ORIGINAL ARTICLE

Examining validity evidence for a simulation-based assessment tool for basic robotic surgical skills

Maria Cecilie Havemann[1] · Torur Dalsgaard[1] · Jette Led Sørensen[1] · Kristin Røssaak[1] · Steffen Brisling[2] · Berit Jul Mosgaard[1] · Claus Høgdall[1] · Flemming Bjerrum[3]

Abstract

Increasing focus on patient safety makes it important to ensure surgical competency among surgeons before operating on patients. The objective was to gather validity evidence for a virtual-reality simulator test for robotic surgical skills and evaluate its potential as a training tool. Surgeons with varying experience in robotic surgery were recruited: novices (zero procedures), intermediates (1–50), experienced (> 50). Five experienced surgeons rated five exercises on the da Vinci Skills Simulator. Participants were tested using the five exercises. Participants were invited back 3 times and completed a total of 10 attempts per exercise. The outcome was the average simulator performance score for the 5 exercises. 32 participants from 5 surgical specialties were included. 38 participants completed all 4 sessions. A moderate correlation between the average total score and robotic experience was identified for the first attempt (Spearman $r = 0.58$; $p = 0.0004$). A difference in average total score was observed between novices and intermediates [median score 61% (IQR 52–66) vs. 83% (IQR 75–91), adjusted $p < 0.0001$], as well as novices and experienced [median score 61% (IQR 52–66) vs. 80 (IQR 69–85), adjusted $p = 0.002$]. All three groups improved their performance between the 1st and 10th attempts ($p < 0.00$). This study describes validity evidence for a virtual-reality simulator for basic robotic surgical skills, which can be used for assessment of basic competency and as a training tool. However, more validity evidence is needed before it can be used for certification or high-stakes assessment.

Keywords Robotic surgery · Simulation · Virtual reality · Validity · Assessment

Introduction

Introducing new surgical methods can initially lead to longer operating time and increased risk of adverse events, which can pose a risk to patient safety [1, 2].

Over the last decade, robotic-assisted laparoscopic surgery (RALS) has become increasingly common, which has led to a need for structured training programmes and training tools to learn RALS [3]. Several randomised studies have shown that simulation-based training in laparoscopy shortens the learning curve for inexperienced surgeons, reduces operating time and reduces the risk of adverse events during the initial procedures [4–6].

✉ Maria Cecilie Havemann
mchavemann@dadlnet.dk

Torur Dalsgaard
torur.dalsgaard@regionh.dk

Jette Led Sørensen
jette.led.Soerensen@regionh.dk

Kristin Røssaak
krossaak@dadlnet.dk

Steffen Brisling
skbr@regionsjaelland.dk

Berit Jul Mosgaard
berit.jul.mosgaard.01@regionh.dk

Claus Høgdall
claus.hogdall@regionh.dk

Flemming Bjerrum
fbjerrum@gmail.com

1 Department of Gynaecology, Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet-Glostrup University Hospital, University of Copenhagen, Blegdamsvej 9, 2100 Copenhagen, Denmark

2 Department of Surgery, Zealand University Hospital, University of Copenhagen, Copenhagen, Denmark

3 Department of Surgery, Herlev Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark

Whereas much research has focused on simulation-based training of laparoscopic surgery, RALS has received comparatively little attention [7–9].

An important aspect of modern surgical education is that it utilises competency-based training instead of time-based training or numbers of performed procedures. Therefore, the implementation of tools to assess surgical skills is important and necessary. Simulation-based assessments can ensure that novices possess basic skills before operating on real patients, which may help increase patient safety [3, 7]. However, before using a simulation-based test to assess competency, it is imperative that the assessment tool is examined to ensure its relevance, validity and reliability [10].

