situations requiring conversion between surgical approaches. Therefore, the goal of modern surgical education extends beyond achieving technical proficiency on a single platform and includes the development of transferable psychomotor skills that can be applied across different minimally invasive environments.

To provide an objective and standardised assessment of basic laparoscopic competence, the European Basic Laparoscopic Urological Skills (E-BLUS) curriculum was developed and validated as a urology-specific training and assessment programme [12]. Although designed for conventional laparoscopy, its tasks assess fundamental visuospatial, bimanual, and instrument-handling skills relevant across minimally invasive surgical platforms. Peg Transfer and Needle Guidance have also been used in previous E-BLUS-based educational research as objective measures of trainee performance [13]. Selected E-BLUS tasks therefore provided a common and reproducible framework for comparing initial performance with 2DL, 3DL, and RAS.

Previous studies involving inexperienced participants have generally compared 2DL with 3DL [14] or conventional laparoscopy with RAS [15]. Direct comparisons of all three platforms under the same standardised protocol remain limited and have usually focused on an isolated task or included participants with different levels of surgical experience [16]. Gaming experience has been proposed as a possible correlate of novice laparoscopic performance [17], while similar associations have been investigated for musical-instrument experience [18]; however, the available evidence remains inconsistent. The present study therefore directly compared initial task performance across 2DL, 3DL, and RAS in participants without previous practical laparoscopic or robotic experience. In addition to completion time, errors, and attainment of predefined proficiency criteria, we assessed subjective workload, including mental and physical demands, and physical discomfort. We also explored associations between task performance and subjective workload, physical discomfort, gaming exposure, and musical-instrument experience.

MATERIAL AND METHODS

Study design and participants

This prospective, multicentre, three-arm parallel-group, comparative, simulation-based study was conducted between January and June 2026 at three academic centres. Participants were adult volunteers, predominantly students enrolled in medical and health-related programmes, with no previous practical experience in laparoscopic or robot-assisted surgery.

The study protocol comprised two selected E-BLUS tasks – Peg Transfer and Needle Guidance – performed using a single assigned platform: 2DL, 3DL, or RAS.

Platform allocation was determined by equipment availability during each study session and was randomised only when more than one platform was simultaneously available; consequently, allocation was only partially randomised, and not every participant had an equal opportunity to be assigned to each of the three study platforms. All three platforms were never simultaneously available within a single study session. Centre 1 enrolled 55 participants (2DL, $n = 12$; 3DL, $n = 27$; RAS, $n = 16$), Centre 2 enrolled 45 (2DL, $n = 19$; 3DL, $n = 7$; RAS, $n = 19$), and Centre 3 enrolled 16 (2DL, $n = 8$; RAS, $n = 8$); 3DL was not available at Centre 3. Overall, 86 of 116 participants (74.1%) were randomised between the two platforms available during the corresponding session, whereas the remaining 30 participants (25.9%) were allocated according to equipment availability. The distribution of the full allocated cohort by centre and platform is presented in Supplementary Table S1. Therefore, the study was considered a prospective comparative simulation-based study rather than a fully randomised parallel-group trial and was reported with reference to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement, supplemented by the reporting guidelines for health care simulation research [19, 20]. Instructors were centre-specific because of the geographical separation of the participating sites. Before study initiation, all instructors were trained by an E-BLUS examiner and followed the same standardised E-BLUS examination procedure. All participants viewed the same official E-BLUS instructional video, received the same standardised pre-test instructions, and received no coaching or performance-related feedback during the recorded attempt. Equipment was standardised as far as possible. The same three-dimensional laparoscopic system, IMAGE1 S (KARL STORZ, Tuttlingen, Germany), was used for 3DL, multiport da Vinci surgical systems (Intuitive Surgical, Sunnyvale, CA, USA) were used for RAS, and standard two-dimensional laparoscopic video equipment available at each centre was used for 2DL. The same platform-specific instrument sets were used across centres, and the E-BLUS task configuration was standardised. Baseline data included age, sex, dominant hand, height, weight, visual correction, gaming exposure, and experience playing a musical instrument. Gaming exposure was assessed according to estimated weekly gaming time during the preceding year, duration of previous gaming involvement, and game type. Musical experience was defined as playing a musical instrument for more than six months during the preceding 3 years.

Study procedure and assessment

Prior to the test session, all participants received an email containing links to official European School of Urology instructional videos demonstrating the selected E-BLUS tasks [21]. Before task performance, study participants underwent a standardised 15-minute instructor-led introduction covering the procedure, task objectives, and basic handling of the instruments. This was followed by a 5-minute hands-on familiarisation period allowing free manipulation of the instruments on the assigned platform, without completion of either study task.

Two standardised tasks from the European Basic Laparoscopic Urological Skills programme were chosen: Peg Transfer and Needle Guidance [22]. These tasks were selected because the investigators considered them feasible for complete novices following brief instruction and familiarisation. They were performed in a predefined fixed order, with Peg Transfer followed by Needle Guidance. The same sequence was maintained for all participants to ensure procedural consistency. Needle Guidance was performed second because it was considered the more technically demanding of the two tasks and more dependent on spatial orientation and depth perception.

The protocol specified one recorded attempt at each task, without performance coaching. Completion time was recorded by an instructor for each completed task. In accordance with the E-BLUS assessment rules, errors were recorded only for Peg Transfer, as Needle Guidance does not include a separate error count.

Under the standard E-BLUS procedure, the first Peg Transfer error requires the task to be restarted, whereas a second error results in failure. In the present study, the task was not interrupted or restarted after an error. Instead, participants were allowed to complete the attempt, and the total number of errors was recorded to enable an additional quantitative assessment of accuracy.

The published proficiency thresholds were used to classify task performance: completion within 126 seconds with no more than one error for Peg Transfer and completion within 268 seconds for Needle Guidance. Owing to the modified administration of Peg Transfer, criterion attainment represented fulfilment of the published time and error thresholds under the study protocol rather than a formal E-BLUS examination result.

Immediately after completing the second task, participants completed the Raw National Aeronautics and Space Administration Task Load Index

(NASA-TLX) questionnaire, comprising six domains: mental demand, physical demand, temporal demand, perceived performance, effort, and frustration [23, 24]. Each domain was rated on a scale from 0 to 100, with higher scores indicating greater workload or poorer perceived performance. The unweighted version was used without the pairwise weighting procedure of the original instrument, and the overall Raw NASA-TLX score was calculated as the arithmetic mean of the six domain scores [24].

Physical and ergonomic discomfort was additionally rated in four anatomical regions using an 11-point numerical rating scale ranging from 0, indicating no discomfort, to 10, indicating the greatest possible discomfort. Discomfort ratings were analysed separately and were not included in the Raw NASA-TLX score.

Study outcomes

The primary outcomes were task completion time and attainment of the predefined proficiency criterion for each E-BLUS task. Combined completion time was calculated as the sum of Peg Transfer and Needle Guidance completion times. Criterion attainment was assessed separately for Peg Transfer and Needle Guidance and as a combined outcome, defined as meeting the criteria for both tasks during the same study session. The number of Peg Transfer errors was analysed as an additional objective measure of accuracy.

The principal secondary outcome was the overall Raw NASA-TLX score. The overall discomfort score, calculated from ratings in the four anatomical regions, was considered an additional secondary outcome. The six individual NASA-TLX domains and the discomfort ratings for each anatomical region were analysed as exploratory outcomes. Associations between subjective workload, discomfort, and objective task performance were also explored.

