



# Laparoscopic skills training: the effects of viewing mode (2D vs. 3D) on skill acquisition and transfer

Kirsty L. Beattie<sup>1</sup> · Andrew Hill<sup>1,2,3</sup> · Mark S. Horswill<sup>1</sup> · Philip M. Grove<sup>1</sup> · Andrew R. L. Stevenson<sup>4,5</sup>

Received: 14 May 2020 / Accepted: 17 August 2020 / Published online: 2 September 2020  
© The Author(s) 2020

## Abstract

**Background** Three-dimensional (3D) visual displays have been suggested to aid laparoscopic skills training by providing the depth cues not present in traditional two-dimensional (2D) displays. However, few studies have robustly investigated the impact of viewing mode (2D vs. 3D) on learning outcomes.

**Purpose** To examine how viewing mode (2D vs. 3D) impacts the acquisition and transferability of basic laparoscopic skills by comparing performance between transfer and control groups on a complete proficiency-based training program.

**Method** A counterbalanced between-subjects design was employed. Each participant was randomly allocated to one of four groups, comprising two transfer groups (trained in one viewing mode and tested in the alternate mode: the 2D → 3D and 3D → 2D groups) and two control groups (trained and tested in one viewing mode: the 2D → 2D and 3D → 3D groups). Participants completed proficiency-based training in six laparoscopic training tasks. Testing included two further repetitions of all tasks under test conditions. Objective performance measures included the total number of repetitions to reach proficiency, and total performance scores (i.e. time + error penalties across all repetitions) in training and testing.

**Results** The groups trained in 3D demonstrated superior training performance (i.e. less time + errors) and took fewer repetitions to reach proficiency than the groups trained in 2D. The groups tested in 3D also demonstrated superior test performance compared to those tested in 2D. However, training mode did not yield significant test differences between the groups tested in 2D (i.e. 2D → 2D vs. 3D → 2D), or between the groups tested in 3D (i.e. 3D → 3D vs. 2D → 3D).

**Conclusion** Novices demonstrate superior performance in laparoscopic skills training using a 3D viewing mode compared to 2D. However, this does not necessarily translate to superior performance in subsequent testing or enhanced learning overall. Rather, test performance appears to be dictated by the viewing mode used during testing, not that of prior training.

**Keywords** 3D laparoscopy · 2D laparoscopy · Surgical skills · Training · Novices · Learning

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00464-020-07923-8>) contains supplementary material, which is available to authorised users.

✉ Kirsty L. Beattie  
kirsty.beattie@uq.net.au

- <sup>1</sup> School of Psychology, The University of Queensland, St Lucia, Brisbane 4072, Australia
- <sup>2</sup> Clinical Skills Development Service, Metro North Hospital and Health Service, Brisbane, Australia
- <sup>3</sup> Minerals Industry Safety and Health Centre, Sustainable Minerals Institute, The University of Queensland, Brisbane, Australia
- <sup>4</sup> School of Medicine, The University of Queensland, Brisbane, Australia
- <sup>5</sup> Department of Colon and Rectal Surgery, Royal Brisbane and Women's Hospital, Brisbane, Australia

## Background

Laparoscopy is currently considered the gold standard approach for a range of surgical treatments [1]. While these minimally invasive procedures offer significant advantages in patient recovery, they also require complex surgical skills that are difficult for novices to learn and highly vulnerable to error. Moreover, research indicates that 80% of major laparoscopic complications occur early on in a surgeon's career (often within the first 100 cases), and can lead to disastrous patient outcomes including permanent injury or fatality [2–4]. Consequently, it is important for contemporary training methods to accelerate the learning curve and provide trainees with a platform to further optimise laparoscopic skills within the training period [5]. It has been suggested that the recent introduction of three-dimensional (3D)

visual displays has the potential to accelerate the learning curve in training by overcoming the perceptual constraints associated with traditional two-dimensional (2D) displays (i.e. by increasing depth perception) [6]. However, not all contemporary operating theatres have access to 3D displays to support surgical tasks, and different equipment is often employed across sites (e.g. displays of differing brand, size, and picture quality) [7]. As a result, laparoscopic surgeons must be able to demonstrate surgical proficiency with any and all display types.

Evidence suggests that superior performance during training in simulated laparoscopic tasks (i.e. faster completion times and fewer errors) can be achieved when novices use a 3D viewing mode rather than a 2D display or viewing mode [6–9]. However, several studies identified that the superior skill acquisition associated with training in 3D was not reflected in subsequent performance during testing in 2D [7, 10–12]. Further, those trained in 2D performed at a superior level when subsequently tested in 3D [7, 10–12]. Such results have been interpreted as an indication of ineffective skill transfer from 3D to 2D, and effective or increased skill transfer from 2D to 3D [7, 10–12]. If this interpretation is correct, it may be because 2D training forces trainees to attend to and rely upon secondary depth cues (e.g. shadows, object occlusion, image size, and changes with motion) [13] to overcome the perceptual constraints of 2D displays (e.g. an absence of the convergence and stereopsis that normal vision relies on) [13]. Thus, despite the extra difficulty associated with 2D leading to increased training time, subsequent performance in 3D is enhanced as trainees are able to use both primary and secondary depth cues. Similarly, it could also be argued that those who train in 3D may not learn to exploit secondary depth cues due to the availability of primary depth cues, and subsequently experience greater difficulty performing tasks in 2D when primary depth cues are no longer available.

It is also important to note that some studies have found opposing results regarding skill transfer (i.e. transferring from 3D to 2D appeared to improve performance, while transferring from 2D to 3D appeared to reduce performance [14–17]). However, unlike the research discussed above, these studies employed a defined number of repetitions or a fixed length of time for participant training, rather than the attainment of a criterion or “proficiency” level [14–17]. As a result, the 2D- and 3D-trained groups did not reach an equivalent level of performance during training, and thus any comparisons of subsequent performance in alternate viewing modes should be interpreted with caution [14–17]. Unfortunately, drawing clear conclusions from prior research into 2D/3D skill transfer is ultimately problematic due to the omission of appropriate control groups. Even in studies where participants were trained to a proficiency criterion, it is impossible to determine the extent to which observed

performance differences on transfer tasks were due to the viewing mode itself or the additional practice generally received by those trained in 2D. In addition, previous research has typically failed to account for potentially confounding variables, such as participants’ stereoacuity (i.e. the smallest detectable difference in depth that they can perceive using binocular vision) [6, 18, 19], their psychomotor abilities (e.g. manual dexterity), or potential “crosstalk” (i.e. double vision due to the use of suboptimal 3D viewing positions, which results in both eyes viewing a combination of the image intended for one eye and a portion of the image intended for the other eye) [19].

The aim of the present study was to overcome the limitations of prior research and draw robust conclusions regarding the impact of viewing mode on laparoscopic training and performance. Specifically, the current study examines (1) whether using a 3D (versus 2D) viewing mode enhances the acquisition of laparoscopic skills, and (2) to what extent the skills acquired using a 3D viewing mode transfer to 2D, and vice versa. Furthermore, as research has yet to determine whether differences in transfer performance can be attributed to the change in viewing mode itself or to the different lengths of exposure during prior proficiency-based practice, this study also examines (3) whether differences in test performance exist between those trained and tested in alternate viewing modes and those who experience only one viewing mode in training and testing. The answers to these questions carry significant implications for the hospitals and clinics currently training laparoscopic surgeons, and for the trainee surgeons themselves. For instance, if training in 3D is found to enhance skill acquisition, then training centres may wish to invest in the latest 3D technology to accelerate the learning curve for trainees. However, if the skills acquired under 3D conditions do not adequately transfer to subsequent performance in 2D, then this could be disadvantageous for trainees who learn in 3D but subsequently work at sites with different (i.e. 2D-only) equipment.