It is especially important to determine a relevant pass/fail level using relevant standard setting methods to ensure a relevant level of competency. However, there is currently no consensus about how and which method to apply [11]. One example of a simulation-based test for robotic surgical skills is the fundamentals of robotic surgery, which is used for dry-lab training. However, no other simulation-based assessment tools have gained widespread use, despite several virtual reality robotic simulators being available [12, 13]. The da Vinci Skills Simulator (dVSS) is the only robotic simulator available today that uses the actual robotic console. Therefore, it is believed to be the simulator system that is closest to the real-life experience of performing surgery in the da Vinci robotic system.

The objective of the present study was to examine validity evidence for the dVSS as an assessment tool for robotic skills and to evaluate the usefulness of the dVSS as a training tool. We also wished to assess the implications of setting a pass/fail level based on the initial performance of experienced surgeons. This was done by comparing the level of participants before and after simulator training, based on which a relevant pass/fail level was determined.

Materials and methods

We designed a validation study examining the da Vinci Skills Simulator and used the unitary framework to describe the validity evidence [10, 14].

Participants and settings

Surgeons from five different specialties at three hospitals in the eastern part of Denmark (Rigshospitalet-Glostrup University Hospital, Herlev-Gentofte University Hospital and Zealand University Hospital) were invited to participate in the study. The inclusion criterion was surgeons (in their final year of residency or specialists) who were about to or were already performing RALS procedures. They were grouped according to the number of previously performed robotic procedures: novice surgeons (0 procedures), intermediate surgeons (1–50 procedures) and experienced surgeons (> 50 procedures). Novices with more than three hours of simulation-based robotic surgical training were excluded.

Description of the simulator and selection of exercises

The dVSS (Intuitive Surgical®, California, USA) was used in combination with the da Vinci® Si™ console.

The simulator has a built-in scoring system consisting of metrics such as time to complete exercise, economy of motion and excessive instrument force, as well as task-specific metrics. Based on a weighted average of these parameters, a total performance score was calculated for each exercise. All data were stored automatically on the simulator.

An experienced robotic surgeon evaluated all 40 simulator exercises on the dVSS and selected the 10 most relevant. Subsequently, 5 experienced robotic surgeons ranked the 10 exercises from 1 to 10, 1 being the most relevant and 10 the least relevant. The five exercises with the highest average ranking were chosen for further testing.

Testing of exercises

All participants completed a questionnaire on baseline demographics. All participants received a short oral introduction to the dVSS and performed an exercise (Peg board 2 exercise) to familiarise themselves with the simulator. They, then completed the five exercises once to familiarise themselves with the particular exercises. Finally, they completed all five exercises again and this was used to assess their baseline performance (first attempt).

Participants were then invited back for three additional sessions on the dVSS with an interval of 1–4 weeks between sessions. At each session, they completed three repetitions of the five exercises. Between sessions, the experts and intermediates were not allowed to practice on the dVSS and novices were not allowed to perform robotic-assisted surgery.

After each repetition of the five exercises, the participants filled out a Subjective Mental Effort Questionnaire (SMEQ) to measure their cognitive load. Using the SMEQ, they rated the amount of mental effort they felt they had used, on a scale of 0 to 150 [15, 16]. The scale included nine markers with corresponding statements from "not at all hard to do" to "tremendously hard to do".

An investigator (MC or FB) was present during all sessions and was only allowed to provide technical support to the participants. The investigator did not provide feedback or instructions on how to complete the exercises.

Outcomes

The primary outcome was the average total performance score for the five exercises for the first attempt. Secondary outcome was the average total performance score for the five exercises for the 10th attempt. Exploratory outcomes were the SMEQ score used to assess cognitive load after the 1st and 10th attempts.

Statistical analysis

The Kruskal–Wallis test was used to examine differences between the three groups of surgeons with different level of robotic experience. Group-wise comparisons, using the Wilcoxon Rank Sum test with Benjamini-Hochberg correction, were performed to identify differences between novice and intermediate, novice and experienced, and intermediate and experienced surgeons [17]. The Spearman rank correlation coefficient (r_s) was calculated to examine the correlation between the number of robotic procedures performed and the average total performance score.