Statistical analysis

All statistical tests were two-sided, with a significance level of 0.05. Effect estimates were reported with 95% confidence intervals (CIs). Continuous variables were summarised as means with standard deviations or medians with interquartile ranges (IQRs), depending on their distribution, whereas categorical variables were presented as counts and percentages. Baseline characteristics were summarised descriptively, and differences between platform groups were assessed using standardised mean differences.

Because completion times were right-skewed, they were log-transformed and analysed using linear regression models with robust standard errors. The primary models included platform group and study centre, while adjusted models additionally included age, sex, and current weekly gaming exposure. Results were reported as ratios of geometric means with 95% confidence intervals and as model-based absolute differences in completion time expressed in seconds. Kruskal–Wallis tests followed by Dunn’s pairwise comparisons were performed as sensitivity analyses.

Peg Transfer error counts were analysed using Poisson or negative binomial regression, as appropriate. Attainment of the predefined proficiency criteria was compared between platform groups using appropriate categorical-data methods. Analyses of task performance, proficiency, workload, and discomfort were based on participants who completed the study protocol. Non-completion was analysed separately as an outcome of the allocated cohort. Non-completion of the protocol was additionally compared between platforms using the Fisher exact test, with Newcombe risk differences and Firth penalised logistic regression.

The interaction between task and platform was examined using a mixed-effects model with a participant-level random intercept. Raw NASA-TLX and discomfort scores were compared between platform groups using regression models, with rank-based methods used as sensitivity analyses. Associations of workload and discomfort with task performance were assessed using Spearman correlations, including partial correlations adjusted for platform, study centre, age, sex, and current weekly gaming exposure. The Holm procedure was applied to confirmatory comparisons and the Benjamini–Hochberg procedure to exploratory analyses. Missing covariate data were addressed by multiple imputation in sensitivity analyses.

No formal a priori sample-size calculation was performed. All analyses were performed in R version 4.6.1 (R Foundation for Statistical Computing, Vienna, Austria).

Bioethical standards

The study protocol was approved by the relevant institutional Bioethics Committee (approval no. 24/2025, dated 22 December, 2025). All participants received written information about the study and provided written informed consent before enrolment. Individual study identification numbers were assigned, and all data used for analysis were anonymised.

RESULTS

Participants and baseline characteristics

Of 116 participants without previous practical experience in laparoscopic or robot-assisted surgery, 39 were allocated to 2DL, 34 to 3DL, and 43 to RAS. Overall, 104 completed the protocol and were included in analyses of completion time, Peg Transfer errors, proficiency, workload, and ergonomic discomfort (Figure 1). Twelve participants withdrew: four from 2DL (4/39; 10.3%), two during each task, and eight from 3DL (8/34; 23.5%), all during Needle Guidance. All RAS participants completed both tasks (0/43 withdrawals). Non-completion differed across platforms ($p = 0.001$); compared with RAS, from which no participant withdrew, the risk of non-completion was higher by 23.5 percentage points with 3DL (95% CI: 9.7–40.0) and by 10.3

points with 2DL (Table 4). Non-completion is reported here as a separate outcome of the allocated cohort, distinct from performance analyses, which were based on the 104 participants who completed the study protocol. The analysed cohort had a median age of 23 years (IQR: 21–24), and 68 of 103 participants (66.0%) were female. Most participants were right-handed (98/104; 94.2%); 46 (44.2%) used visual correction, 66 (63.5%) reported gaming exposure, and 39 (37.5%) reported musical-instrument experience. Baseline characteristics varied across platform groups, partly reflecting centre-specific platform availability (Table 1).

Task completion time

Completion times differed across platforms for Peg Transfer, Needle Guidance, and both tasks combined ($p < 0.001$ for each; Figure 2 and Table 2).

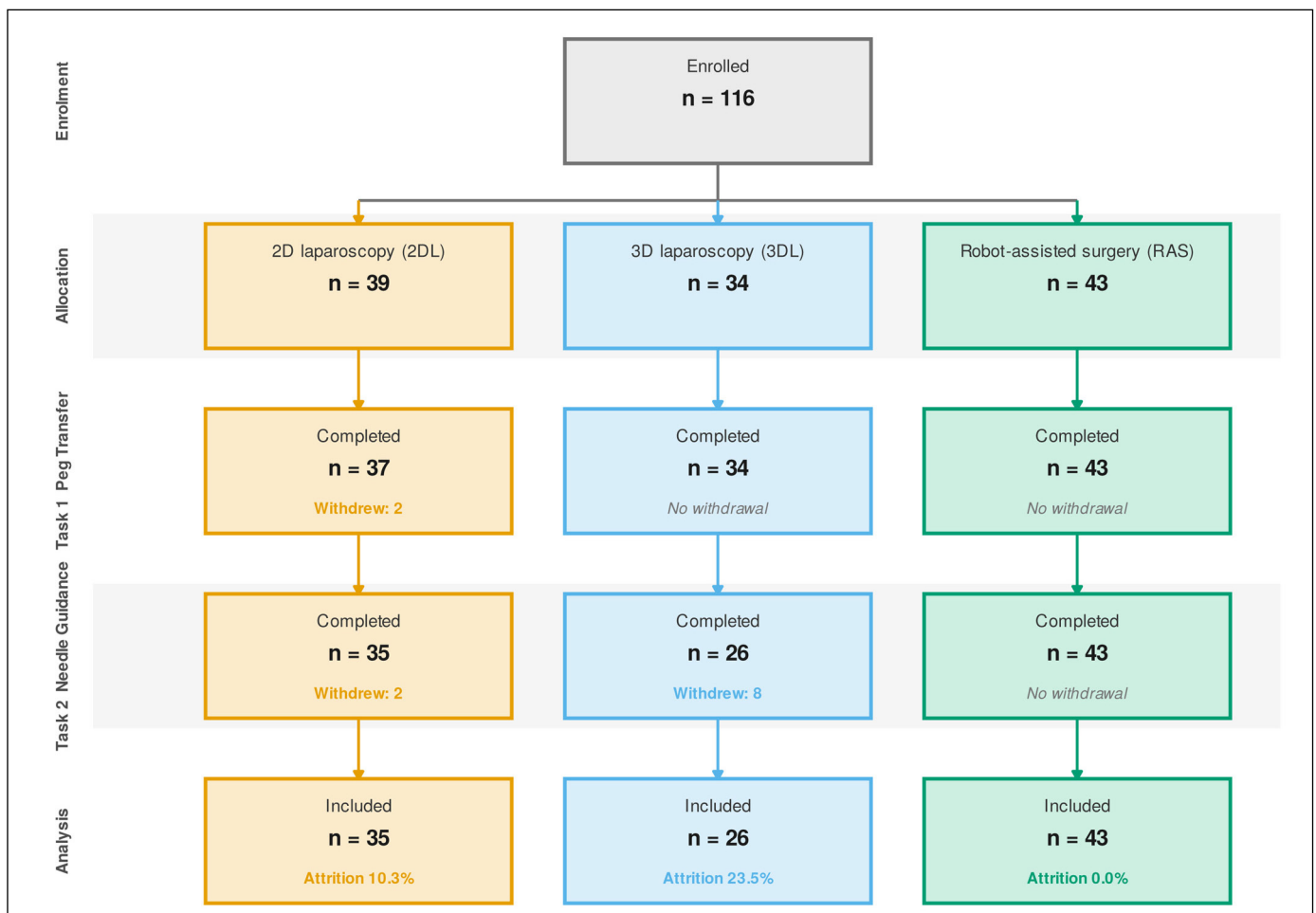


Figure 1. Participant flow through the study.