## Materials and methods

### Participants

Sixty novices (male,  $n=32$ ; female,  $n=28$ ) with a mean age of 24.78 ( $SD=3.24$ , range = 19–34) voluntarily participated in the study. Participants were current medical students (between first and fourth year of medical school), recruited via an advertisement placed on the University of Queensland’s surgical interest group (Incision UQ) noticeboard. All participants reported normal or corrected-to-normal vision, normal stereoacuity, and no prior laparoscopic experience (including no formalised laparoscopic skills training with a simulator, and no hands-on laparoscopic experience in an

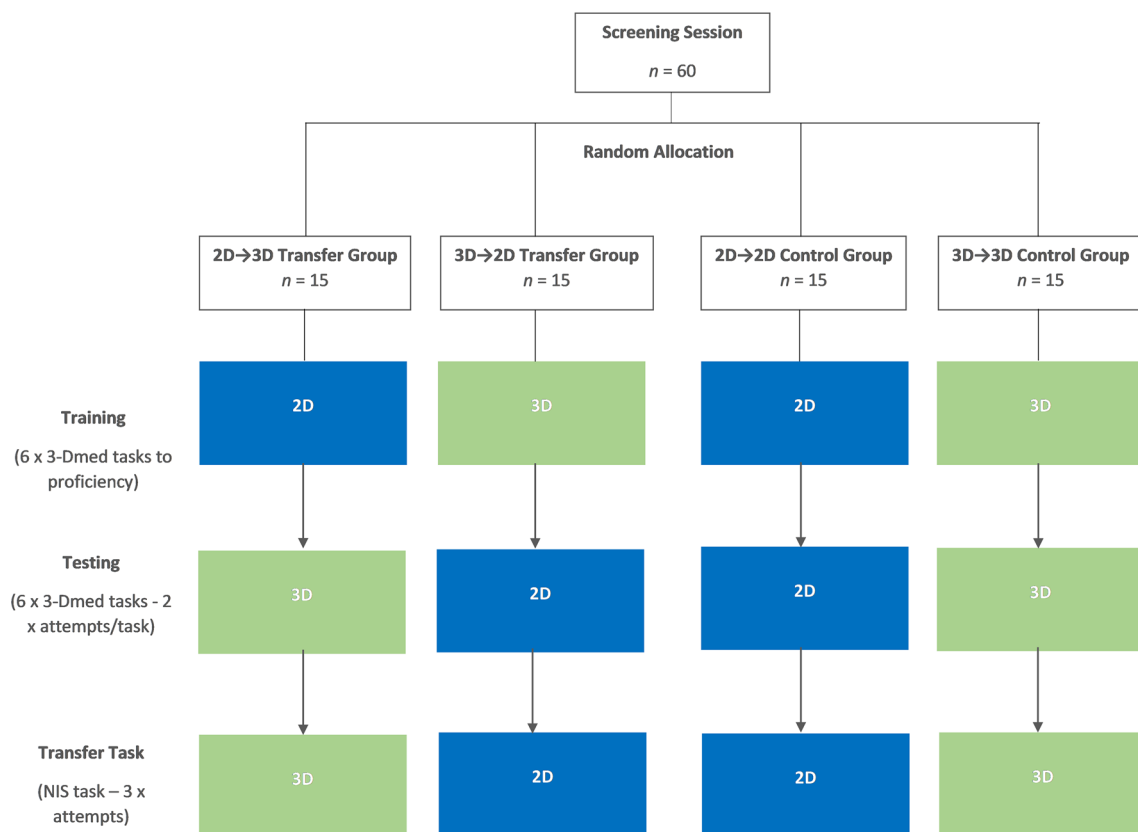
operational context, e.g. as a surgical assistant). Fifty-six participants were right-handed and four were left-handed. Participants received no compensation for their involvement. The study was approved by the relevant institutional review boards (i.e. the Human Research Ethics Committees of the Royal Brisbane and Women's Hospital (RBWH) and The University of Queensland).

## Design

A counterbalanced between-subjects design was employed. Each participant was randomly allocated to one of four experimental groups (in blocks of four participants). This comprised two transfer groups (*trained* in one viewing mode  $\rightarrow$  *tested* in the alternate mode: the  $2D \rightarrow 3D$  and  $3D \rightarrow 2D$  groups) and two control groups (*trained*  $\rightarrow$  *tested* in one viewing mode: the  $2D \rightarrow 2D$  and  $3D \rightarrow 3D$  groups). Given that some studies have found significant performance differences between sexes during initial laparoscopic skills training [15, 20, 21], a conservative strategy of randomising separately for males and females was used to avoid significantly uneven sex distributions across the groups.

## Procedure

Each participant attended a total of four one-on-one sessions with an experimenter. All sessions were conducted in a research lab in a clinical simulation centre at the RBWH, Brisbane, Australia. This included a screening session where background information was obtained and laparoscopically relevant skills and abilities were assessed, followed by two separate training sessions and one testing session (Fig. 1). In the training sessions, the participant completed a series of six laparoscopic training tasks using either a 2D or 3D viewing mode, depending on their experimental group. The six tasks were practised in a set order (i.e. the first three tasks in session one and the last three in session two). Each task was performed repeatedly until the participant was able to complete the task to a pre-defined criterion level of proficiency (on two non-consecutive attempts) before moving on to the next task. During the testing session, the participant performed each of the six tasks twice under test conditions (in the same order as in training), followed by three attempts at a novel task. Participants were also asked to rate their experiences (i.e. perceived workload and physical/visual comfort) following each training and testing session.



**Fig. 1** Overview of the study design and procedure showing the viewing mode (2D or 3D) used at each stage of the experiment

## Screening

Participants completed a demographic questionnaire to document their age, gender, hand dominance, use of corrective glasses, occupation, level of medical study, interest in surgery, and their experience in surgery, laparoscopy, suturing, 3D displays, video games, and snooker/billiards. To account for any attrition that may have occurred during the study, a validated measure of psychological grit [22] was also administered. Grit is a characteristic found to positively correlate with residents' longevity, performance, and well-being during medical training (with higher scores reflecting more grit) [23–25]. Overall, the study did not encounter any such problems with attrition, and grit was not found to be associated with any measure of laparoscopic performance.

Participants' visual-spatial ability, manual dexterity, stereoacuity, and visual acuity, abilities that have all been associated with novices' laparoscopic performance [26, 27], were also measured to detect any outliers or significant group-level differences that could potentially confound the results. Visual-spatial ability was assessed using an online adaptation of the Mental Rotations Test (MRT-A) [28]. The MRT-A evaluates participants' ability to visualise and mentally manipulate a 2D image of a 3D object in space, and is a well-established test of visual-spatial ability that has been utilised in prior laparoscopic research [29–31]. Manual dexterity was measured using the Purdue Pegboard (PP) test (Lafayette Instrument Co). The PP includes four subtests that involve the placement of small pins into holes on a board, and the assembly of pins and washers to assess fine finger dexterity. The reliability and validity of this test has been well established [32, 33]. For both the MRT-A and PP test, higher scores reflect a higher level of ability.