The 9th and 10th attempt was used to calculate the intraclass correlation coefficient (ICC), with single measures and absolute agreement definition, to evaluate the internal structure.

To investigate the consequence of testing and to define an acceptable pass/fail level, we examined how many in the novice and experienced group would pass or fail at their 1st and 10th attempts, respectively. We calculated the pass/fail levels using the experienced group average performance score for all five exercises for the experienced group minus one standard deviation. Based on this, we decided on a relevant pass/fail level. This was reported using both frequencies and percentages.

To examine the effect of simulator training, we calculated the Wilcoxon signed-rank test and effect size, comparing the 1st attempt with the 10th attempt. To assess whether simulator training had an impact on cognitive load, the SMEQ score for the 1st and 10th attempts were compared (Fig. 1).

We used SPSS® version 21.0 (IBM, Armonk, New York, USA) for statistical analysis. Complete case analysis was performed and participants who did not complete all four training sessions were excluded from analysis of the secondary and exploratory outcomes. A two-sided significance level of 0.05 was used.

Ethics and funding

The study was submitted to the Regional Scientific Ethical Committee (Ref. No. H-1-2013-FSP-73), which found that no approval was necessary to carry out the study. All

Fig. 1 Surgeon practising on the da Vinci Skills Simulator

Results

The 32 participants were comprised of 11 novices, 11 intermediates and 10 experienced surgeons from the following specialties: gynaecology, urology, gastrointestinal surgery, thoracic surgery and paediatric surgery. The participants' baseline demographics are shown in Table 1. 28 participants completed all training sessions; 4 participants did not complete all sessions and only data from their first attempt were included.

The 10 exercises evaluated for relevance and the 5 final exercises included in the test are shown in Table 2 and Fig. 2, respectively.

There was a significant difference in the average total score between the three groups ($p < 0.0001$) (Table 3). Group-wise comparisons revealed a significant difference between novices and intermediates ($p < 0.0001$), as well as between novices and experienced surgeons ($p = 0.002$) for the first attempt. No significant difference was found between intermediates and experienced surgeons ($p = 0.36$).

Table 1 Participant baseline demographics

	Novices ($n=11$)	Intermediates ($n=11$)	Experienced ($n=10$)
Sex (men/women)	3/8	5/6	5/5
Position			
Senior resident	3	0	0
Specialist/consultant	8	11	10
Number of robotic procedures performed	0 (0–0)	40 (13–45)	125 (50–300)

Values are in median and interquartile range

Table 2 The da Vinci Skills Simulator Exercises ranked for inclusion in the final test

Exercise name	Average ranking (1 being most relevant and 10 least relevant)
Camera targeting 2	6.4
Ring walk 3[a]	3.2
Energy dissection 2[a]	6
Suture sponge 3[a]	3.2
Tubes	6.4
Peg board 2	7
Match board 3	6.2
Ring and rail 2	6.6
Needle targeting[a]	3.8
Thread the rings[a]	6.2

[a]Exercises included in the final test

A moderate correlation was identified between the average total score for the first attempt and the number of robotic procedures performed ($r_s = 0.58$; $p = 0.0004$).

All groups showed a significant improvement in their average performance score between the 1st and 10th attempts and effect sizes varied between 0.60 and 0.63.

The SMEQ score was reduced in all three groups from the 1st to 10th attempts, although the reduction for the intermediate group was minimal and not significant. The results of the intragroup comparisons are shown in Table 4.

The intraclass correlation coefficient was calculated to 0.825 ($p < 0.0001$).

To explore the consequences of testing, we calculated a pass/fail level using the 1st and 10th attempts for the experienced group. Using the first attempt, the average performance score had to be above 69% to pass, and when using their 10th attempt the level was set at 80%. As shown in Table 5, using the first attempt resulted in a pass rate of 18% for novices and 80% for experienced users; using the 10th attempt instead resulted in a pass rate of 0% for novices and 50% for experienced surgeons.