The diagram shows the number of participants allocated to each platform, completing Peg Transfer and Needle Guidance, withdrawing during the study, and contributing to the final analyses.

2DL – two-dimensional laparoscopy; 3DL – three-dimensional laparoscopy; RAS – robot-assisted surgery

RAS was faster than both laparoscopic platforms for both tasks, whereas no clear difference between 2DL and 3DL was demonstrated.

For Peg Transfer, the median completion time was shortest with RAS at 125 seconds (IQR: 103–151), compared with 184 seconds (IQR: 154–230) with 2DL and 198 seconds (IQR: 146–279) with 3DL. After adjustment, completion times were 1.58 times longer with 2DL than with RAS (95% CI: 1.35–1.85; $p < 0.001$) and 1.56 times longer with 3DL than with RAS (95% CI: 1.32–1.85; $p < .001$). The two laparoscopic platforms were similar (ratio 1.01, 95% CI: 0.81–1.25; $p = 0.92$).

The difference between RAS and the laparoscopic platforms was more pronounced for Needle Guidance. Median completion times were 211 seconds (IQR: 166–267) with RAS, 328 seconds (IQR: 288–408)

with 3DL, and 394 seconds (IQR: 285–644) with 2DL. Adjusted completion times were 1.60 times longer with 3DL than with RAS (95% CI: 1.34–1.91; $p < 0.001$) and 2.14 times longer with 2DL than with RAS (95% CI: 1.75–2.62; $p < 0.001$). The comparison between 2DL and 3DL remained inconclusive: the adjusted parametric model favoured 3DL (ratio 1.34, 95% CI: 1.07–1.67; $p = 0.03$), whereas the prespecified rank-based analysis did not show a difference ($p = 0.66$).

Peg Transfer errors and proficiency attainment

Under the E-BLUS assessment criteria, errors are recorded for Peg Transfer only; Needle Guidance is assessed by completion time alone. Participants using 2DL made more errors than those using

Table 1. Baseline characteristics of the 104 participants who completed the study protocol, by allocated platform

Characteristic	Overall (n = 104)	2DL (n = 35)	3DL (n = 26)	RAS (n = 43)	2DL vs 3DL	2DL vs RAS	3DL vs RAS
Demographic and anthropometric characteristics							
Age, years, median (IQR)	23.0 (21.0–24.0)	22.0 (20.5–23.0)	23.5 (22.0–27.5)	23.0 (21.0–24.0)	0.604	0.328	0.362
Sex, female, n (%)	68 (66.0)	23 (65.7)	16 (61.5)	29 (69.0)	0.087	0.071	0.158
Body mass index, median (IQR)	22.6 (20.1–24.7)	22.7 (20.2–24.6)	22.6 (20.3–24.0)	22.5 (20.2–24.8)	0.193	0.213	0.012
Physical and sensory attributes							
Handedness, n (%)					0.297	0.173	0.386
Right-handed	98 (94.2)	33 (94.3)	23 (88.5)	42 (97.7)			
Left-handed	5 (4.8)	2 (5.7)	2 (7.7)	1 (2.3)			
Ambidextrous	1 (1.0)	0 (0.0)	1 (3.8)	0 (0.0)			
Visual status, corrected, n (%)	46 (44.2)	21 (60.0)	7 (26.9)	18 (41.9)	0.708	0.369	0.318
State on the day of the session							
Sleep on the preceding night, h, mean (SD)	6.4 (1.5)	6.5 (1.4)	5.9 (1.5)	6.6 (1.4)	0.395	0.093	0.483
Caffeine before the session, yes, n (%)	52 (50.0)	19 (54.3)	13 (50.0)	20 (46.5)	0.086	0.156	0.070
Nicotine before the session, yes, n (%)	16 (15.5)	6 (17.1)	7 (26.9)	3 (7.1)	0.238	0.310	0.545
Prior exposure relevant to psychomotor skill							
Any gaming exposure, yes, n (%)	66 (63.5)	24 (68.6)	16 (61.5)	26 (60.5)	0.148	0.170	0.022
Current gaming, h/wk, median (IQR)	1.5 (0.0–5.0)	2.0 (0.0–4.5)	0.5 (0.0–4.0)	2.0 (0.0–6.5)	0.145	0.201	0.355
Gaming experience, y, median (IQR)	3.5 (0.0–10.0)	4.0 (0.0–8.5)	3.0 (0.0–8.0)	4.5 (0.0–10.0)	0.190	0.102	0.297
Cumulative gaming exposure, h, median (IQR)	364 (0–2,574)	312 (0–1,846)	104 (0–2,496)	468 (0–3,224)	0.189	0.196	0.409
Gaming exposure category, cumulative, n (%)					0.619	0.306	0.306
No exposure	41 (40.2)	12 (34.3)	12 (48.0)	17 (40.5)			
Moderate	31 (30.4)	15 (42.9)	4 (16.0)	12 (28.6)			
High	30 (29.4)	8 (22.9)	9 (36.0)	13 (31.0)			
Gaming exposure category, duration, n (%)					0.097	0.350	0.318
No exposure	36 (35.3)	11 (31.4)	9 (36.0)	16 (38.1)			
Low	28 (27.5)	12 (34.3)	8 (32.0)	8 (19.0)			
High	38 (37.3)	12 (34.3)	8 (32.0)	18 (42.9)			
Plays a musical instrument, yes, n (%)	39 (37.5)	14 (40.0)	12 (46.2)	13 (30.2)	0.125	0.206	0.332

Data are presented as mean (SD), median (IQR), or n (%), as appropriate. Percentages are based on non-missing denominators. SMDs are absolute pairwise values; for variables with more than two levels, the multivariate SMD is shown on the header row.