The Randot ® Stereo Test (Stereo Optical, Chicago, IL) was used to assess stereoacuity and ensure that all participant

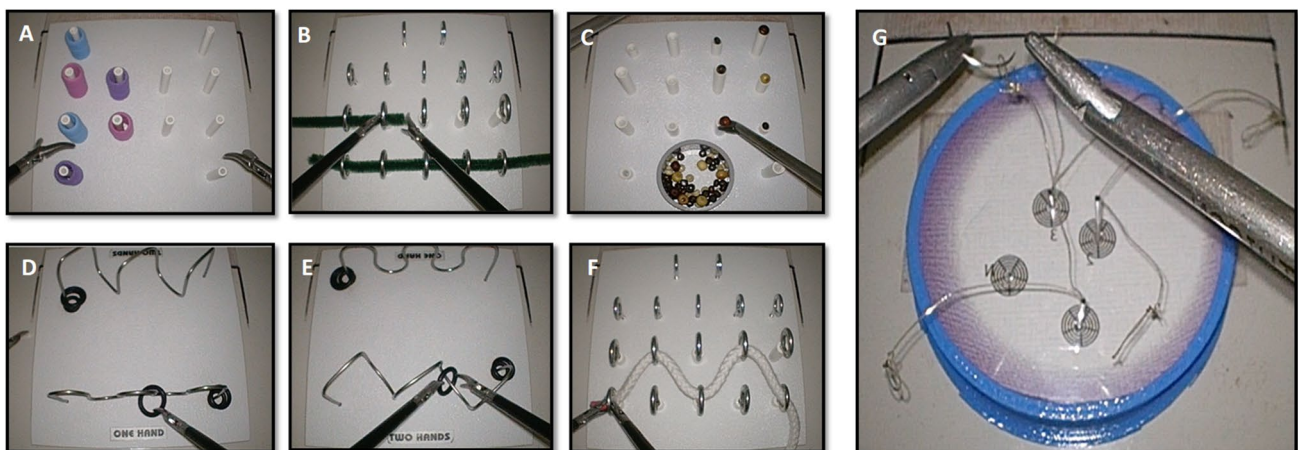
stereopsis was sufficient (between 20 and 200 arc/sec) [34] to effectively use the 3D display. This test presents ten sets of three circles, with each set containing a target circle that appears closer than the others (crossed disparity) when viewed through cross-polarised glasses [18]. The viewer's task is to identify the target circle. In each successive set, the disparity decreases, and lower scores reflect greater stereoacuity. This test has been found to be a reliable, valid, and sensitive measure of stereoacuity [35, 36].

Participants' right and left visual acuity were each assessed using the logMAR visual acuity chart (National Vision Research Institute, Melbourne, Australia) at a distance of three metres. The logMAR involves the presentation of a series of five-letter rows, incrementally decreasing in size and spaced in a standardised manner to determine 2D spatial discrimination [37]. A visual acuity score exceeding 9.5 m (or 3/9.5) was considered to be non-normal (as the World Health Organization defines low visual acuity as a logMAR value exceeding 0.5 [38], or a score greater than 9.5 m with a chart placed at a distance of three metres).

## Training

### Laparoscopic tasks

Participants performed all six tasks from the 3-Dmed program (3-Dmed®, Franklin, OH, US) with laparoscopic graspers (Fig. 2). The 3-Dmed tasks (i.e. post and sleeve, loops and wire, pea on a peg, wire chaser (one hand), wire chaser (two hands), and zig-zag loop) are intended to train and evaluate skills such as hand–eye coordination, manual dexterity, laparoscopic depth perception, and interactions between the dominant and non-dominant hand through touching, grasping, transferring, placing, navigating, and manoeuvring with laparoscopic graspers [39]. To assess the



**Fig. 2** Laparoscopic images of the 3-Dmed tasks used in training and testing **A** Post and Sleeve. **B** Loops and Wire. **C** Pea on a Peg. **D** Wire Chaser (one hand). **E** Wire Chaser (two hands). **F** Zig-zag loop, and the novel task used in testing (**G** Navigating in Space)

validity of metrics derived from the use of these 3-Dmed tasks, Schreuder et al. [39] compared the performance scores (i.e. time + error penalties) of groups of laparoscopic experts, intermediates, and novices on the six exercises. They found that the tasks effectively discriminated and reflected the different levels of expertise as the expert group performed significantly better than the novice group on all exercises [39]. Therefore, we employed the average performance scores from this expert group as the criterion level of proficiency in the current study, and all training instructions and error parameters defined by Schreuder et al. [39] were adopted here (see Appendix 1 for task descriptions and criterion proficiency scores).

## Testing

### Laparoscopic tasks

In addition to repeating the six 3-Dmed tasks twice, participants completed a novel task developed and validated by Sakata et al. [18, 40], known as “Navigating in Space” (NIS). The objective of this task is to measure fine dexterity and instrument control. It was used here to assess whether the training or testing viewing mode impacted novices’ ability to transfer previously acquired skills to a novel task. The task required participants to hold a needle-holder in each hand and pass a curved needle through six 2 mm loops at the tip of a monofilament suture in a pre-defined sequence [18, 40] (Fig. 2). This task was completed a total of three times at the end of the testing session. This included a practice attempt to allow participants to become familiar with the task, followed by two attempts under test conditions. A 10-min time limit was applied to each attempt at this task.

## Outcome measures

### Repetitions

The total number of repetitions taken to reach the pre-defined level of proficiency across all six tasks was calculated as a measure of training efficiency.

### Performance scores

The total time taken and the total time penalties incurred for errors across all training repetitions were combined to provide a total performance score for training (in seconds). The total time and the total time penalties incurred across all 12 test attempts (two attempts per 3-Dmed task) were also combined to provide a total performance score for testing (in seconds). Lower scores indicated more efficient and more accurate performances.

## Novel task performance

Performance on the novel (NIS) task was measured by the total time taken to pass the suture through all six loops, up to a maximum of 600 s (10 min). The times from the two test attempts were then combined to provide a total NIS score (in seconds).

## Simulator comfort

The Simulator Sickness Questionnaire (SSQ) [41] was administered after each training and testing session, to investigate the impact of 2D and 3D viewing modes on physical/visual comfort. The SSQ was chosen as it is the only validated measure that addresses the symptoms specific to simulation use [41].

## Workload

With stress and mental workload found to negatively impact laparoscopic task performance [42], the NASA Task Load Index (NASA-TLX) [43] was administered at the end of each training and testing session to measure the self-perceived workload associated with the tasks. The NASA-TLX comprises six rating scales measuring mental demand, physical demand, temporal demand, performance, effort, and frustration. This tool has been extensively validated across a number of industries (including the medical sector) and is now the gold standard for workload assessment [40, 44].