Based on these results, we decided on a pass/fail level on 75%. The number of participants who passed or failed in each group at each of these levels is summarised in Table 5.

Discussion

In this study, we have evaluated the validity evidence of the dVSS simulation-based assessment tool for basic robotic surgical skills and found that it can be used for training and assessment of basic robotic skills.

We have identified five content relevant exercises, which were further evaluated as a combined test.

We chose the number of robotic procedures to assess a relationship to an external variable, and demonstrated a moderate correlation with test score. We also found a significant difference between the novices and the intermediate and experienced surgeons, respectively.

We examined the internal structure of the test, which revealed a high intraclass correlation coefficient; this indicates high reliability of the average performance score and supports its use as a reliable assessment tool.

Furthermore, the automated registration of data and the standardised nature of the simulator ensured a uniform data collection and minimised bias in the response process.

We decided that the pass/fail level would be an average score of 75% for the five exercises. This level can be used for formative assessment and as a training goal for proficiency-based training, and also to determine whether a surgeon has acquired the necessary robotic skills, before moving on to more advanced simulator tasks and supervised operations.

We observed a reduction in cognitive load for novices and experienced surgeons, as assessed by the SMEQ, although there was no significant change for the intermediate group. This supports the use of simulators as a tool for reducing the psychological stress, when learning new skills or procedures. This has also been observed in another study [18]. As shown in Table 3, the novices' SMEQ score after 10 repetitions had been reduced to approximately the starting level for the experienced surgeons. The reduction in cognitive load and the improvement in performance for all three groups
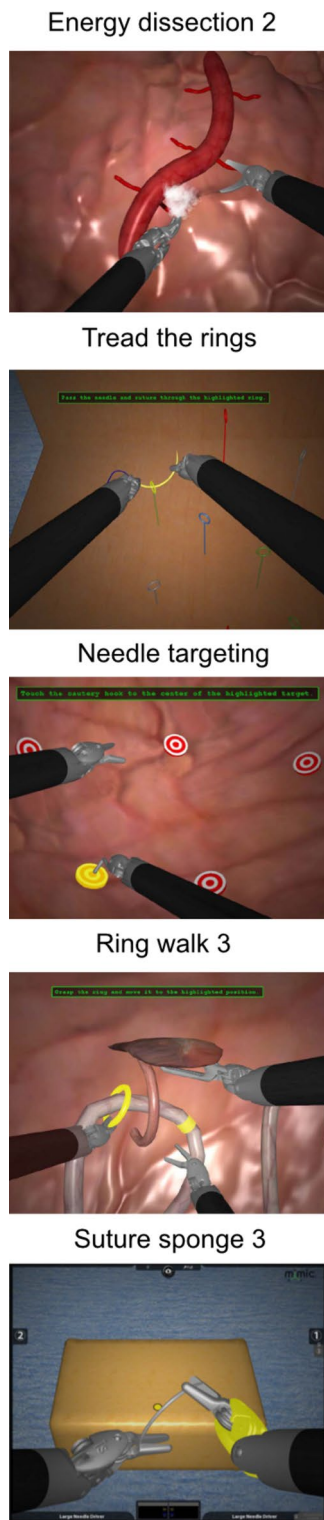
Energy dissection 2

Tread the rings

Needle targeting

Ring walk 3

Suture sponge 3

Fig. 2 Screenshots of the five exercises included in the final test. Used with permission of Intuitive Surgical®, California, USA

demonstrates that the simulator has potential as a training tool. Interestingly, the experienced group also reduced their

cognitive load while practising, which could indicate that they also needed further familiarisation with the simulator.