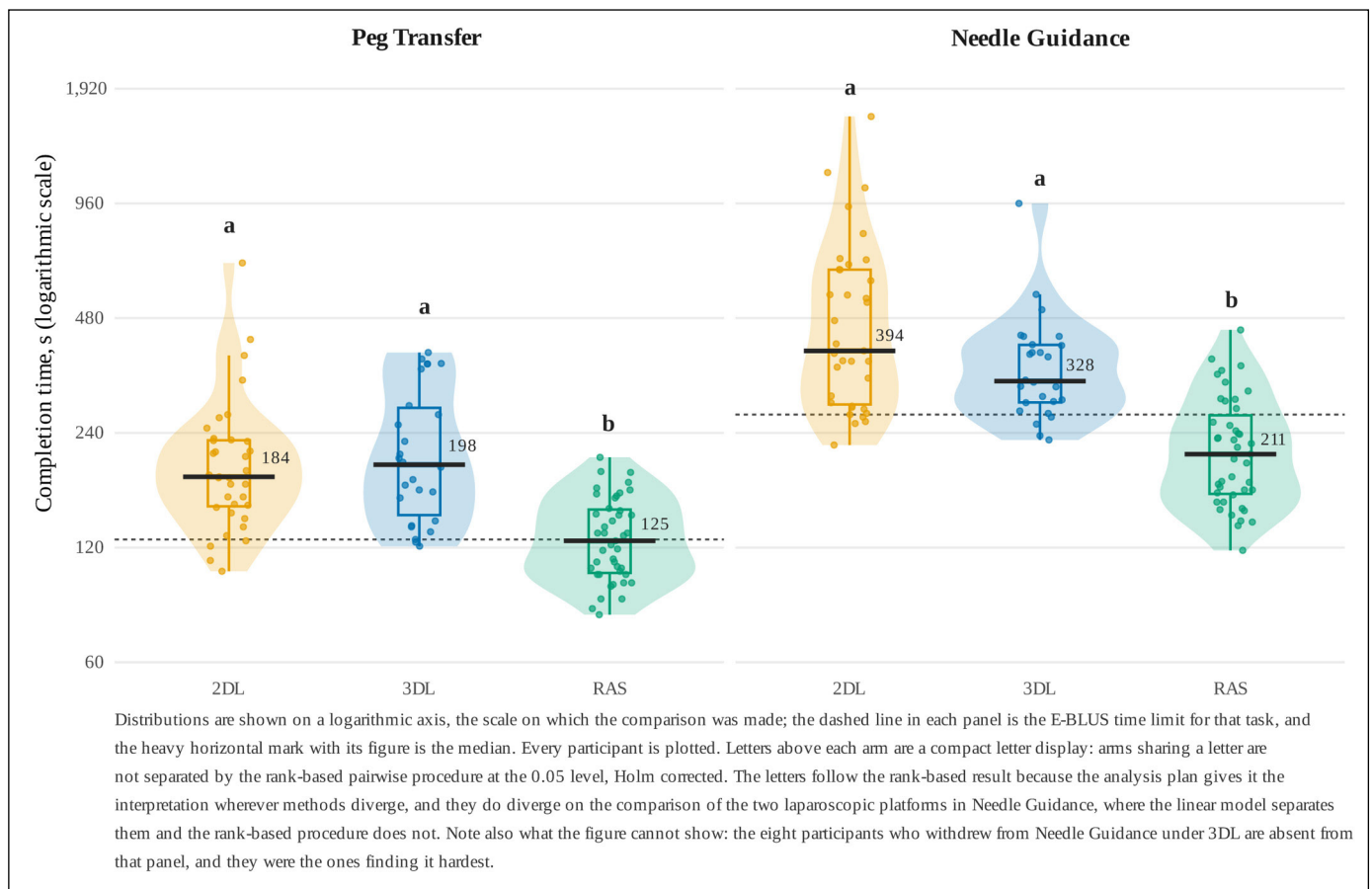
2DL – two-dimensional laparoscopy; 3DL – three-dimensional laparoscopy; IQR – interquartile range; RAS – robot-assisted surgery; SD – standard deviation;

SMD – standardised mean difference

Table 2. Task completion time and its pairwise contrasts, crude and adjusted

Contrast	Crude ratio (95% CI)	p-value	Adjusted ratio (95% CI)	p-value	Cliff delta (95% CI)
Peg Transfer time, median (IQR) 2DL 184 (154–230)^a, 3DL 198 (146–279)^a, RAS 125 (103–151)^b; across arms p <0.001					
2DL vs 3DL	0.94 (0.77–1.15)	.557	1.01 (0.81–1.25)	.920	–0.07 (–0.36–0.24)
2DL vs RAS	1.55 (1.33–1.81)	<.001	1.58 (1.35–1.85)	<.001	0.69 (0.48–0.83)
3DL vs RAS	1.65 (1.39–1.96)	<.001	1.56 (1.32–1.85)	<.001	0.71 (0.49–0.85)
Needle Guidance time, median (IQR) 2DL 394 (285–644)^a, 3DL 328 (288–408)^a, RAS 211 (166–267)^b; across arms p <0.001					
2DL vs 3DL	1.29 (1.04–1.59)	.056	1.34 (1.07–1.67)	.030	0.26 (–0.04–0.51)
2DL vs RAS	2.13 (1.75–2.59)	<.001	2.14 (1.75–2.62)	<.001	0.81 (0.66–0.91)
3DL vs RAS	1.65 (1.41–1.94)	<.001	1.60 (1.34–1.91)	<.001	0.74 (0.53–0.87)
Combined task time, median (IQR) 2DL 625 (463–827)^a, 3DL 576 (487–666)^a, RAS 335 (278–404)^b; across arms p <0.001					
2DL vs 3DL	1.15 (0.97–1.37)	.218	1.22 (1.02–1.45)	.063	0.18 (–0.13–0.46)
2DL vs RAS	1.94 (1.65–2.28)	<.001	1.96 (1.66–2.31)	<.001	0.85 (0.71–0.94)
3DL vs RAS	1.68 (1.48–1.91)	<.001	1.61 (1.39–1.86)	<.001	0.85 (0.68–0.94)

Ratios >1 indicate longer completion time with the first-listed platform. Adjusted models included centre, age, sex, and current weekly gaming exposure. P-values were Holm-adjusted; confidence intervals were unadjusted. Cliff delta is reported with bias-corrected and accelerated bootstrap confidence intervals. 2DL – two-dimensional laparoscopy; 3DL – three-dimensional laparoscopy; CI – confidence interval; IQR – interquartile range; RAS – robot-assisted surgery

**Figure 2.** Task completion times by platform among the 104 participants who completed the protocol.

Individual observations are shown on a logarithmic scale. Violin plots represent the distribution, boxes the interquartile range, bold horizontal lines the median, and dashed lines the E-BLUS time limits of 126 seconds for Peg Transfer and 268 seconds for Needle Guidance. Different letters indicate statistically significant differences between platforms in the prespecified rank-based pairwise analysis after correction for multiple comparisons.

2DL – two-dimensional laparoscopy; 3DL – three-dimensional laparoscopy; RAS – robot-assisted surgery; E-BLUS – European Basic Laparoscopic Urological Skills

either 3DL or RAS, whereas error rates were similar between 3DL and RAS. Mean error counts were 1.14 ± 1.26 with 2DL, 0.54 ± 0.65 with 3DL, and 0.37 ± 0.82 with RAS ($p = 0.005$). Compared with 3DL and RAS, the error rate with 2DL was 2.12-fold higher (95% CI: 1.19–3.79; $p = 0.02$) and 3.07-fold higher (95% CI: 1.46–6.45; $p = 0.009$), respectively; 3DL and RAS did not differ significantly (ratio 1.45, 95% CI: 0.66–3.19; $p = 0.36$).

Among participants who completed the protocol, proficiency was attained substantially more often with RAS than with either laparoscopic platform, whereas attainment rates were similarly low with 3DL and 2DL. For Peg Transfer, proficiency was

achieved by 48.8% of participants with RAS (21/43), 11.5% with 3DL (3/26), and 11.4% with 2DL (4/35). For Needle Guidance, the corresponding rates were 74.4% (32/43), 15.4% (4/26), and 14.3% (5/35), respectively. Both criteria were met during the same session by 41.9% of participants with RAS (18/43), 3.8% with 3DL (1/26), and 0% with 2DL (0/35). Overall between-platform differences were significant for Peg Transfer, Needle Guidance, and combined criterion attainment ($p < 0.001$ for each).

The task-by-platform interaction was therefore examined separately to assess whether the relative platform effect changed with increasing task demands.