## Equipment and setup

A PC was used to present instructional videos on how to complete the 3-Dmed simulation-based exercises (derived from Schreuder et al. 2011) [39]. All tasks were performed in a box trainer using two laparoscopic graspers (3-Dmed tasks) or needle-holders (NIS task). An Olympus Endoeye Flex 3D Imaging System (Olympus Corporation, Tokyo, Japan) was used to provide the light source, laparoscope, video processors, recording device, and monitor. This system allowed either a 2D or 3D image to be displayed, ensuring consistency in the size and clarity of images shown across conditions. A one metre distance between the display and the participant was continuously maintained to standardise picture clarity and focus across participants. This distance was chosen as it has been found to optimise performance relative to longer distances (e.g. three metres) [18]. Additionally, the height of the monitor was adjusted to ensure that each participant’s natural gaze was perpendicular to the centre of the screen to minimise crosstalk (as per the findings and recommendations of Sakata et al. 2016) [45]. A stopwatch was used to record participants’ time on each task.

## Statistical analysis

Statistical analyses were conducted using SPSS® version 25 (IBM Corp, 2017, Armonk, NY: USA), with  $\alpha$  set at 0.05. A chi-square test was used to evaluate categorical data (i.e. gender). A series of one-way ANOVAs was used to identify differences between the four groups on continuous demographic/screening variables, simulator comfort, workload, repetitions, and performance scores during training and testing. One-way ANOVAs were employed to ensure that all appropriate comparisons could be made between training and test performance (e.g. conducting baseline checks for training differences that could have a confounding effect on test results). All significant main effects were followed up with a series of post hoc *t* tests using the Bonferroni–Holm sequential adjustment method [46]. However, the main effect of a demographic variable that violated the ANOVA assumption of equal group variances (i.e. right visual acuity) was followed up using the Games–Howell adjustment method [47]. Effect sizes ( $\eta^2$ ) were calculated to further interpret the main effects ( $\eta^2 \geq 0.01$ , 0.06, and 0.14 reflect small, medium, and large effects, respectively) [48].

Following normality checks and visual examination of the data, one participant was defined as an “extreme” outlier based on Tukey’s (1977) Boxplot rule [49], with scores falling more than three times the length of the interquartile range from the group mean (or equivalent to 4.67 standard deviations) [50]. Given the moderate sample size, and a lack of significant skew in the group’s data to account for this deviation, further investigations were conducted. Subsequent analyses revealed that the participant’s manual dexterity score was more than three standard deviations below the normative mean of their respective age/sex group [51]. Consequently, as this was the only case of non-normal manual dexterity within the sample, the outlier was removed from the substantive analyses to avoid unexplained physical limitations from influencing the results.

## Results

### Participant characteristics

Overall, there were no significant differences between the four groups across any of the demographic variables or individual difference measures (Table 1).

### Training

#### Repetitions

Analyses revealed a significant main effect of group on the total number of repetitions taken to reach proficiency in training (Table 2).

**Baseline checks** As expected, post hoc tests indicated that there was no significant difference in the number of repetitions between the two groups trained in 3D (the 3D → 3D and 3D → 2D groups), or between the two groups trained in 2D (the 2D → 2D and 2D → 3D groups) (Table 2).

**2D vs. 3D training** Further post hoc tests revealed that the 3D-trained groups (the 3D → 3D and 3D → 2D groups) took significantly fewer repetitions to reach proficiency than the 2D-trained groups (the 2D → 2D and 2D → 3D groups) (Table 2).

#### Total performance score

A significant main effect of group on total performance scores (time to complete tasks + time penalties for errors) in training to proficiency was also revealed (Fig. 3, Table 2).

**Baseline checks** As expected, post hoc tests indicated that there was no significant difference in total performance scores between the two groups trained in 3D (the 3D → 3D and 3D → 2D groups), or between the two groups trained in 2D (the 2D → 2D and 2D → 3D groups) (Table 2).

**2D vs. 3D training** Further post hoc tests revealed that the 3D-trained groups (the 3D → 3D and 3D → 2D groups) performed significantly better during training (i.e. lower total performance scores) than the 2D-trained groups (the 2D → 2D and 2D → 3D groups) (Table 2).

#### Simulator comfort and workload

There was no significant main effect of group on reported simulator sickness or experienced workload following the two training sessions, indicating that these variables could not account for the group differences in performance (Table 2).

## Test

#### Total performance score

There was a significant main effect of group on total performance scores (time to complete tasks + time penalties for errors) at test (Fig. 4, Table 3).

**Table 1** Participant characteristics

Variable	2D→3D Transfer ( <i>n</i> =15) M (SD)	3D→2D Transfer ( <i>n</i> =15) M (SD)	2D→2D Control ( <i>n</i> =15) M (SD)	3D→3D Control ( <i>n</i> =15) M (SD)	<i>F</i>	<i>p</i>
Gender ( <i>n</i> = male:female) <sup>a</sup>	9:6	8:7	7:8	8:7	0.00	1.00
Age (years)	24.73 (3.94)	24.93 (3.11)	24.33 (2.61)	25.13 (3.44)	0.16	0.923
Year of medical study (0–4 years)	1.80 (0.68)	1.73 (0.80)	2.13 (0.74)	1.93 (0.88)	0.77	0.516
Surgical interest (low “1”–high “5”)	3.67 (0.72)	3.87 (0.99)	4.13 (0.99)	3.87 (0.74)	0.72	0.542
No. of surgeries observed	0.73 (0.46)	0.47 (0.52)	0.87 (0.35)	0.67 (0.49)	1.45	0.239
No. of dummy sutures completed	3.07 (2.79)	2.33 (3.11)	5.60 (7.39)	3.47 (5.33)	1.18	0.328
No. of live sutures completed	1.60 (4.14)	0.33 (0.82)	3.60 (12.85)	0.60 (0.99)	0.72	0.546
3D experience (h/last year)	0.93 (1.71)	1.80 (2.34)	2.67 (7.39)	8.60 (25.73)	0.99	0.401
Video-game play (h/last year)	169.27 (244.19)	86.13 (143.73)	88.80 (153.03)	84.47 (112.25)	0.89	0.454
Snooker play (h/last year)	11.33 (22.37)	13.60 (30.25)	33.87 (63.20)	11.07 (21.24)	1.24	0.305
Grit (score from 8 to 40) <sup>b</sup>	28.07 (4.03)	30.73 (3.69)	31.27 (3.73)	28.87 (4.81)	2.06	0.116
Visual-spatial ability (score out of 24) <sup>b</sup>	12.00 (5.35)	12.53 (4.67)	13.27 (5.50)	13.93 (3.97)	0.44	0.722
Dexterity—right (score out of 25) <sup>b</sup>	15.07 (1.67)	15.33 (1.54)	15.47 (1.89)	15.07 (1.94)	0.19	0.901
Dexterity—left (score out of 25) <sup>b</sup>	14.40 (1.68)	14.00 (1.69)	14.60 (1.76)	14.07 (2.28)	0.34	0.795
Dexterity—both (score out of 25) <sup>b</sup>	12.00 (1.73)	12.27 (1.28)	12.07 (1.67)	12.07 (1.83)	0.07	0.974
Dexterity—assembly (score out of 100) <sup>b</sup>	43.00 (5.48)	41.40 (3.85)	43.40 (4.67)	44.00 (9.89)	0.45	0.718
Dexterity – L + R + both (score out of 75) <sup>b</sup>	41.47 (4.27)	41.60 (3.62)	42.13 (4.55)	41.20 (5.53)	0.11	0.953
Stereoacuity (20–400) <sup>c</sup>	37.67 (21.29)	29.67 (13.56)	34.67 (15.86)	31.33 (15.52)	0.67	0.572
Visual Acuity—right (1.5–30) <sup>c</sup>	2.40 (0.46)	2.81 (0.60)	3.51 (1.99)	2.37 (0.48)	3.53	<b>0.021*</b>
Visual acuity—left (1.5–30) <sup>c</sup>	2.47 (0.62)	3.07 (0.87)	3.50 (2.44)	2.40 (0.49)	2.24	0.094