Overall, the proposed simulation-based test on the dVSS can be used both as an assessment tool and as a training tool. It demonstrates high reliability and is able to discriminate between groups with different levels of experience, although only a moderate correlation is observed with robotic experience. We believe that the test is currently best suited for formative assessment of skills before starting to perform RALS in the operating room. Before the assessment tool can be recommended for certification of surgeons (that is, "high stakes" assessment), more studies examining validity evidence are needed. The participants' simulation-based performance improved with training and a reduction in cognitive load was also observed, suggesting that it can be used as a training tool.

We used the contemporary framework for validity, as proposed by Messick, which is a strength of this study [10, 14]. Most earlier validity studies on robotic surgery have used older frameworks [3, 9].

Another strength of the present study is that a relative high number of participants were surgeons, and even the novices were all surgical residents with a certain amount of surgical experience. Some previous validity studies have used medical students or inexperienced residents as participants [9]. This is important for the validity argument, because the novices in the present study represent the actual target group for robotic simulator training and competency assessment.

Because we used basic skills exercises, the proposed test is more generalisable. This enabled us to use it for multiple surgical specialties and thereby include a higher number of robotic surgeons, than some of the previous studies [9].

A further strength of this study is that we used a relevant pass/fail level. We used performance over time, instead of simply relying on the initial performance of experienced surgeons. Although the chosen level would fail 60% of the experienced surgeons, they would quickly pass after gaining familiarity with the simulator.

We did not examine the effect of simulation-based training on actual procedures. The number of procedures was used as a surrogate for surgical experience, which is a major limitation of this study. However, it was not feasible to assess the effect of simulation-based training on the actual procedures in the setup for the present study, as it involved multiple institutions. Furthermore, the novices had not undergone the formal robotic course, which meant they were not yet permitted to perform real operations. Because multiple specialties were included, a specific procedure as post-test could not be used, and procedural modules were not available on the simulator.

Also, using the comparison of different groups with varying levels of experience is a commonly used measure of

Table 3 Primary, secondary and exploratory outcomes

Outcome	Novices	Intermediates	Experienced
Average total score at 1st attempt (%)	61.4 (52–66)	83.2 (75.4–90.8)	80 (69.3–85.4)
Average total score at 9th attempt (%)[a]	81.1 (72.8–89.3)	87 (83.4–92.6)	88 (83.7–91.5)
Average total score at 10th attempt (%)	82 (77.5–88.7)	89.4 (83.7–92.2)	85.8 (82.6–94.9)
SMEQ score at 1st attempt (points)	56 (26–70)	26 (26–56)	33.5 (19–56)
SMEQ score at 10th attempt (points)**	26.5 (12.25–38)	25 (13.5–45)	18 (3.5–27.5)

All results are reported in median and interquartile range

**SMEQ score is rated on a scale from 0 to 150 points

[a]Used to calculate intraclass correlation coefficient

Table 4 Intragroup comparisons for average total performance score and Subjective Mental Effort Questionnaire when comparing 1st and 10th attempts

	Novices ($n = 10$)	Intermediates ($n = 9$)	Experienced ($n = 9$)
Average total performance score (Z value/p value/r value)[a]	$Z = -2.803$ $p = 0.005$ $r = 0.63$	$Z = -2.549$ $p = 0.011$ $r = 0.60$	$Z = -2.666$ $p = 0.008$ $r = 0.63$
SMEQ score (Z value/p value/r value)[a]	$Z = -2.654$ $p = 0.008$ $r = 0.59$	$Z = -0.141$ $p = 0.888$ $r = 0.03$	$Z = -2380$ $p = 0.017$ $r = 0.56$

Note: Only participants who completed all four training sessions were included in this analysis

[a]Calculated using Wilcoxon signed-rank test

Table 5 Comparison of pass/fail rates for novices and experienced surgeons for their first attempt for different pass/fail settings

	Novices ($n = 11$)	Experienced ($n = 10$)
Based on 1st attempt (number of participants passed/total)[a]	2/11	8/10
Based on 1st attempt (% passed)[a]	18%	80%
Based on 10th attempt (number of participants passed/total)[b]	0/11	5/10
Based on 10th attempt (% passed)[b]	0%	50%
Based on pass/fail score of 75% (number of participants passed/total)	1/11	6/10
Based on pass/fail score of 75% (% passed)	9%	60%

[a]Pass/fail level of 69%

[b]Pass/fail level of 80%

validity, even though the relevance of this has been questioned [19].