Table 3. Pairwise comparison of the relative increase in completion time from Peg Transfer to Needle Guidance across platforms

Contrast	Ratio of difficulty penalties (95% CI)	P-value, uncorrected	P-value, corrected
2DL vs 3DL	1.37 (1.10–1.69)	0.004	0.008
2DL vs RAS	1.37 (1.14–1.65)	<0.001	0.003
3DL vs RAS	1.00 (0.82–1.23)	0.985	0.985

Ratios above 1 indicate a greater difficulty penalty for the first-listed platform. P values were corrected for multiple pairwise comparisons using the Holm procedure. CI – confidence interval; 2DL – two-dimensional laparoscopy; 3DL – three-dimensional laparoscopy; RAS – robot-assisted surgery

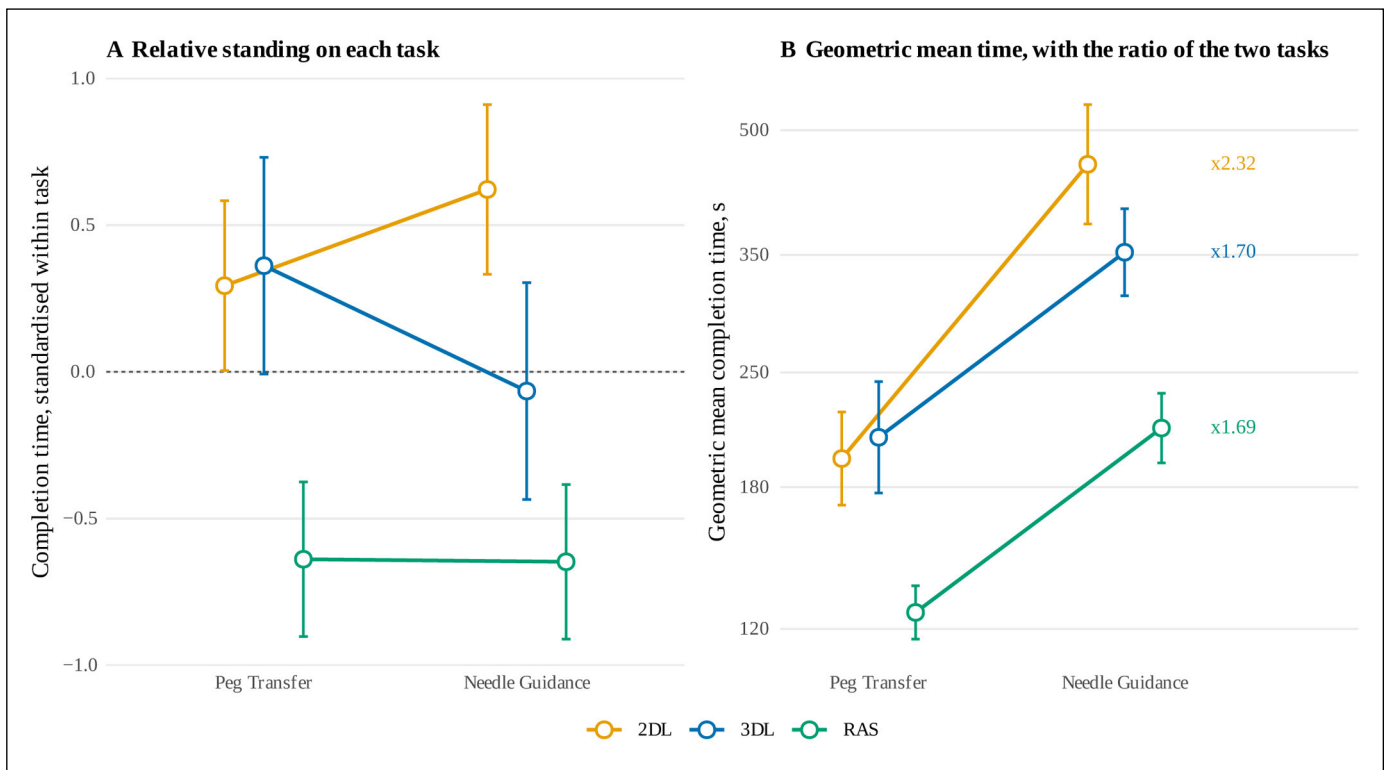


Figure 3. Change in relative performance from the easier task to the harder one, by platform.

Panel A presents completion times standardised within each task, showing changes in the relative performance of each platform. Panel B presents geometric mean completion times with 95% confidence intervals and the within-platform Needle Guidance-to-Peg Transfer time ratios: 2.32 for 2DL, 1.70 for 3DL, and 1.69 for RAS. The task-by-platform interaction was significant ($\chi^2 = 12.26$, $df = 2$; $p = 0.002$).

CI – confidence interval; 2DL – two-dimensional laparoscopy; 3DL – three-dimensional laparoscopy; df – degrees of freedom; RAS – robot-assisted surgery

Change in relative performance with increasing task difficulty

As task difficulty increased from Peg Transfer to Needle Guidance, which is more dependent on spatial orientation, the relative performance of 3DL improved, whereas that of 2DL worsened. RAS was faster in both tasks, but its relative advantage remained broadly similar. This differential change across platforms was significant (task-by-platform interaction: $\chi^2 = 12.26$, 2 df; $p = 0.002$; Figure 3 and Table 3). These two tasks differ not only in depth-perception demand but also in duration, motor sequence, precision requirements, and fixed order, Needle Guidance always being performed second; the observed interaction therefore reflects several dimensions of task difficulty rather than depth perception alone, and it is exploratory. Needle Guidance took 2.32 times longer than Peg Transfer with 2DL, compared with 1.70 times with 3DL and 1.69 times with RAS. The relative increase in completion time was 37% greater with 2DL than with either 3DL (ratio, 1.37; 95% CI: 1.10–1.69; adjusted $p = 0.008$) or RAS (ratio, 1.37; 95% CI: 1.14–1.65; adjusted $p = 0.003$), whereas it did not differ between 3DL and RAS (ratio, 1.00; 95% CI: 0.82–1.23; adjusted $p = 0.985$).

Subjective workload and ergonomic discomfort

Subjective workload was highest with 2DL, intermediate with 3DL, and lowest with RAS (overall comparison, $p < 0.001$; Figure 4). After adjustment, NASA-TLX workload was 17.3 points higher with 2DL than with RAS (95% CI: 9.1–25.6; adjusted $p < 0.001$), whereas the remaining pairwise comparisons were not significant after correction for multiple comparisons. Platform-related differences were observed for physical demand, effort, and self-rated performance, but not for mental demand, temporal demand, or frustration. Self-rated performance was the only domain in which 3DL differed significantly from 2DL.

Ergonomic discomfort was also lowest with RAS (overall comparison, $p = 0.002$; Figure 5). However, after correction for multiple comparisons, only shoulder discomfort differed significantly across platforms. Compared with RAS, the odds of a higher shoulder discomfort score were greater with 2DL (odds ratio [OR], 4.51; 95% CI: 1.91–10.65; adjusted $p = 0.007$) and 3DL (OR, 6.50; 95% CI: 2.56–16.49; adjusted $p = 0.003$). No significant ergonomic advantage of 3DL over 2DL was observed.