*df*=(3, 56) for all analyses

<sup>a</sup>Chi-square test

<sup>b</sup>Higher score reflects a greater level of the ability

<sup>c</sup>Lower score reflects a greater level of acuity

\*Games–Howell post hoc tests revealed no significant group differences (all *p*'s > 0.10)

**Control vs. control** Post hoc tests revealed a significant difference between the two control groups (3D→3D and 2D→2D), with those trained and tested in 3D performing significantly better than those trained and tested in 2D (Table 3).

**Transfer vs. transfer** Post hoc tests comparing the two transfer groups revealed that the 2D→3D group performed significantly better during testing than the 3D→2D group (Table 3).

**Table 2** Training data: mean repetitions, total performance scores, simulator comfort, and perceived workload across the four groups

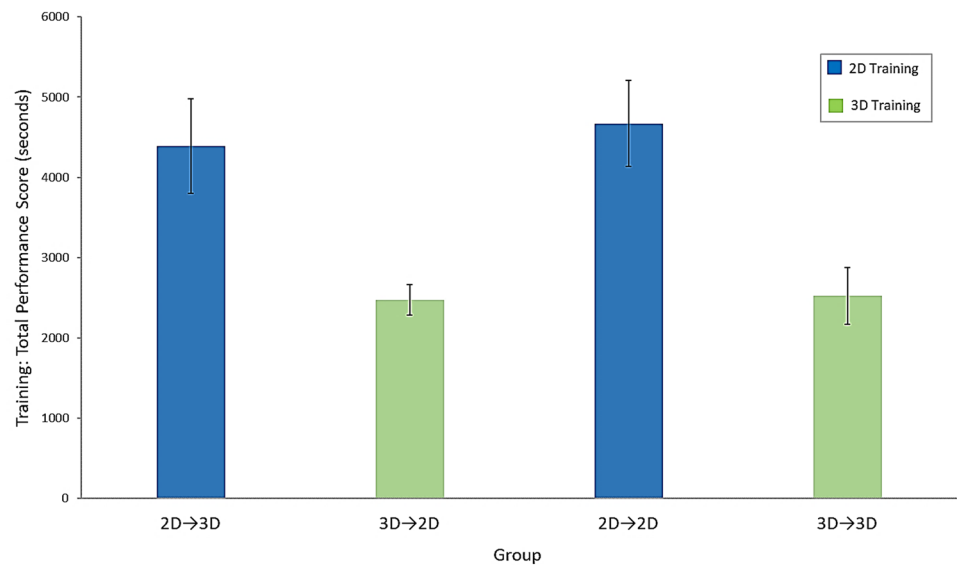
Measure	2D → 3D transfer M (SD)	3D → 2D transfer M (SD)	2D → 2D control M (SD)	3D → 3D control M (SD)	F	p	$\eta^2$	Post hoc tests	Comparison	Result	$p^a$
Total repetitions	34.07 (11.99)	22.40 (5.60)	35.67 (12.51)	23.40 (8.19)	7.27	<.001	.28	Baseline checks	2D → 3D vs 2D → 2D 3D → 2D vs 3D → 3D	N/A N/A	1.00 1.00
								2D vs. 3D training	2D → 3D vs 3D → 2D 2D → 3D vs 3D → 3D	2D > 3D 2D > 3D	.008 .015
									3D → 2D vs 2D → 2D 2D → 2D vs 3D → 3D	2D > 3D 2D > 3D	.006 .006
Total performance score (s)	4389.67 (2286.89)	2470.53 (730.25)	4670.33 (2083.22)	2522.47 (1360.36)	6.99	<.001	.27	Baseline checks	2D → 3D vs 2D → 2D 3D → 2D vs 3D → 3D	N/A N/A	1.00 1.00
								2D vs. 3D training	2D → 3D vs 3D → 2D 2D → 3D vs 3D → 3D	2D > 3D 2D > 3D	.006 .006
Simulator sickness score (1 <sup>st</sup> training)	2.07 (2.12)	1.60 (1.50)	2.47 (2.56)	2.07 (2.19)	.417	.742	.02	N/A	N/A	N/A	N/A
Simulator sickness score (2 <sup>nd</sup> training)	1.93 (1.87)	1.00 (1.25)	1.80 (1.97)	2.47 (2.61)	1.40	.254	.07	N/A	N/A	N/A	N/A
NASA-TLX score (1 <sup>st</sup> training)	26.07 (5.17)	22.53 (4.56)	27.60 (5.12)	24.80 (5.62)	2.61	.060	0.12	N/A	N/A	N/A	N/A
NASA-TLX score (2 <sup>nd</sup> training)	24.13 (4.73)	23.67 (6.40)	25.00 (5.37)	25.60 (5.83)	0.36	0.785	0.02	N/A	N/A	N/A	N/A

Bolded items mean group differences are significant at a level of  $p < 0.05$ ; and italicised headings reflect the purpose of each group of post-hoc tests  $df = (3, 56)$  for all analyses

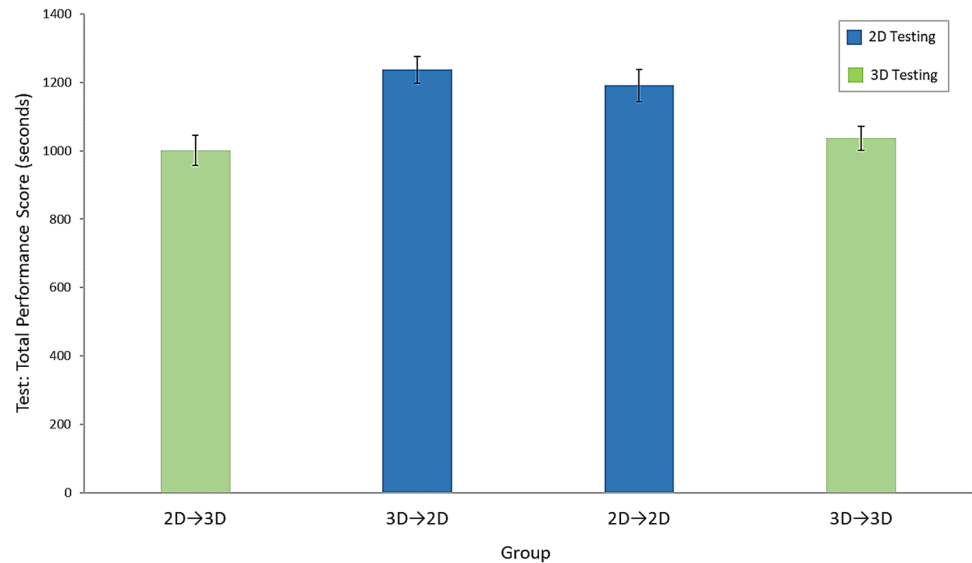
<sup>a</sup> Adjusted  $p$  values based on Bonferroni-Holm Sequential Method ( $\alpha = .05$ )



**Fig. 3** Training data: mean total performance scores across the four groups. Error bars represent standard errors



**Fig. 4** Test data: mean total performance scores across the four groups. Error bars represent standard errors



**Same training, different testing** Among the groups that trained in 2D, the 2D → 3D transfer group performed significantly better than the 2D → 2D control group during testing. Among those that trained in 3D, the 3D → 2D transfer group performed significantly worse compared to the 3D → 3D control group (Table 3).