We used the simulator generated performance score as the primary outcome. We decided to use the average score for all five exercises, thereby letting the test express the overall performance. However, using the automated score from the simulator could be seen as a limitation. One of the challenges with the dVSS is that users cannot program the pass/fail settings for each parameter for the different exercises. Therefore, to apply a test that was feasible to administer, we decided to use the total score. Choosing different measures for surgeons' performance could have changed the outcome.

The intermediate and experienced surgeons in the study all had some previous simulation-based training experience prior to the study, as they had completed the formalised Intuitive Surgical® robotic surgery training course. However, all of the participants had completed this course a long time before our study and all had subsequently obtained actual operating experience; therefore, we believe the possible bias due to this was minimal.

We defined pass/fail levels as the experts average perfomance minus one standard deviation, although the data, although the data was not normally distributed, which is a limitation. However, we did not find it appropriate to use the contrasting groups method. There is no consensus about which method should be used to define a pass/fail level and the choice of level is ultimately an institutional decision. It is important to emphasise that simulator training is only one aspect of training prior to initiating surgery on actual operations [11].

Simulation-based assessment of robotic surgical skills has great potential for use in the ongoing development of robotic surgical curricula and certification. Virtual reality simulators possess the advantage of standardisation and can generate many different exercises. Unlike laparoscopic simulators, the lack of haptic feedback does not pose a problem, because there is no haptic feedback during robotic surgery.

While no standardised curriculum for robotic surgical training has yet gained widespread use, many have been described and are currently being developed. One example of a curriculum that includes assessment of technical skills is the Fundamentals of Robotic Surgery, which uses real training exercises on the da Vinci © system [12]. Industrial-held courses are currently the standard for certification, but these courses are not evidence-based and relevant training goals on the simulator have not been defined. To optimise training and increase patient safety, it is essential that training is proficiency-based and relevant training goals and exercises are identified. The same process has taken place for laparoscopic simulation-based training over many years and the experiences from these studies can help accelerate the implementation for robotic surgery.

An advantage of the dVSS is that it can be used with an actual da Vinci Surgeon console. This provides the opportunity to practice using the same equipment as in the operating theatre [20]. However, a limiting factor is that we could only test basic skills exercises and not use simulated surgical procedures, as none were available for this simulator.

Our study looked at basic technical skills for robotic surgery. With the further development of simulators, the incorporation of procedural modules should be emphasised after basic skills training and the training of non-technical skills should also be prioritised [21]. A full surgical curriculum is needed to combine surgeon technical skills training with theoretical education, followed by inter-professional scenario training in which the entire team learns how to dock the robot, handle equipment and handle complications and unforeseen events.

Conclusion

The da Vinci Skills Simulator can be used to assess basic robotic skills and can be used as a first step in training, before moving on to more advanced tasks. Although the proposed test can be used for assessment of competency and formative feedback, further validity evidence is needed before it can be used for certification and high-stakes assessment.