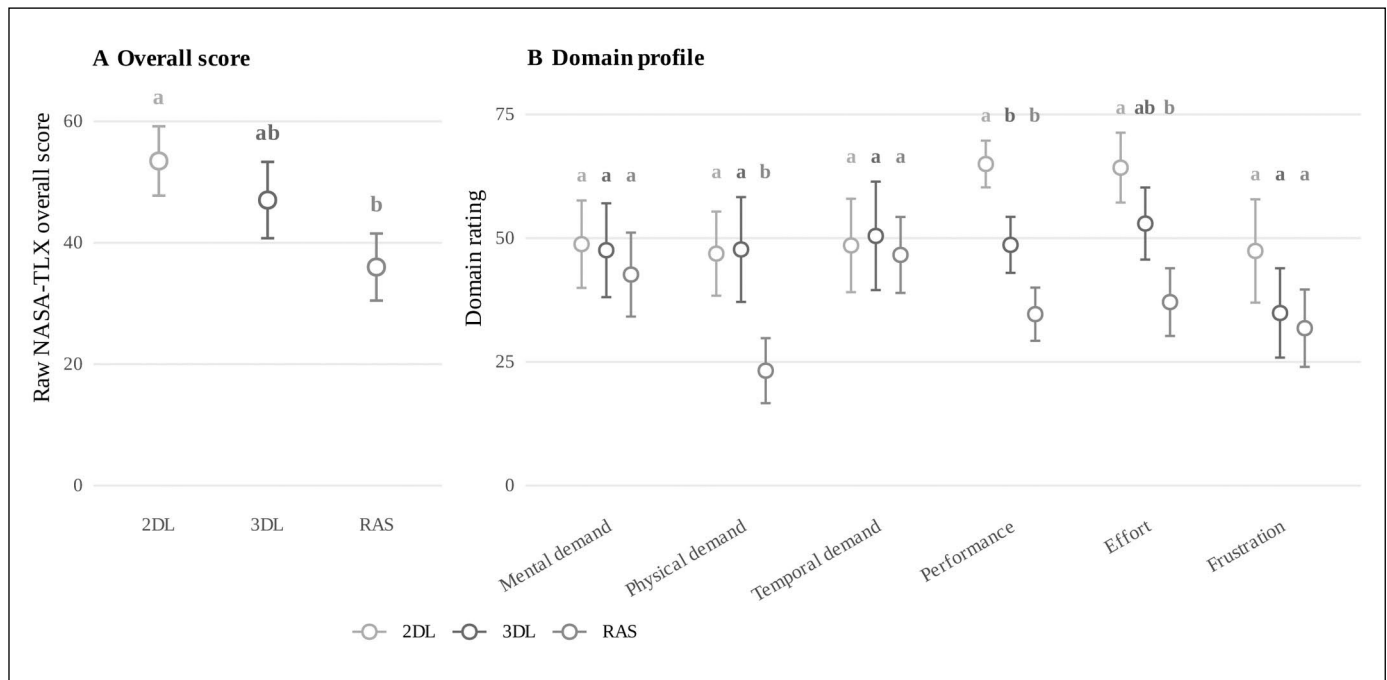


Figure 4. Subjective workload across platforms.

Panel A shows the overall raw NASA-TLX score. Panel B shows the six individual workload domains. Points represent means and error bars represent 95% confidence intervals. Different letters indicate statistically significant pairwise differences after correction for multiple comparisons; groups sharing a letter did not differ significantly.

CI – confidence interval; NASA-TLX – National Aeronautics and Space Administration Task Load Index; 2DL – two-dimensional laparoscopy; 3DL – three-dimensional laparoscopy; RAS – robot-assisted surgery

Associations and sensitivity analyses

Higher workload remained associated with longer Needle Guidance and combined completion times after adjustment for platform, centre, and participant characteristics, while greater discomfort remained associated with longer combined completion time. Age, sex, current gaming exposure, and previous musical instrument experience were not independently associated with the primary outcomes after correction for multiple comparisons. Comparisons

of RAS with both laparoscopic platforms were consistent across all prespecified sensitivity analyses, whereas comparisons between 2DL and 3DL varied by analytical method and remained inconclusive.

DISCUSSION

In this prospective multicentre comparative study of complete novices, RAS provided the most favourable overall initial performance, with shorter task completion times, more frequent attainment

Table 4. Non-completion of the study protocol, by allocated platform

Platform	Allocated, n	Non-completion, n (%)	Risk difference vs RAS, pp (95% CI)	Firth OR vs RAS (95% CI)
RAS	43	0 (0.0)	Reference	Reference
2DL	39	4 (10.3)	10.3 (−0.0 to 23.6)	11.0 (1.1 to 1,480)
3DL	34	8 (23.5)	23.5 (9.7 to 40.0)	27.9 (3.3 to 3,663)

Non-completion was compared across platforms using Fisher's exact test ($P = .001$). Risk differences are Newcombe hybrid-score intervals; odds ratios are Firth penalised-likelihood estimates with profile-likelihood intervals, used because no participant allocated to RAS failed to complete the protocol. The two laparoscopic platforms did not differ significantly (risk difference, 13.3 percentage points; 95% CI: −4.1 to 30.9). Non-completion should be interpreted with caution because participants were free to withdraw at any time under the informed-consent procedure and reasons for non-completion were not uniformly available; therefore, non-completion cannot be interpreted uniformly as task failure attributable to the allocated platform.

2DL – two-dimensional laparoscopy; 3DL – three-dimensional laparoscopy; CI – confidence interval; OR – odds ratio; pp – percentage points; RAS – robot-assisted surgery

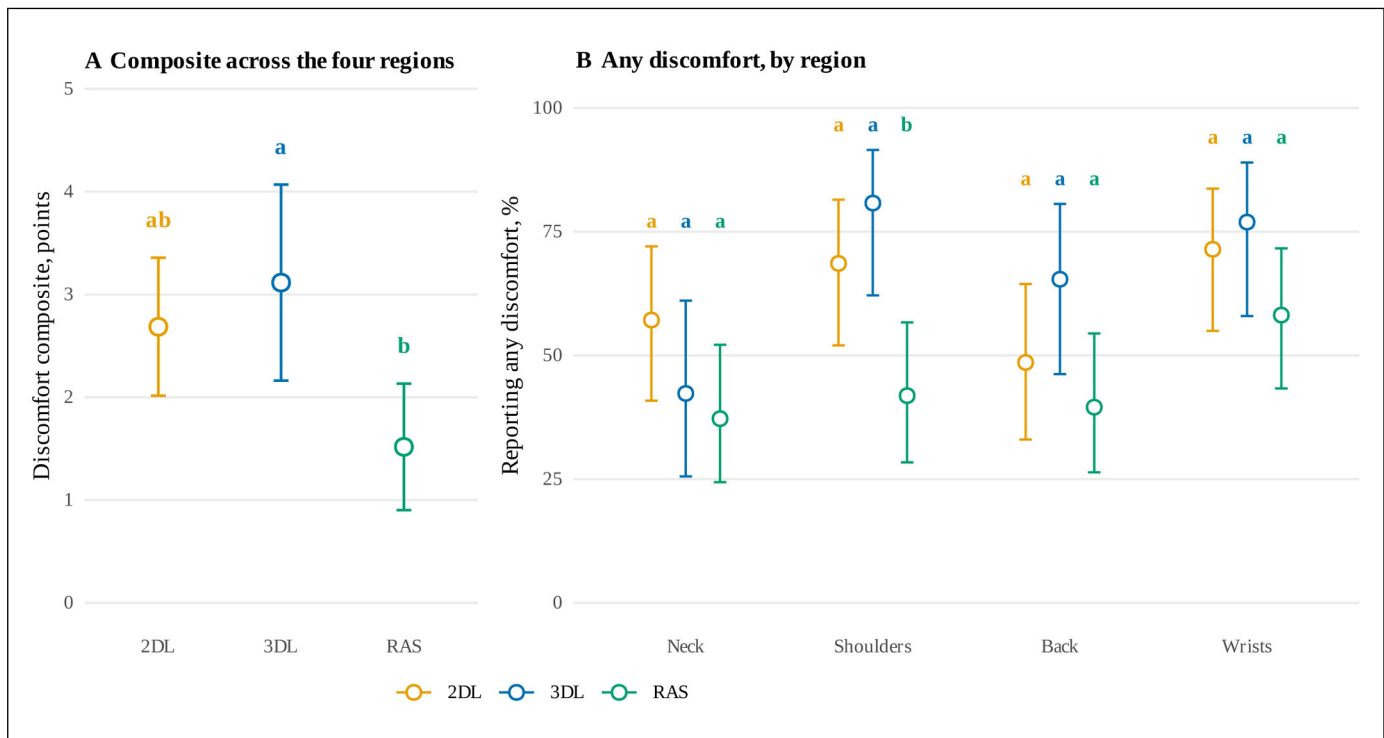


Figure 5. Ergonomic discomfort across platforms.