**Different training, same testing** Regardless of the viewing mode used in training, there were no significant differences found between the test scores of the two groups tested in 2D (the 2D → 2D and 3D → 2D groups), and between the two groups tested in 3D (the 3D → 3D and 2D → 3D groups) (Table 3).

### Novel task

There was a significant main effect of group on the total time taken to complete the novel (NIS) task. Specifically, those tested in 3D (the 3D → 3D and 2D → 3D groups) took significantly less time to complete the novel task than those tested in 2D (the 2D → 2D and 3D → 2D groups). All other NIS comparisons were non-significant (Table 3).

### Simulator comfort and workload

While there was no significant main effect of group on reported simulator sickness during testing, there was a significant main effect of group on perceived workload (Table 3). The only significant difference was between the 2D → 2D control group

**Table 3** Test data: mean total performance scores, NIS task times, simulator comfort, and perceived workload across the four groups

Measure	2D→3D transfer M (SD)	3D→2D transfer M (SD)	2D→2D control M (SD)	3D→3D control M (SD)	F	p	$\eta^2$	Post hoc tests	Comparison	Result	$p^a$
								Purpose			
Total performance score (s)	1000.80 (168.73)	1237.00 (151.92)	1191.33 (182.51)	1036.67 (137.31)	7.69	<0.001	0.29	Control vs. control	2D→2D vs 3D→3D	2D>3D	0.033
								<i>Transfer vs. transfer</i>	2D→3D vs 3D→2D	2D>3D	<0.001
								<i>Same training, different testing</i>	2D→3D vs 2D→2D	2D>3D	0.008
								<i>Different training, same testing</i>	3D→2D vs 3D→3D	2D>3D	0.005
									2D→3D vs 3D→3D	N/A	0.882
									3D→2D vs 2D→2D	N/A	0.882
								<i>Control vs. control</i>	2D→2D vs 3D→3D	2D>3D	<0.001
Navigating in space task—total time (s)	368.27 (159.33)	873.73 (170.80)	794.93 (304.75)	405.93 (162.80)	23.44	<0.001	0.56	<i>Transfer vs. transfer</i>	2D→3D vs 3D→2D	2D>3D	<0.001
								<i>Same training, different testing</i>	2D→3D vs 2D→2D	2D>3D	<0.001
								<i>Different training, same testing</i>	3D→2D vs 3D→3D	2D>3D	<0.001
									2D→3D vs 3D→3D	N/A	0.623
									3D→2D vs 2D→2D	N/A	0.610
Simulator sickness score	2.80 (2.76)	2.33 (2.16)	3.60 (2.64)	2.33 (2.09)	0.91	0.444	0.05	N/A	N/A	N/A	N/A
NASA-TLX score	22.87 (4.03)	27.20 (5.86)	28.33 (4.35)	25.87 (6.94)	2.84	0.046	0.13	Control vs. control	2D→2D vs 3D→3D	N/A	0.654
								<i>Transfer vs. transfer</i>	2D→3D vs 3D→2D	N/A	0.165
								<i>Same training, different testing</i>	2D→3D vs 2D→2D	2D>3D	0.048
								<i>Different training, same testing</i>	3D→2D vs 3D→3D	N/A	1.00
									2D→3D vs 3D→3D	N/A	0.540
									3D→2D vs 2D→2D	N/A	1.00

Bolded items mean group differences are significant at a level of  $p < 0.05$ ; and italicised headings reflect the purpose of each group of post-hoc tests

$df = (3, 56)$  for all analyses

<sup>a</sup>Adjusted  $p$  values based on Bonferroni-Holm Sequential Method ( $\alpha = 0.05$ )

and the 2D→3D transfer group. Specifically, those who transferred from 2D to 3D reported significantly lower perceived workload in testing than those who used 2D only. All other comparisons were non-significant.

## Discussion

Previous studies into the effects of 2D versus 3D viewing modes on laparoscopic skills training were subject to various methodological problems that limit the interpretability of the data. To overcome these limitations, the current study employed control groups to provide greater insights into the effects of viewing mode on laparoscopic performance. In doing so, we confirmed that using a 3D viewing mode did enhance the efficiency of laparoscopic skill acquisition during proficiency-based training compared to 2D. Additionally, the results aligned with the limited research into skill transfer between viewing modes [7, 10–12], with those trained in 2D performing more efficiently and effectively during a post-training test in 3D, compared to those trained in 3D and tested in 2D. Previously, such results have been interpreted as a lack of effective skill transfer from 3D to 2D, and effective or improved skill transfer from 2D to 3D viewing modes [7, 10–12]. However, the omission of control groups in these studies made it impossible to determine whether performance differences in transfer tasks were due to the viewing mode itself, or to the additional practice commonly received in proficiency-based training with 2D. Fortunately, with the inclusion of control groups, we can now rule out the latter interpretation of these results. Specifically, despite receiving the same amount of training in 2D and reaching the same level of proficiency, those who shifted to the 3D viewing mode (2D→3D group) for testing outperformed those who continued using the 2D viewing mode (2D→2D group). This suggests that the change in viewing mode significantly impacted performance. More importantly, the current study highlights that, while training to proficiency in 3D may be more efficient than in 2D, it does not necessarily increase or reduce subsequent performance in 2D (i.e. there were no significant differences in performance between the 3D→2D and 2D→2D groups in testing). This lack of difference between the respective groups implies that the skills obtained during training are transferred to the alternate mode, but only within the limits of the viewing mode itself. In other words, participants who train in 3D can still interpret and make sufficient use of secondary depth cues when later tested in 2D, as they perform just as well as those trained in 2D. Rather, the data suggest that it is the viewing mode itself that defines performance at any stage.

In order to further broaden the investigation of skill transfer, we also set out to assess whether the skills acquired during training could be transferred to a novel task, and whether the viewing mode impacted this ability. Our results aligned with the findings of Sakata et al. [18], as participants completed the NIS task more efficiently when using the 3D viewing mode compared to 2D, regardless of their training mode. This finding was unsurprising given the beneficial impact of 3D displays in complex tasks/environments and the considerable difficulty of the novel NIS task (with fine precision and perceptual discrimination required to navigate and complete the task successfully) [6, 7, 14, 52]. Furthermore, this may help to explain the one significant pairwise difference in subjective workload found between the groups (i.e. 2D→3D vs. 2D→2D) following testing. Specifically, participants who switched to 3D (2D→3D group) likely found the 3-Dmed tasks easier in testing compared to training (as a result of the additional depth cues), and could subsequently exploit the primary depth cues available to complete the complex novel task with less effort than those who continued using 2D (2D→2D group). Further, our adherence to the strict methodological recommendations from Sakata et al.'s [19] review paper on 3D viewing (e.g. maintaining optimal viewing positions to avoid potential crosstalk, screening for stereoacuity to ensure participants could successfully use a 3D display) may have prevented any group differences in perceived simulator sickness during training and testing.