Compliance with ethical standards

References

1. Aggarwal R, Mytton OT, Derbrew M et al (2010) Training and simulation for patient safety. Qual Saf Health Care 19 Suppl 2:i34–i43. https://doi.org/10.1136/qshc.2009.038562
2. Carreyrou J (2010) Botched operation using Da Vinci robot spurs lawsuit. Wall Street J. https://www.wsj.com/articles/SB10001424 052748703341904575266952674277806. Accessed 30 Apr 2018
3. Schreuder HWR, Wolswijk R, Zweemer RP et al (2012) Training and learning robotic surgery, time for a more structured approach: a systematic review. BJOG 119:137–149. https://doi.org/10.1111/j.1471-0528.2011.03139.x
4. Larsen CR, Oestergaard J, Ottesen BS, Sorensen JL (2012) The efficacy of virtual reality simulation training in laparoscopy: a systematic review of randomized trials. Acta Obstet Gynecol Scand 91:1015–1028. https://doi.org/10.1111/j.1600-0412.2012.01482.x
5. Ahlberg G, Heikkinen T, Iselius L et al (2002) Does training in a virtual reality simulator improve surgical performance? Surg Endosc 16:126–129. https://doi.org/10.1007/s00464-001-9025-6
6. Zendejas B, Cook DA, Hernández-Irizarry R et al (2012) Mastery learning simulation-based curriculum for laparoscopic TEP inguinal hernia repair. J Surg Educ 69:208–214. https://doi.org/10.1016/j.jsurg.2011.08.008
7. Bric JD, Lumbard DC, Frelich MJ, Gould JC (2016) Current state of virtual reality simulation in robotic surgery training: a review. Surg Endosc 30:2169–2178. https://doi.org/10.1007/s00464-015-4517-y
8. Aydin A, Shafi AMA, Shamim Khan M et al (2016) Current status of simulation and training models in urological surgery: a systematic review. J Urol 196:312–320. https://doi.org/10.1016/j.juro.2016.01.131
9. Moglia A, Ferrari V, Morelli L et al (2016) A systematic review of virtual reality simulators for robot-assisted surgery. Eur Urol 69:1065–1080. https://doi.org/10.1016/j.eururo.2015.09.021
10. Downing SM, Yudkowsky R (2009) Assessment in health professions education. Routledge, New York
11. Thinggaard E, Bjerrum F, Strandbygaard J et al (2016) Ensuring competency of novice laparoscopic surgeons-exploring standard setting methods and their consequences. J Surg Educ 73:986–991. https://doi.org/10.1016/j.jsurg.2016.05.008
12. Smith R, Patel V, Satava R (2014) Fundamentals of robotic surgery: a course of basic robotic surgery skills based upon a 14-society consensus template of outcomes measures and curriculum development. Int J Med Robot 10:379–384. https://doi.org/10.1002/rcs.1559
13. Foote JR, Valea FA (2016) Robotic surgical training: where are we? Gynecol Oncol 143:179–183. https://doi.org/10.1016/j.ygyno.2016.05.031

14. Downing SM (2003) Validity: on meaningful interpretation of assessment data. Med Educ 37:830–837

15. Zijlstra F (1993) Efficiency in work behaviour: a design approach for modern tools. Delft University Press

16. Bharathan R, Vali S, Setchell T et al (2013) Psychomotor skills and cognitive load training on a virtual reality laparoscopic simulator for tubal surgery is effective. Eur J Obstet Gynecol Reprod Biol 169:347–352. https://doi.org/10.1016/j.ejogrb.2013.03.017

17. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate—a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol 57:289–300

18. Yang K, Zhen H, Hubert N et al (2017) From dV-trainer to real robotic console: the limitations of robotic skill training. J Surg Educ 74:1074–1080. https://doi.org/10.1016/j.jsurg.2017.03.006

19. Cook DA (2015) Much ado about differences: why expert-novice comparisons add little to the validity argument. Adv Health Sci Educ Theory Pract 20:829–834. https://doi.org/10.1007/s10459-014-9551-3

20. Julian D, Tanaka A, Mattingly P et al (2018) A comparative analysis and guide to virtual reality robotic surgical simulators. Int J Med Robot 14:e1874. https://doi.org/10.1002/rcs.1874

21. Ahmed K, Khan R, Mottrie A et al (2015) Development of a standardised training curriculum for robotic surgery: a consensus statement from an international multidisciplinary group of experts. BJU Int 116:93–101. https://doi.org/10.1111/bju.12974