Panel A shows the composite discomfort score across four anatomical regions. Panel B shows the proportion of participants reporting any discomfort in the neck, shoulders, back, and wrists. Error bars represent 95% confidence intervals. Different letters indicate statistically significant pairwise differences after correction for multiple comparisons; groups sharing a letter did not differ significantly.

CI – confidence interval; 2DL – two-dimensional laparoscopy; 3DL – three-dimensional laparoscopy; RAS – robot-assisted surgery

of the predefined E-BLUS-derived performance criteria, and lower perceived workload and physical discomfort than either laparoscopic platform. Participants using 2DL also made more Peg Transfer errors, whereas error rates were similar with 3DL and RAS. By contrast, differences between 2DL and 3DL were less consistent and varied according to task and analytical method. Notably, the relative increase in completion time from Peg Transfer to the more spatially demanding Needle Guidance task was greater with 2DL than with either 3DL or RAS. These findings are consistent with a lower initial technical and subjective burden with RAS during basic minimally invasive tasks, while stereoscopic visualisation may be one contributor to the larger between-platform differences observed as spatial and precision demands increase.

The favourable initial performance with RAS is consistent with previous studies involving inexperienced operators. Stefanidis et al. [15] reported faster performance, higher objective scores, fewer errors, and lower NASA-TLX workload with robotic compared with conventional laparoscopic suturing in medical students. Park et al. [16] similarly found that novice operators who failed to complete a suturing task using either laparoscopic platform completed it with RAS, whereas platform differences narrowed with increasing experience. Together, these findings suggest that the relative advantage of RAS may be greatest during initial exposure and may change as conventional laparoscopic skills develop.

Technological advancement alone does not necessarily facilitate initial performance. Uysal et al. [25] found slower E-BLUS performance and more Needle Guidance failures with motorised articulating instruments than with conventional instruments. These findings underscore that novice performance may depend not only on the technical capabilities of a platform but also on the cognitive and motor demands of its interface. By contrast, RAS integrates multiple visual and instrument-control features that may reduce several psychomotor constraints encountered by novices.

Previous comparisons of 2DL and 3DL generally favoured 3DL but showed task- and experience-dependent effects. Gómez-Gómez et al. [14] reported shorter completion times, fewer errors, and lower workload with 3DL in medical students, whereas Cicione et al. [26] found that the magnitude of the 3DL advantage differed by task and operator experience.

In the present study, 3DL did not demonstrate a consistent overall time advantage over 2DL, but it was associated with fewer Peg Transfer errors and a smaller relative increase in completion time

during Needle Guidance. Differences between the studies may reflect task selection, study design, preliminary training, and whether cumulative or task-specific outcomes were examined. The present findings are consistent with task-dependent differences between 2DL and 3DL, with greater use of depth information as one possible contributor [27]. However, because Peg Transfer and Needle Guidance also differed in duration, motor requirements, precision demands, and fixed order, the observed interaction likely reflects several dimensions of task complexity and does not isolate stereoscopic visualisation as the causal mechanism.

Objective performance differences were accompanied by differences in perceived workload and physical discomfort. Overall workload was highest with 2DL, intermediate with 3DL, and lowest with RAS, although after adjustment the clearest pairwise difference was between 2DL and RAS. Differences were mainly observed in physical demand, effort, and self-rated performance. These findings are consistent with previous studies reporting platform-related differences in ergonomics and workload [7, 8, 14, 15]. Physical discomfort was also lowest with RAS, with shoulder discomfort showing the clearest regional differences. Higher workload remained associated with longer Needle Guidance and combined completion times, while greater discomfort was associated with longer combined task time, supporting an association between subjective burden and initial technical performance.

E-BLUS is a structured training and assessment programme that has been used across international educational settings [22], and its published proficiency standards are demanding even for advanced trainees. Brinkman et al. [12] reported that only 4.2% of final-year urology residents met the combined validated time and quality criteria across all 5 examination tasks. Against this background, attainment of the selected thresholds by some complete novices in a single session is noteworthy, particularly with RAS. Criterion attainment in the present study is therefore best interpreted as performance relative to E-BLUS-derived benchmarks rather than formal E-BLUS certification.

Neither current gaming exposure nor previous musical-instrument experience was independently associated with the primary outcomes, consistent with the mixed evidence reported in previous studies [17, 18].

The present findings indicate meaningful differences in initial task performance and perceived difficulty between platforms and may inform future research on training pathways that combine platform-specific competence with transferable

minimally invasive skills. This single-session comparison measures the initial intuitiveness and ease of use of each platform for a complete novice rather than the rate of learning, retention, or transfer; a platform that yields the best first attempt is not necessarily the one on which training should begin, so any implication for training pathways is hypothesis-generating and requires longitudinal evaluation. Educational programmes should also balance the potential benefits of emerging technologies with equipment availability, training capacity, costs, and access to advanced platforms.

We hypothesise, as a direction for future educational research, that access to a stereoscopic assistant display during bedside assistance might support novices' early grasp of spatial relationships; this was not evaluated in the present study.

The principal strengths of this study include its prospective multicentre design, direct comparison of three platforms, inclusion of a relatively large and homogeneous cohort of complete novices, use of a common standardised protocol, and combined assessment of completion time, errors, predefined performance criteria, workload, and physical discomfort. The inclusion of two tasks also enabled evaluation of whether platform effects changed as task demands increased.

Several considerations should be taken into account when interpreting the findings. Platform allocation was only partially randomised and reflected equipment availability, which may have introduced centre- or session-related variation. Because the third centre operated no three-dimensional station, platform is entangled with centre by design, and adjustment for centre cannot fully separate a platform effect from a centre or session effect. Baseline characteristics of the 12 participants who withdrew were not captured in the analytical dataset, so baseline imbalance at allocation could not be tabulated over the full enrolled cohort. The parallel-group design and single recorded attempt were well suited to evaluating initial performance but did not examine longitudinal learning or transfer between platforms. Because the participants were predominantly students without surgical experience rather

than urology residents, these results describe the initial psychomotor performance of complete novices and should not be extrapolated to trainees who already possess transferable operative, endoscopic, visuospatial, or tissue-handling skills; the inclusion of complete novices is nonetheless a strength for isolating initial performance. The fixed task order may also have contributed to the differences observed between Peg Transfer and Needle Guidance. Differential non-completion, particularly during Needle Guidance, may have biased completion-time comparisons among protocol completers. The brief simulation setting may not reflect all aspects of prolonged operative performance. Finally, the E-BLUS-derived criteria served as common comparative benchmarks and were not intended to represent platform-specific certification.