Limitations of the study include an inability to re-assess participants following an extended delay to determine whether the viewing mode impacted long-term skill retention. Additionally, while we were able to address the transfer of skills to the alternate viewing mode and compare this to control groups, assessing subsequent performance upon switching back to the original viewing mode may have provided greater insight into how exposure to alternate modes further impacts performance. Finally, given the difficulty of the NIS task, particularly with 2D viewing, it may have proven valuable to use multiple novel tasks with varying levels of difficulty to more thoroughly assess skill generalisability.

The current findings provide additional support for the use of 3D displays to increase the efficiency of proficiency-based training in basic laparoscopic surgical skills. However, unlike prior research, the present data allowed us to confirm that the viewing mode used in training does not significantly impact subsequent performance in an alternate mode. Overall, the current study provides a robust platform for further research into the structure of training to ultimately optimise the early acquisition of laparoscopic skills.

**Acknowledgements** The authors would like to thank the leadership team and staff of the Clinical Skills Development Service, Metro North Hospital and Health Service, for supporting this research including access to the laparoscopic equipment. Particular thanks to Daniel Host for his technical assistance with the apparatus.

## Compliance with ethical standards

**Disclosures** Kirsty L. Beattie, Andrew Hill, Mark S. Horswill, Philip M. Grove, and Andrew R. L. Stevenson have no conflicts of interest or financial ties to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Agrusa A, Buono G, Chianetta D, Sorce V, Citarrella R, Galia M et al (2016) Three-dimensional (3D) versus two-dimensional (2D) laparoscopic adrenalectomy: a case-control study. *Int J Surg* 28(1):114–117. <https://doi.org/10.1016/j.ijvsu.2015.12.055>
- Archer SB, Brown DW, Smith CD, Branum GD, Hunter JG (2001) Bile duct injury during laparoscopic cholecystectomy: results of a national survey. *Ann Surg* 234(4):549–559
- Moore MJ, Bennett CL (1995) The learning curve for laparoscopic cholecystectomy. The Southern Surgeons Club. *Am J Surg* 170(1):55–59
- Öztürk T (2011) Risks associated with laparoscopic surgery. In: Shamsa A (ed) *Advanced laparoscopy*. InTech, Rijeka, pp 1–27. <https://doi.org/10.5772/18033>
- Dawidek MT, Roach VA, Ott MC, Wilson TD (2017) Changing the learning curve in novice laparoscopists: incorporating direct visualization into the simulation training program. *J Surg Educ* 74(1):30–36. <https://doi.org/10.1016/j.jsurg.2016.07.012>
- Alaraimi B, Bakbak WEI, Sarker S, Makkiyah S, Almarzouq A, Goriparthi R et al (2014) A Randomized prospective study comparing acquisition of laparoscopic skills in three-dimensional (3D) vs. Two-dimensional (2D) laparoscopy. *World J Surg* 38:2746–2752. <https://doi.org/10.1007/s00268-014-2674-0>
- Poudel S, Kurashima Y, Watanabe Y, Ebihara Y, Tamoto E, Murakami S et al (2017) Impact of 3D in the training of basic laparoscopic skills and its transferability to 2D environment: a prospective randomized controlled trial. *Surg Endosc* 31:1111–1118. <https://doi.org/10.1007/s00464-016-5074-8>
- Sorensen SM, Savran MM, Konge L, Bjerrum F (2016) Three-dimensional versus two-dimensional vision in laparoscopy: a systematic review. *Surg Endosc* 30(1):11–23. <https://doi.org/10.1007/s00464-015-4189-7>
- Arezzo A, Vettoretto N, Francis NK, Bonino MA, Curtis NJ, Amparore D et al (2019) The use of 3D laparoscopic imaging systems in surgery: EAES consensus development conference 2018. *Surg Endosc* 33(10):3251–3274. <https://doi.org/10.1007/s00464-018-06612-x>
- Harada H, Kanaji S, Nishi M, Otake Y, Hasegawa H, Yamamoto M et al (2018) The learning effect of using stereoscopic vision in the early phase of laparoscopic surgical training for novices. *Surg Endosc* 32(2):582–588
- Blavier A, Gaudissart Q, Cadiere GB, Nyssen AS (2007) Comparison of learning curves and skill transfer between classical and robotic laparoscopy according to the viewing conditions: implications for training. *Am J Surg* 194(1):115–121. <https://doi.org/10.1016/j.amjsurg.2006.10.014>
- Gómez-Gómez E, Carrasco-Valiente J, Valero-Rosa J, Campos-Hernández JP, Anglada-Curado FJ, Carazo-Carazo JL et al (2015) Impact of 3D vision on mental workload and laparoscopic performance in inexperienced subjects. *Actas Urol Español (English Edition)* 39(4):229–235. <https://doi.org/10.1016/j.acuroe.2015.03.006>
- Sweet BT, Kaiser MK (2011) Depth perception, cueing, and control. In *AIAA modeling and simulation technologies conference and exhibit*, Portland, OR. AIAA 2011–6424.
- Ashraf A, Collins D, Whelan M, O'Sullivan R, Balfe P (2015) Three-dimensional (3D) simulation versus two-dimensional (2D) enhances surgical skills acquisition in standardised laparoscopic tasks: A before and after study. *Int J Surg* 14:12–16. <https://doi.org/10.1016/j.ijvsu.2014.12.020>
- Lin CJ, Cheng C, Chen H, Wu K (2017) Training performance of laparoscopic surgery in two- and three-dimensional displays. *Surg Innov* 24(2):162–170. <https://doi.org/10.1177/1553350617692638>
- Ozsoy M, Kallidonis P, Kyriazis I, Panagopoulos V, Vasilas M, Sakellaropoulos GC, Liatsikos E (2015) Novice surgeons: do they benefit from 3D laparoscopy? *Lasers Med Sci* 30(4):1325–1333. <https://doi.org/10.1007/s10103-015-1739-0>
- Votanopoulos K, Brunnicardi FC, Thornby J, Bellows CF (2008) Impact of three-dimensional vision in laparoscopic training. *World J Surg* 32(1):110–118. <https://doi.org/10.1007/s00268-007-9253-6>
- Sakata S, Grove PM, Hill A, Watson MO, Stevenson ARL (2017) Impact of simulated three-dimensional perception on precision of depth judgements, technical performance and perceived workload in laparoscopy. *Br J Surg* 104(8):1097–1106. <https://doi.org/10.1002/bjs.10528>
- Sakata S, Watson MO, Grove PM, Stevenson ARL (2016) The conflicting evidence of three-dimensional displays in laparoscopy: a review of systems old and new. *Ann Surg* 263(2):234–239. <https://doi.org/10.1097/SLA.0000000000001504>
- Nomura T, Matsutani T, Hagiwara N, Fujita I, Nakamura Y, Kanazawa Y et al (2018) Characteristics predicting laparoscopic skill in medical students: nine years' experience in a single center. *Surg Endosc* 32(1):96–104. <https://doi.org/10.1007/s00464-017-5643-5>
- White MT, Welch K (2012) Does gender predict performance of novices undergoing Fundamentals of Laparoscopic Surgery (FLS) training? *Am J Surg* 203(3):397–400
- Duckworth AL, Peterson C, Matthews MD, Kelly DR (2007) Grit: perseverance and passion for long-term goals. *J Personal Soc Psychol* 9:1087–1101
- Dam A, Perera T, Jones M, Haughy M, Gaeta T (2019) The relationship between grit, burnout, and well-being in emergency medicine residents. *AEM Educ Training* 3(1):14–19
- Kelly AM, Townsend KW, Davis S, Nouryan L, Bostrom MPG, Felix KJ (2018) Comparative assessment of grit, conscientiousness, and self-control in applicants interviewing for residency positions and current orthopaedic surgery residents. *J Surg Educ* 75(3):557–563