CONCLUSIONS

The present study showed that, among complete novices, RAS was associated with the most favourable initial performance, including shorter task completion times, more frequent attainment of predefined E-BLUS-derived performance criteria, and lower overall workload and physical discomfort than either laparoscopic platform. The advantage of 3DL over 2DL was less consistent, although 3DL was associated with fewer Peg Transfer errors and a smaller relative increase in completion time during the more demanding Needle Guidance task. Further longitudinal research may clarify how these early differences relate to learning, skill retention, transfer between platforms, and the organisation of minimally invasive surgical training.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ETHICS APPROVAL STATEMENT

The study protocol was approved by the relevant institutional Bioethics Committee (approval no. 24/2025, dated 22 December, 2025).

SUPPLEMENTARY MATERIAL

Supplementary Table S1. Composition of the full allocated cohort by centre and allocated platform (n = 116)

Centre	2DL	3DL	RAS	Total
Centre 1	12	27	16	55
Centre 2	19	7	19	45
Centre 3	8	0	8	16
Total	39	34	43	116

Centre 3 had no three-dimensional laparoscopic platform available, and platform availability was therefore partly associated with centre by design. The table describes all 116 allocated participants.

2DL – two-dimensional laparoscopy; 3DL – three-dimensional laparoscopy; RAS – robot-assisted surgery

References

- Lai TJ, Roxburgh C, Boyd KA, Bouttell J. Clinical effectiveness of robotic versus laparoscopic and open surgery: an overview of systematic reviews. *BMJ Open*. 2024; 14: e076750.
- Amiri R, Zwart MJW, Jones LR, et al. Surgeon Preference and Clinical Outcome of 3D Vision Compared to 2D Vision in Laparoscopic Surgery: Systematic Review and Meta-Analysis of Randomized Trials. *Ann Surg Open*. 2024; 5: e415.
- Sørensen SMD, Savran MM, Konge L, Bjerrum F. Three-dimensional versus two-dimensional vision in laparoscopy: a systematic review. *Surg Endosc*. 2016; 30: 11-23.
- Zwimpfer TA, Stiegeler N, Müller PC, et al. Comparison of 2D 4K vs. 3D HD laparoscopic imaging systems in bariatric surgery: study protocol for a randomized controlled prospective trial. *Trials*. 2024; 25: 140.
- Rivero-Moreno Y, Echevarria S, Vidal-Valderrama C, et al. Robotic Surgery: A Comprehensive Review of the Literature and Current Trends. *Cureus*. 2023; 15: e42370.
- Longmore SK, Naik G, Gargiulo GD. Laparoscopic Robotic Surgery: Current Perspective and Future Directions. *Robotics*. 2020; 9: 42.
- Lee GI, Lee MR, Clanton T, Sutton E, Park AE, Marohn MR. Comparative assessment of physical and cognitive ergonomics associated with robotic and traditional laparoscopic surgeries. *Surg Endosc*. 2014; 28: 456-465.
- Berguer R, Smith W. An Ergonomic Comparison of Robotic and Laparoscopic Technique: The Influence of Surgeon Experience and Task Complexity. *J Surg Res*. 2006; 134: 87-92.
- Checucci E, Puliatto S, Pecoraro A, et al. ESUR-ESU-YAU_URTECH Survey on Urology Residents Surgical Training: Are We Ready for Simulation and a Standardized Program? *Eur Urol Open Sci*. 2024; 61: 18-28.
- Mantica G, Carrion GD, Pang KH. The definition of ideal training of a urology resident from two different perspectives: trainees vs professors. Is there agreement in their idea of good training? *Cent European J Urol*. 2023; 76: 162-166.
- Lavi A, Tzemah S, Hussein A, et al. The Urology Residency Program in Israel – Results of a Residents Survey and Insights for the Future. *Rambam Maimonides Med J*. 2017; 8: e0039.
- Brinkman WM, Tijam IM, Schout BMA, et al. Results of the European Basic Laparoscopic Urological Skills Examination. *Eur Urol*. 2014; 65: 490-496.
- Anacleto S, Mota P, Fernandes V, et al. Can narration and guidance in video-enhanced learning improve performance on E-BLUS exercises? *Cent European J Urol*. 2021; 74: 131-138.
- Gómez-Gómez E, Carrasco-Valiente J, Valero-Rosa J, et al. Impact of 3D vision on mental workload and laparoscopic performance in inexperienced subjects. *Actas Urol Esp*. 2015; 39: 229-235.
- Stefanidis D, Wang F, Korndorffer JR, Dunne JB, Scott DJ. Robotic assistance improves intracorporeal suturing performance and safety in the operating room while decreasing operator workload. *Surg Endosc*. 2010; 24: 377-382.
- Park YS, Oo AM, Son SY, et al. Is a robotic system really better than the three-dimensional laparoscopic system in terms of suturing performance?: comparison among operators with different levels of experience. *Surg Endosc*. 2016; 30: 1485-1490.
- Gupta A, Lawendy B, Goldenberg MG, Grober E, Lee JY, Perlis N. Can video games enhance surgical skills acquisition for medical students? A systematic review. *Surgery*. 2021; 169: 821-829.
- Arbel EJ, Kozłowski K, Reese AD, Burns O, Lamboy JA, Rosser JC. The influence of musical experience on laparoscopic surgical skill performance: an original investigation and literature review. *Surg Endosc*. 2024; 38: 6736-6742.
- Von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *Ann Intern Med*. 2007; 147: 573-577.
- Cheng A, Kessler D, Mackinnon R, Chang TP, Nadkarni VM, Hunt EA, et al. Reporting Guidelines for Health Care Simulation Research: Extensions to the CONSORT and STROBE Statements. *Simul Healthc J Soc Simul Healthc*. 2016; 11: 238-248.
- European Association of Urology. Uroweb [Internet]. [cited 2026 Jul 22]. Laparoscopic urological skills- Step 1 (LUSs1) – European School of Urology (ESU). Available at: <https://uroweb.org/training/laparoscopic-training/laparoscopic-urological-skills-step-1>.
- Somani BK, Van Cleynenbreugel B, Gözen AS, et al. Outcomes of European

- Basic Laparoscopic Urological Skills (EBLUS) Examinations: Results from European School of Urology (ESU) and EAU Section of Uro-Technology (ESUT) over 6 Years (2013-2018). *Eur Urol Focus*. 2020; 6: 1190-1194.
23. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: *Advances in Psychology* [Internet]. Elsevier; 1988 [cited 2026 Jul 28]; 139-183. Available at: <https://linkinghub.elsevier.com/retrieve/pii/S0166411508623869>. doi:10.1016/S0166-4115[08]62386-9.
24. Said S, Gozdzik M, Roche TR, et al. Validation of the Raw National Aeronautics and Space Administration Task Load Index (NASA-TLX) Questionnaire to Assess Perceived Workload in Patient Monitoring Tasks: Pooled Analysis Study Using Mixed Models. *J Med Internet Res*. 2020; 22: e19472.
25. Uysal D, Gasch C, Behnisch R, et al. Evaluation of new motorized articulating laparoscopic instruments by laparoscopic novices using a standardized laparoscopic skills curriculum. *Surg Endosc*. 2021; 35: 979-988.
26. Cicione A, Autorino R, Breda A, et al. Three-dimensional vs Standard Laparoscopy: Comparative Assessment Using a Validated Program for Laparoscopic Urologic Skills. *Urology*. 2013; 82: 1444-1450.
27. Yamazaki Y, Kanaji S, Harada H, et al. Three-dimensional laparoscopic vision improves forceps motion more in the depth direction than in the horizontal direction: An analysis of data from prospective randomized controlled trials. *Asian J Endosc Surg*. 2020; 13: 265-271. ■