25. Verrier ED (2017) The elite athlete, the master surgeon. *J Am Coll Surg* 224(3):225–235
26. Gallagher AG, Cowie R, Crothers I, Jordan-Black JA, Satava RM (2003) PicSOR: an objective test of perceptual skill that predicts laparoscopic technical skill in three initial studies of laparoscopic performance. *Surg Endosc* 17(9):1468. <https://doi.org/10.1007/s00464-002-8569-4>
27. Stefanidis D, Korndorffer JR, Black FW, Dunne JB, Sierra R, Touchard CL et al (2006) Psychomotor testing predicts rate of skill acquisition for proficiency-based laparoscopic skills training. *Surgery* 140(2):252–262. <https://doi.org/10.1016/j.surg.2006.04.002>
28. Peters M, Laeng B, Latham K, Jackson M, Zaiyouna R, Richardson C (1995) A redrawn Vandenberg & Kuse mental rotations test: different versions and factors that affect performance. *Brain Cogn* 28:39–58
29. Brandt MG, Davies ET (2006) Visual-spatial ability, learning modality and surgical knot tying. *Can J Surg* 49(6):412–416
30. Mistry M, Roach VA, Wilson TD (2013) Application of stereoscopic visualization on surgical skill acquisition in novices. *J Surg Educ* 70(5):563–570. <https://doi.org/10.1016/j.jsurg.2013.04.006>
31. Roach VA, Brandt MG, Moore CC, Wilson TD (2012) Is three-dimensional videography the cutting edge of surgical skill acquisition? *Anat Sci Educ* 5(3):138–145. <https://doi.org/10.1002/ase.1262>
32. Wang YC, Magasi SR, Bohannon RW, Reuben DB, McCreath HE, Bubela DJ et al (2011) Assessing dexterity function: a comparison of two alternatives for the NIH toolbox. *J Hand Ther* 24(4):313–321. <https://doi.org/10.1016/j.jht.2011.05.001>
33. Yancosek KE, Howell D (2009) A narrative review of dexterity assessments. *J Hand Ther* 22(3):258–269. <https://doi.org/10.1016/j.jht.2008.11.004>
34. Fawcett SL (2005) An evaluation of the agreement between contour-based circles and random dot-based near stereoacuity tests. *J Am Assoc Pediatr Ophthalmol Strabismus* 9(6):572–578. <https://doi.org/10.1016/j.jaapos.2005.06.006>
35. Simons K (1981) A comparison of the Frisby, random-dot E, TNO, and randot circles stereotests in screening and office use. *Arch Ophthalmol* 99(3):446–452
36. Wang J, Hatt SR, O'Connor AR, Drover JR, Adams R, Birch EE, Holmes JM (2010) Final version of the distance randot stereotest: normative data, reliability, and validity. *J AAPOS* 14(2):142–146. <https://doi.org/10.1016/j.jaapos.2009.12.159>
37. Bailey IL, Lovie-Kitchin JE (2013) Visual acuity testing. From the laboratory to the clinic. *Vision Res* 90:2–9
38. Virgili G, Acosta R, Bentley SA, Giacomelli G, Allcock C, Evans JR (2018) Reading aids for adults with low vision. *Cochrane Database Syst Rev* 4:CD003303. <https://doi.org/10.1002/14651858.CD003303.pub4>
39. Schreuder H, van den Berg C, Hazebroek E, Verheijen R, Schijven M (2011) Laparoscopic skills training using inexpensive box trainers: which exercises to choose when constructing a validated training course. *BJOG Int J Obstetr Gynaecol* 118(13):1576–1584. <https://doi.org/10.1111/j.1471-0528.2011.03146.x>
40. Sakata S, Grove PM, Watson MO, Stevenson ARL (2017) The impact of crosstalk on three-dimensional laparoscopic performance and workload. *Surg Endosc* 31(10):4044–4050. <https://doi.org/10.1007/s00464-017-5449-5>
41. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG (1993) Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *Int J Aviation Psychol* 3(3):203–220
42. Yurko YY, Scerbo MW, Prabhu AS, Acker CE, Stefanidis D (2010) Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. *Simul Healthcare* 5(5):267–271. <https://doi.org/10.1097/SIH.0b013e3181e3f329>
43. Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock PA, Meshkati N (eds) *Human Mental Workload*. North Holland, Amsterdam, pp 139–183
44. Hubert N, Gilles M, Desbrosses K, Meyer J, Felblinger J, Hubert J (2013) Ergonomic assessment of the surgeon's physical workload during standard and robotic assisted laparoscopic procedures. *Int J Med Robot Comput Assist Surg* 9(2):142–147. <https://doi.org/10.1002/rcs.1489>
45. Sakata S, Grove PM, Hill A, Watson MO, Stevenson ARL (2016) The viewpoint-specific failure of modern 3D displays in laparoscopic surgery. *Langenbeck's Arch Surg* 401(7):1007–1018. <https://doi.org/10.1007/s00423-016-1495-z>
46. Holm S (1979) A simple sequential rejective method procedure. *Scand J Stat* 6:65–70
47. Games PA, Howell JF (1976) Pair wise multiple comparison procedures with unequal n's and/or variances. *J Educ Stat* 1:13–125
48. Cohen J (1988) *Statistical power analysis for the behavioral sciences*. Routledge Academic, New York
49. Tukey JW (1977) *Exploratory data analysis*. Addison-Wesley, Reading
50. Dawson R (2011) How significant is a boxplot outlier? *J Stat Educ* 19(2). Retrieved October 29, 2019 from <https://www.amstat.org/publications/jse/v19n2/dawson.pdf>
51. Yeudall LT, Fromm D, Reddon R, Stefanyk WO (1986) Normative data stratified by age and sex for 12 neuropsychological tests. *J Clin Psychol* 42:918–946
52. Sorensen SMD, Konge L, Bjerrum F (2017) 3D vision accelerates laparoscopic proficiency and skills are transferable to 2D conditions: a randomized trial. *Am J Surg* 214(1):63–68. <https://doi.org/10.1016/j.amjsurg.2017.03.001>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.